Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

5-24-2012

Comparative and Functional Genomic Analysis of the Methanobrevibacter smithii Pan Genome

Elizabeth Hansen
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

Recommended Citation
Hansen, Elizabeth, "Comparative and Functional Genomic Analysis of the Methanobrevibacter smithii Pan Genome" (2012). *All Theses and Dissertations (ETDs)*. 695.
https://openscholarship.wustl.edu/etd/695

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Molecular Microbiology and Microbial Pathogenesis

Dissertation Examination Committee:
Jeffrey I. Gordon, Chairperson
Jan Amend
Tamara Doering
Daniel Goldberg
Elaine Mardis
Clay Semenkovich

Comparative and Functional Genomic Analysis of the *Methanobrevibacter smithii* Pan Genome

by

Elizabeth Erica Hansen

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2012

Saint Louis, Missouri

Copyright by

Elizabeth Erica Hansen

2012

ABSTRACT OF THE DISSERTATION

Comparative and Functional Genomic Analysis of the
Methanobrevibacter smithii pan genome

by

Elizabeth E Hansen

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Microbiology and Microbial Pathogenesis

Washington University in St. Louis, 2012

Professor Jeffrey I Gordon, Chairperson

The human gastrointestinal tract hosts trillions of microorganisms that impact human health. Although members of the domain Bacteria dominate our gut bioreactors, members of Archaea and Eukarya, and their viruses, are also represented. *Methanobrevibacter smithii* is the dominant Archaeon found in human gut, although only a subset of humans harbor this methanogen. Together with other hydrogen-consuming organisms (acetogens, sulfate reducing bacteria), *M. smithii* plays an important role in determining the efficiency of fermentation of dietary polysaccharides. Thus, deciphering the interactions between methanogens, acetogens and SRB and other members of the gut microbiota offers an important opportunity to gain new insights about how host energy balance is regulated, and new approaches for microbiome-directed attempts to control the partitioning of energy and nutrients from diet to host. This thesis addresses a number of basic questions: what forces determine if *M. smithii* is in a human gut microbiota; what other species in the gut microbial community co-occur with it; how does its genome evolve within an individual and between individuals within and between families. To address these questions, I used qPCR and 16S rRNA-directed pyrosequencing plus a variety of computational tools to examine the representation of *M. smithii* and SRB and other taxa in the fecal microbiota of healthy

adult female monozygotic (MZ) and dizygotic (DZ) twin pairs, developed methods for isolating *M. smithii* from their frozen fecal samples, sequenced the genomes of 20 isolates recovered from two families, one composed of a MZ twin pair and the other of DZ twin pair, and characterized the isolates' transcriptomes using RNA-Seq during in vitro growth under a variety of conditions. My studies revealed (i) carriage and levels of colonization correlate to a significant degree in MZ but not DZ twins, (ii) a core set of genes conserved between isolates; (iii) that based on variations in SNP and gene content/function, strains cluster according to individual host and then family, and (iv) suites of genes, including adhesin-like proteins, that are likely elements driving the evolutionary adaptation of this organism within and between human gut ecosystems. These results should help future efforts to develop strategies for manipulating the hydrogen economy of the gut in ways that promote health.

Acknowledgements

This thesis has been a collaborative effort in a laboratory that cultivates interaction, free exchange of ideas, and incredible support. Jeffrey Gordon's lab is a true reflection of his enthusiasm, dedication and ability to see and do what others think is out of reach. Jeff's mentorship has guided me through a project that has provided amazing learning experience, with the patient support and encouragement that was needed to complete this work. I am so grateful for the support he has given me, both with kind words when I needed them, and with the experimental tools to explore ideas, build new tools, generate vast amounts of data, and the time to learn how to analyze that data. Jeff not only provides the sustenance we all need to do science, but the environment that makes it so much fun to do. The people who are drawn to this lab have incredible qualities – qualities that reflect the talent and drive in our mentor. The lab is full of amazing people with a remarkable array of skills, fresh ideas, a willingness to help a fellow lab-mate through a difficult time, provide feedback for works-in-progress, and of course, many laughs.

Our lab managers, Jill Manchester and Sabrina Wagoner, are essential to the day-to-day workings of the lab, and to the collaborative, friendly nature of the lab. They provide us with the support and guidance we need to find our way through new experiments, and make sure that everyone has a special treat on their birthday. I can't imagine the lab without them. Jill Manchester has been a great help in many aspects of my thesis, including biochemical tests and preparing samples for sequencing. Sabrina Wagoner makes many of the unglamorous jobs required for our type of science look easy, and she is truly willing to do whatever it takes to get the job done, including accompanying me to a butcher shop to retrieve stomach contents from a freshly slaughtered cow, so that I could supplement my media with rumen fluid. She has also saved many knitting projects from disaster, been a comforting shoulder to cry on, and a steady source of advice and wisdom.

Stephanie Amen is a wonderfully calm, competent and amazingly helpful person, without

whom we would all likely be very lost. I am very grateful to her for help with so many tasks, including scheduling thesis updates, keeping track of correspondences and paperwork, and submission of manuscripts. Sharon Armon assists with many organizational tasks as well, and has such a sunny and cheerful personality. Laura Kyro has been a wonderful help organizing figures and tables for manuscripts, preparing PowerPoint slides, giving Illustrator tips, and putting all of that together for manuscripts and even this thesis.

We are all lucky to be park of the Center for Genome Sciences and Systems Biology, a collaborative and smart group of people with diverse backgrounds but a common interest in genomic sciences and big-picture thinking. We have superb technical help from Eric Martin and Brian Koebbe who solve computer problems and maintain our computational cluster.

Janaki Guruge has been incredibly helpful in the isolation of *M. smithii* strains by anaerobic culturing in our custom built PHAT tanks, as well as with verification of purity of strains. Her cheerful attitude and entertaining stories have gotten me through some of the longer days working on culturing. David O'Donnell is part of the duo that maintains our world-class gnotobiotic facility, and together with Maria Karlsson, they provide germ-free mice for large-scale experiments, assist with these experiments, and while doing all that, they remain friendly, upbeat and extremely patient with us, even during repeated sampling of the mice.

There are many alums of the Gordon lab who have been a great influence on my science as well as my life in St Louis. Erica Sonnenburg mentored me through my rotation and taught me new techniques as well as introducing me to the world of *B. theta*. She became a great friend and mentor through our years overlapping the lab. Helene Kling Backhed and Fredrik Backhed were incredibly collaborative scientifically, and also shared their culture with us by hosting annual midsummer parties to celebrate the longest day of the year, and introduced us to traditional Swedish food and games. I still hope to visit them

in Sweden and meet their adorable girls. Daniel Peterson taught me immunological and mouse techniques, gave advice as a graduate of the MSTP, and was ever willing to bounce around ideas and look at results. Peter Crawford is another wonderful MSTP graduate who took the time to give me advice on my professional development as well as medical issues. Chris Reigstad made sure everyone in the lunchroom had at least one good laugh a day, and was always up for looking at data and talking about results. Justin Sonnenburg was always willing to explain a method, or discuss science, and demonstrated with ease how the transition from post-doc to principle investigator was done. My fellow MSTPs, Marios Giannakis and Michael Mahowald are still a source of great advice and inspiration going through graduate school and then back into the clinics. Herb Chaing, Jung Oh and Joe Ippolito, all MSTP alums, also shared helpful advice and insight. Swaine Chen had the great task of introducing me to Linux and scripting, and very patiently showed me how to work on the command line. I am so grateful for this introduction and now know that I have to pay it forward to the next Linux novice. He also assisted with many statistical analyses, and laid the groundwork for the type of pan-genome analysis we have done. Eric Martens and John F. Rawls are both incredibly gifted and careful scientists who showed me that doing difficult experiments was worth the effort with their elegant work. I am so glad that I met Ruth Ley, who has become one of my best friends – the kind of friend you hope to know your whole life. She is a true inspiration to me in so many ways, and it is incredibly gratifying to see her succeed in her lab at Cornell. I fondly remember dinners with Ruth and her husband, Lars Angenent, who talked with me about science and about life. Their hilarious and delightfully charming son, Miles, is my honorary nephew, and I miss them all exceedingly.

With all the sad goodbyes to these former members, there have been as many hellos to the newer members of the lab. Jeff attracts incredible people to his lab, at a scientific level and a personal level, and it is a joy to work beside all of them. Nathan McNulty has been a fantastic bay-mate. I know I can always turn around to find clever, witty and helpful advice on almost any question. He will even tell me when I'm wrong, which is most appre-

ciated. He has also developed computational tools, together with Ansel Hsiao, which were used in my studies of *M. smithii* transcriptional analysis. Both Nate and Ansel were always more than willing to help me adapt their methods to my data. Federico Rey is incredibly smart and hard working, and also friendly, down-to-earth and generous with his time, expertise and knowledge. I have learned a lot from him, and can always count on him for great advice. Henning Seedorf is another fantastic source of advice, and knowledge about methanogens and the hydrogen world. He is also a wonderful person, and I'm grateful for his friendship. Gabriel Simon is the person I most love to argue with at cookie-time, making me re-think the ideas I thought were true.

My fellow MSTP, Brian Muegge, has developed computational tools that have been incredibly helpful for everyone in the lab, and is always willing to help out with a problem or question with thoughtful ideas. Andrew Kau is another MSTP alum who has clearly embodied the physician scientist role, and is a natural teacher, quizzing me about pneumonia while I did benchwork. Vitas Wagoner and Nick Griffin are our ecology experts, and Nick is a wonderful resource for statistical analyses, and bouncing around ideas with. I first met Tanya Yatsunenko before she joined our lab, when she gave me some Archaeal cultures from Yellowstone; we share a love for these fascinating microbes.

Andrew Goodman is a great role model, and has a keen scientific mind. He is helpful and funny, and sometimes gives me a hard time ("please avoid gratuitous use of red tape!"), but is always a pleasure to work with. Mark Gonzalez and Philip Ahern recently joined our lab and have already made the lab more fun. We are especially enjoying Philip's Irish accent and great sense of humor. I think my lab bench will be in good hands after I leave, and I hope he enjoys the window seat! Vanessa Ridaura and Michelle Smith, while not my actual sisters, are sister-like in my affection for them. I am grateful for their friendship, and feel lucky that they will not only talk to me about science, but about everything else too.

Alejandro Reyes has been a great help in many computational aspects of my project, and is just an all around great person. Jeremiah Faith also assisted with many computational tasks, and has developed a lot of tools the rest of the lab now rely on. J and his wife, Ilaria, are grounded people, inspiring in the way they live their lives, and the admirable in things they accomplish. I am very happy to know them both (and their son Eliot!) Meng Wu is an extremely gifted computational biologist who provided me with several scripts used in this work, and is always willing to help out with a computational problem. I very much appreciate her help. Several rotation and undergraduate summer students have been a part of this project, and learned some of the techniques including q-PCR and anaerobic culturing. Erica Scholler, Jonathan Goodfellow, Jonathan Xia, and Tony Jianqing Li all contributed.

Steven Salzberg, Daniella Puiu and Arthur Dechler, as well as other members of the CBCB at the University of Maryland, generously hosted me for a week in their lab, and taught me how to use their software to create the best genome assemblies possible from my sequencing data.

Catherine Lozupone has collaborated with me on the co-occurrence part of my project, and has taught me a lot about phylogenetics, and competition and cooperation between microbes. It is refreshing and extremely helpful to have another perspective on the data, and her approach in this project has opened up a lot of doors for future work.

The incredible team that runs our MSTP has been invaluable over the years, with paperwork and administrative help, general advice, and organization of social events. Brian Sullivan, Andrew Richards, Christy Durbin, and Liz Bayer are assets to our program and our university, and the MSTP would not be the success it is without them. The MSTP has had two directors during my time in the program, Dan Goldberg and Wayne Yokoyama, both of whom are amazing scientists and clearly very dedicated to the MSTP students. David Haslam, my undergraduate research mentor, trusted me to have my own project as a

new biology student, spent time explaining the most basic techniques as well as the bigger picture scientific ideas, and encouraged me to go on to a career in academic medicine and science. I am so honored to be in the MSTP and I'm grateful to everyone who has helped me through the years, and to the MSTP training grant for support.

My thesis committee, including my committee chair, Jan Amend, and my interim chair, Tamara Doering, has been wonderful in providing support and feedback. Other committee members Clay Semenkovich, Daniel Goldberg, and Elaine Mardis, as well as former member Lars Angenent, have provided invaluable feedback and advice as my thesis project progressed and developed.

My family and friends have seen me through this incredible experience of earning a PhD and have given me much needed support, love and wonderful vacations when I needed a bit of rest and relaxation. My parents, Rosie and Fred, have taught me so much, and have been role models to me in so many ways. From their lessons on working out minor tussles with my sister, to their encouragement to try new foods, they prepared me (for the most part) for life as a functioning adult. My sister, Gillian, always seems to know whether I need a laugh or a bit of serious advice, and has been my best friend ever since our parents took us around the world to live in Japan. We later discovered that they were not, in fact, punishing us, but were giving us an amazing opportunity to see the world and learn about another culture. I also thank my extended family – aunts, uncles, cousins, and cousins removed or seconded; I am lucky to have so many amazing relatives. I am grateful for the support and love all of my family has given me. Many friends have helped me through stressful times, kept things in perspective, and talked with me about science and about life. My incredibly smart boyfriend, Chris McCracken, is patient while I need to work, and so much fun when we have time to go hiking or play with our dog. He has turned out to be a huge help in my scientific progress as well, teaching me Linux tricks and shortcuts for command-line scripting.

With the stresses of graduate school and difficult experiments, I have found yoga to be an essential way to step back, look at the big picture, and think more deeply and carefully about a problem. I am lucky to have had many wonderful yoga teachers and co-students, who have helped me grow as a human being, cultivate my breath, live in the present moment, and keep my practice challenging and fun. Thanks to Cindy Dollar, Jane Fitzgerald, Pam Schultze, Agatha Thaller, Kim Winn, Sat Inder Singh, and Brigette Niedringhaus for all that you have taught me, for reminding me that we have teachers all around us, and that we all share a common breath. Namaste.

Dedication

To my family, who has always supported me with laughter, joy and love.

Table of Contents

Chapter 1

Introduction

Chapter 2

Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut

Chapter 3

The pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii* in twins

Chapter 4

Future Directions

List of Figures

Chapter 2

Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut

Chapter 3

The pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*: a study in twins

Chapter 4

Future Directions

List of Tables

Chapter 2

Evolution of symbiotic bacteria in the distal human intestine

Chapter 3

The pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*: a study in twins

Chapter 1

Introduction

Chapter 1

Introduction

The human gastrointestinal tract hosts trillions of microorganisms that impact health. Although members of the domain Bacteria dominate, members of Archaea and Eukarya, and their associated viruses, are also represented in our gut bioreactor. *Methanobrevibacter smithii* is the dominant Archaeon, although only a subset of humans harbor this methanogen. Together with other hydrogen-consuming organisms (acetogens, sulfate reducing bacteria (SRB)), *M. smithii* plays an important role in determining the efficiency of fermentation of dietary polysaccharides. Thus, deciphering the interactions between methanogens, acetogens and SRB and other members of the gut microbiota represents a unique opportunity to gain new insights about factors that affect host energy balance, and new approaches for microbiome-directed attempts to control the partitioning of energy and nutrients from diet to the host.

Role of microbial-host symbiosis in health and disease

Our biosphere is dominated by microbial life. Microorganisms have been in existence for billions of years (Brocks, 1999), and have shaped our planet, impacting atmospheric conditions and changing the landscape to one that supported development of multi-cellular life (Falkowski, 2008). All animals have co-evolved symbiotic relationships with microbes. We are no exception: ~10% of our cells are human; the rest are microbial, with the vast majority of our microbes residing in our distal guts. The gene content of members of our gut microbiota outnumber our own human genes by a factor of at least 100 (Qin *et al*., 2010), and encode metabolic capabilities that are not represented in our *H. sapiens* genomes (Goodman and Gordon, 2010). These metabolic traits include the ability to ferment otherwise indigestible complex dietary plant polysaccharides to absorbable

short-chain fatty acids, and dietary proteins into ammonia, to synthesize essential amino acids and vitamins, and to degrade various xenobiotics. The gut microbiota also plays an important role in our postnatal development, including education of the immune system and intestinal angiogenesis (Hooper and MacPherson, 2010 and Stappenbeck *et al*., 2002)

Host-specific microbiota and the question of the pan-genome

Each person has a unique collection of microorganisms associated with their body: e.g., deep sampling of monozygotic twin pairs has shown that their fecal microbiota shares only 50% of species-level phylotypes (Turnbaugh *et al*., 2010). The fecal microbiota undergoes changes as we age, but interpersonal differences remain greater than intrapersonal variation (Costello *et al*., 2009; Claesson *et al*., 2010). Reciprocal microbiota transplantation experiments involving introduction of a mouse gut microbiota into germ-free zebrafish and a reciprocal transplant of zebrafish gut microbiota into germ-free mice indicate that host factors likely contribute to the selection of gut microbial communities (Rawls *et al*., 2006). Nonetheless, the relative contributions of environmental factors versus host genotype in selecting a (gut) microbiota remain ill defined. Indeed, a comparison of adult monozygotic (MZ) and dizygotic (DZ) twin pairs revealed that the overall phylogenetic structure of the fecal microbiota of MZ pairs was not more similar than that observed in DZ pairs, and that family members (twins and their mothers) shared more similarity in their fecal community composition than unrelated individuals - suggesting the importance of shared early environmental exposures in determining the configuration of a microbiota (Turnbaugh *et al*., 2009a). Furthermore, studies in gnotobiotic mice that have received human fecal microbiota transplants have disclosed the profound influence of diet on community structure (Turnbaugh, 2009b).

Strains of the same species-level phylogenetic type (phylotype) found in two different individuals may have differences in gene content that confer different functions and that allow varied interactions with other members of the microbiota and with the host.

Identifying strain differences within versus between different hosts provides a glimpse at how the microbiota as a whole may be inherited or how it may evolve.

Pan-genome analysis involves comparative genomic studies of strains of the same species-level phylotype: strains are typically initially defined as isolates that share ≥99% identity (ID) in the sequences of their 16S rRNA gene while species bins are defined by isolates sharing ≥97% ID. The aggregate set of genes in all strains of a given species is defined as that organism's pan-genome. Some pan-genomes are more open than others, meaning that some species share a smaller core group of genes that are conserved in all isolates and a larger set of variably represented genes (Medini *et al*. 2005).

A pan-genome study can provide a more comprehensive definition of the niches (professions) occupied by a given species in a community and provide insights about how variations in gene repertoires and functions contributes to the fitness of that species in an ecosystem and to overall ecosystems robustness. Knowledge of gene conservation between strains has potential therapeutic implications: development of strategies for enhancing or reducing the representation of metabolic activities of a given species in a microbiota critically depends upon what functions are invariably represented in that organism's genome.

Microbiota and host energy harvest

A number of recent studies have identified a relationship between the gut microbiota and host energy harvest. Comparisons of germ-free mice, conventionally-raised mice and germ-free mice that have received a gut microbiota from conventionally-raised donors have shown that the presence of a microbiota is associated with a significant increase in host adiposity without an associated increase in food consumption (Backhed *et al*., 2005). This effect appears to reflect a number of mechanisms: (i) increased efficiency of breakdown of otherwise indigestible components of the diet by the microbiota (e.g., the host lacks most of the glycoside hydrolases required to process complex plant glycans whereas

members of the microbiota are richly endowed with these enzymes; Martens *et al*., 2009; Turnbaugh *et al*., 2010); (ii) microbiota-mediated modulation of expression of host genes involved in the metabolism of absorbed energy and their deposition in adipocytes (Backhed *et al*., 2004); and (iii) effects of the microbiota on the immune system (Vijay-Kumar *et al*., 2010).

Transplantation of gut microbial communities from C57Bl/6J mice who are obese because they are homozygous for a null allele in their leptin gene (ob/ob) or C57Bl/6J mice who have consumed a Western diet (high fat) to wild-type C57Bl/6J germ-free recipients produces a greater increase in adiposity than does transplantation of gut communities from wild-type donors or those fed a standard/low fat diet (Turnbaugh *et al*., 2006; 2008). Culture independent, bacterial 16S rRNA-based metagenomic studies of the gut microbiota of wild-type and *ob*/*ob* littermates revealed that the latter contained shifts in the proportional representation of the principal bacterial phyla (the Bacteroidetes were reduced, the Firmicutes and Actinobacteria increased; Ley *et al*., 2005). Furthermore, shotgun sequencing of gut community DNA disclosed that the ob/ob microbiome contained a greater proportional representation of genes involved in the digestion/fermentation of indigestible dietary polysaccharides by human enzymes. Similar differences in microbiome gene content have been observed in lean versus obese human twins (Turnbaugh *et al*., 2009a). Moreover, transplantation of fecal microbiota from MZ or DZ twins who are discordant for obesity into wild type C57Bl/6J germ-free mice has shown that recipients of the obese co-twin's microbiota exhibit greater increases in adiposity than do recipients of the lean co-twins' microbiota (Ridaura *et al*., manuscript in preparation).

These findings raise the possibility that the microbiota could represent a target for therapeutic manipulation in individuals who are obese or who are at risk for obesity. Therapeutic targeting would have to be sufficiently selective so that key microbiota functions that benefit the host are not imperiled. One potential target for selective manipulation is the hydrogen-economy of the gut.

The hydrogen economy of the human gut microbiota

Microbial fermentation of dietary polysaccharides yields short-chain fatty acids (SCFA), organic acids such as formate, and other bacterial end-products, including gasses (hydrogen and carbon dioxide; MacFarlane *et al*., 2003). SCFA can account for up to 10% of daily energy intake in individuals on a Western diet (McNeil *et al*., 1984). If hydrogen is not cleared from the gut ecosystem, fermentation is inhibited: i.e., excess hydrogen inhibits re-oxidation of NADH, which reduces the yield of ATP and SCFA. Intriguingly, a recent study of overweight humans indicated that acetate produced from colonic fermentation also inhibits lipolysis (Ferchaud-Roucher *et al*., 2005). Since clearance of hydrogen from the gut is a key factor in maintaining efficient fermentation, reductions in the representation of hydrogen-consumers or in their metabolic activities would be expected to reduce the efficiency of energy harvest.

The human gut contains three groups of hydrogen-consuming organisms: methanogenic archaea, plus acetogenic and sulfate-reducing bacteria. The only known methane-producing microbes on Earth belong to Archaea. Archaeal diversity is remarkably low in the human gut: the mesophilic methanogenic euryarchaeote, *M. smithii* is the overwhelmingly dominant archaeon in this ecosystem (Eckburg *et al*., 2005). It is possible that manipulating the representation of *M. smithii* and/or other archaeons in our gut microbiota could provide a novel means for preventing obesity or for increasing caloric harvest in undernourished individuals. *M. smithii* is a potential therapeutic target because (i) it is an archaeon and thus possesses features that are fundamentally different from bacteria and it has a lower H_2-utilization threshold than acetogens and thus is likely to be more efficient at depleting H_2 from the gut environment. Studies utilizing methane breath tests to detect the presence of methanogens in the gut have shown that 30-50% of humans are methanogen positive (Bond *et al*., 1971; Levitt *et al*., 2006). A recent study found that compared to obese subjects, levels of methanogens were lower in lean individuals, and in those who had undergone gastric bypass surgery (Zhang *et al*., 2009).

Acetogens consume H_2 and CO_2, or formate, to generate acetate and ATP by means of the Wood-Ljungdahl pathway. One human-gut associated acetogen, *B. hydrogenotrophica*, utilizes aliphatic and aromatic amino acids in gnotobiotic mice colonized with this organism and with a prominent sacchrolytic bacterium, *Bacteroides thetaiotaomicron*. In this system, *B. hydrogenotrophica* consumed reducing equivalents and maintained a higher NAD+/NADH ratio, and boosted acetate production (Rey *et al*., 2010).

Desulfovibrio piger, a member of the delta-Proteobacteria, is the dominant SRB present in the human gut microbiota (Scanlan *et al*., 2009). Sulfate reducers are able to clear hydrogen by generating hydrogen sulfide (H_2S) via anaerobic sulfate respiration. In addition to hydrogen clearance per se, the ability of SRB to produce H_2S may influence energy balance through direct effects on mammalian cell metabolism. For example, H_2S produces features analogous to a hibernation-like state: mice exposed to high levels (80ppm) exhibit a 50% decrease in oxygen consumption within 5 min, and their metabolic rate drops by 90% after 6 h (Blackstone *et al*., 2005).

We do not know whether deliberate reduction in the representation or methanogenic activity of *M. smithii* would be compensated for by increased representation of acetogens, or SRB, and/or by alterations/increases in their metabolic activities. There is some evidence that hydrogen-consumers compete in the gut, with methanogens dominating when present (Pochart *et al.*, 1992, Strocchi *et al.*, 1991). SRB may gain an advantage depending upon the levels of mucin (Gibson *et al.*, 1998), or sulfate (Christl *et al.*, 1992a) that are available, though some workers in the field have not seen those associations (Strocchi *et al.*, 1991). Acetogens may compete more effectively in the absence of methanogens (Dore *et al.*, 1995; Chassard *et al.*, 2008). Methanogens and SRB have also been found to exist together peacefully, with SRB modifying their metabolism to provide hydrogen or formate for methanogens (Raskin *et al.*, 1996, Thiele *et al.*, 1988).

Strategies for targeting hydrogen consumers will likely depend on information gleaned about the representation of methanogens, acetogens and SRB in an in-

dividual prior to treatment, and as noted in Chapter 3 of this thesis could involve class-specific (methanogen, acetogen, sulfate-reducer) or multi-class targeting depending upon knowledge of the initial pre-treatment configuration of the gut microbiota, and knowledge about the responses of all classes to class-specific interventions.

Overview of the dissertation

Given the above considerations, this thesis addresses a number of basic questions. What determines if *M. smithii* is present in a human gut microbiota? What other species in the gut microbial community co-occur with it; how does its genome evolve within an individual and between individuals within and between families.

Chapter 2 describes the full genome sequence of *Methanobrevibacter smithii* type strain, which provides insights into the metabolic adaptations of this organism to the human gut environment, and its interactions with bacterial partners. We compared this genome to other methanogens, including *Methanosphaera stadtmanae*, which is a minor member of the human gut microbiota in some individuals (Eckburg *et al*., 2005). The full 1,853,160-bp genome sequence of the *M. smithii* type strain revealed 1,795 predicted protein-coding genes. *M. smithii* is well equipped to persist in the distal intestine through (i) production of surface glycans resembling those found in the gut mucosa, (ii) a set of novel adhesin-like proteins (ALPs), (iii) consumption of a variety of fermentation products produced by saccharolytic bacteria, and (iv) effective competition for nitrogenous nutrient pools.

Chapter 3 describes experiments using qPCR and bacterial 16S rRNA-directed pyrosequencing plus a variety of computational tools to examine the representation of *M. smithii* and SRB and other taxa in the fecal microbiota of healthy adult female MZ and DZ twin pairs. In addition, I developed methods for isolating *M. smithii* from their frozen fecal samples, sequenced the genomes of 20 isolates recovered from two families, one composed of a MZ twin pair and the other of DZ twin pair, and characterized the isolates' transcriptomes using RNA-Seq during *in vitro* growth in the presence of varying concentrations of

formate, a key fermentation product used as a substrate for methanogenesis. My studies disclosed that carriage and levels of colonization correlate to a significant degree in MZ but not DZ twins providing the first evidence for host genetic factors determining acquisition of a persistent member of the human gut microbiota. Pan-genome analyses defined a core and variable set of genes in *M. smithii*. Based on variations in SNP and gene content/function, strains cluster according to individual host and then family. Adhesin-like proteins (ALPs) likely elements driving the evolutionary adaptation of this organism within and between human gut ecosystems. Finally, RNA-Seq analyses identified strain-specific differences in the transcriptional responses to variations in formate concentrations. These results should help future efforts to develop *M. smithii*-directed strategies for manipulating the hydrogen economy of the gut in ways that promote health.

References

Backhed, F, *et al*. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* **101**:15718-23.

Backhed, F*, et al*. (2005) Host-bacterial mutualism in the human intestine. *Science* 307: 1915-1920.

Blackstone, E*, et al*. (2005) H2S induces a suspended animation-like state in mice. *Science* 308: 518.

Bond, JH Jr, *et al*. (1971) Factors influencing pulmonary methane excretion in man. An indirect method of studying the in situ metabolism of the methane-producing colonic bacteria. *J Exp Med* 133:572-588

Brocks, JJ, *et al*. (1999) Archaean molecular fossils and the early rise of eukaryotes. *Science* **285**:1033-6.

Chassard, C, *et al*. (2008) Assessment of metabolic diversity within the intestinal microbiota from healthy humans using combined molecular and cultural approaches. *FEMS Microbiol Ecol* **66**:496-504.

Claesson, MJ, *et al*. (2010) Microbes and Health Sackler Colloquium: Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci USA* 10.1073/pnas.1000097107.

Costello, EK, *et al*. (2009) Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694-7.

Christl, SU, *et al*. (1992) Cummings, Role of dietary sulphate in the regulation of methanogenesis in the human large intestine. *Gut* **33**: 1234-8.

Dore, J., *et al*. (1995) Most probable number enumeration of H2-utilizing acetogenic bacteria from the digestive tract of animals and man. *FEMS Microbiol Lett* **130**: p. 7-12.

Eckburg, PB, *et al*. (2005) Diversity of the human intestinal microbial flora. *Science* **308**: 1635-8.

Falkowski, PG, *et al*. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**:1034-9.

Ferchaud-Roucher V, *et al*. (2005) Colonic fermentation from lactulose inhibits lipolysis in overweight subjects. *Am J Physiol Endocrinol Metab* 289: E716-720.

Gibson, GR, *et al*. (1998) Dietary modulation of the human gut microflora using prebiotics. *Br J Nutr* **80**: S209-12.

Goodman, AL and JI Gordon (2010) Our unindicted coconspirators: human metabolism from a microbial perspective. *Cell Metab* **12**:111-6.

Hooper LV and AJ Macpherson (2010) Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nature Reviews Immunology* 10, 159-169.

Levitt, MD, *et al*. (2006) Stability of Human Methanogenic Flora Over 35 Years and a Review of Insights Obtained From Breath Methane Measurements. *Clin Gastroenterol Hepatol* 4:123-129.

Ley, RE, *et al*. (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 102: 11070-11075.

Macfarlane, S, *et al*. (2003) Regulation of short-chain fatty acid production. *Proc Nutr Soc* 62: 67-72.

Martens, EC, *et al*. (2009) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* 284: 18445-18457.

McNeil, NI, *et al*. (1984) The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* 39: 338-342.

Medini, D., *et al*. (2007) The microbial pan-genome. *Curr Opin Genet Dev* **15**: 589-94.

Pochart, P, *et al*. (1992) Interrelations between populations of methanogenic archaea and sulfate-reducing bacteria in the human colon. *FEMS Microbiol Lett* **77**: 225-8.

Qin, J*, et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**(7285):59-65.

Raskin, L, *et al*. (1996) Competition and coexistence of sulfate-reducing and methanogenic populations in anaerobic biofilms. *Appl Environ Microbiol* **62**: 3847-57.

Rawls, JF, *et al*. (2006) Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127**: 423-33.

Rey, FE*, et al*. (2010) Dissecting the in vivo metabolic potential of two human gut acetogens. J Biol Chem 285: 22082-22090.

Scanlan, PD*, et al*. (2009) Culture-independent analysis of desulfovibrios in the human distal colon of healthy, colorectal cancer and polypectomized individuals. *FEMS Microbiol Ecol* 69: 213-221.

Stappenbeck, TS, *et al*. (2002) Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* **99**: 15451-5.

Strocchi, A., *et al*. (1991) Competition for hydrogen by human faecal bacteria: evidence for the predominance of methane producing bacteria. *Gut* **32**: 1498-501.

Thiele, JH, *et al*. (1988) Control of Interspecies Electron Flow during Anaerobic Digestion: Role of Floc Formation in Syntrophic Methanogenesis. *Appl Environ Microbiol* **54**:10-19.

Turnbaugh, PJ*, et al*. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444: 1027-1031.

Turnbaugh, PJ, *et al*. (2009a) A core gut microbiome in obese and lean twins. *Nature* **457**: 480-4.

Turnbaugh, PJ, *et al*. (2009b) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**: 6ra14.

Turnbaugh, PJ, *et al*. (2010) Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proc Natl Acad Sci U S A* **107**(16):7503-7508.

Vijay-Kumar, M, *et al*. (2010) Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**: 228-231.

Zhang, H, *et al*. (2009) Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci U S A* **106**:2365-70.

Chapter 2

Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut

Chapter 2

Proc Natl Acad Sci. 2007 June 14;104:10643-10648

Buck S. Samuel[1], Elizabeth E. Hansen[1], Jill K. Manchester[1], Pedro M. Coutinho[3], Bernard Henrissat[3], Robert Fulton[2], Philippe Latreille[2], Kung Kim[2], Richard K. Wilson[1,2], and Jeffrey I. Gordon[1]

[1]Center for Genome Sciences and [2]Genome Sequencing Center, Washington University School of Medicine, St. Louis, MO 63108, and [3]CNRS-UMR6098, Universités Aix-Marseille I & II, Marseille, France

Correspondence to: jgordon@wustl.edu

Introduction

The human gut microbiota is dominated by two divisions of Bacteria, the Bacteroidetes and the Firmicutes, which together encompass >90% of all phylogenetic types (phylotypes). Archaea are also represented, most prominently by a methanogenic Euryarchaeote, *Methanobrevibacter smithii*, which comprises up to 10% of all anaerobes in the colons of healthy adults (1, 2).

Complex dietary polysaccharides (fiber) and proteins are digested by enzymes encoded by genes in the microbial community's collective genome (microbiome) but not in our human genome (3, 4). Bacterial fermentation of polysaccharides yields short chain fatty acids (SCFAs; principally acetate, propionate and butyrate), other organic acids (e.g., formate), alcohols (e.g., methanol, ethanol) and gases [e.g., hydrogen (H_2) and carbon dioxide (CO_2)]. Host absorption of SCFAs provides up to 10% of daily caloric intake, although this value varies depending upon the glycan content of the diet (5). Archaeal methanogenesis improves the efficiency of polysaccharide fermentation in animal gut 'bioreactors' by preventing the buildup of H_2 and other reaction end-products (6).

Several recent observations underscore the importance of delineating the genomic and metabolic underpinnings of *M. smithii*'s contributions to energy balance and, if it is to be a therapeutic target for manipulation, the nature of its adaptations to the gut ecosystem. *First*, comparative metagenomic analysis of the gut microbiomes of genetically obese *ob/ob* mice and their lean *ob*/+ or +/+ littermates revealed that the obese community exhibits increased representation of archaea, division-wide increase in the proportion of Firmicutes relative to Bacteroidetes, and an accompanying enrichment of microbial genes involved in polysaccharide degradation (7). Moreover, transplantation of the gut microbiota from *ob*/*ob* or +/+ donors into lean +/+ germ-free (GF) mouse recipients disclosed that this increased energy harvesting capacity is transmissible: i.e., recipients of an obese donor's gut microbiota gained more body fat than did recipients of a lean donor's microbiota (7). *Sec-*

ond, colonization of adult GF mice with *M. smithii* and/or *Bacteroides thetaiotaomicron*, a prominent sequenced human gut symbiont equipped with a large arsenal of glycoside hydrolases not represented in our human proteome (8), revealed that the methanogen increased the efficiency, and changed the specificity of bacterial digestion of dietary glycans (6). Moreover, co-colonized mice exhibited a significantly greater increase in adiposity compared with mice colonized with either organism alone (6). *Third*, metagenomic studies of the colonic microbiomes of two healthy adults confirmed that *M. smithii* is a prominent component of this community, and that enzymes involved in methanogenesis are well represented (3).

With these observations in mind, we have sequenced the genome of the *M. smithii* type strain PS, compared it to other *M. smithii* strains isolated from the human gut and to other archaeons, and have performed functional genomic and biochemical analyses of its properties in gnotobiotic mice that do or do not harbor *B. thetaiotaomicron*. The results provide insights about *M. smithii*'s niche (profession), its evolved adaptations to its gut habitat, and strategies for identifying targets for development of anti-archaeal agents.

Results and Discussion

The 1,853,160 base pair (bp) genome of the *M. smithii* type strain PS contains 1,795 predicted protein coding genes [**Supporting Information (SI) Tables 1-4** and *Results* in Supporting *Text*), 34 tRNAs, and two rRNA clusters. We compared its proteome with the proteomes of (i) *Methanosphaera stadtmanae,* a methanogenic Euryarchaeote that is a minor and inconsistent member of the human gut microbiota (1), (ii) nine 'non-gut methanogens' recovered from microbial communities in the environment, and (iii) these non-gut methanogens plus an additional 17 sequenced Archaea ('all archaea') **(Supporting Table 5**).

Compared to non-gut methanogens and/or all archaea, *M. smithii* and *M. stadtmanae* are significantly enriched (binomial test, $P < 0.01$) for genes assigned to GO (gene ontology) categories involved in surface variation (e.g., cell wall organization and biogenesis, *see below*), defense (e.g., multi-drug efflux/transport), and processing of bacteria-derived metabolites **(Supporting Tables 6 and 7**).

The *M. smithii* and *M. stadtmanae* genomes exhibit limited global synteny **(Supporting Figure 4**) but share 968 proteins with mutual best BLAST hit E-values $\leq 10^{-20}$ (46% of all *M. smithii* proteins; **Supporting Table 8**). A predicted interaction network of *M. smithii* clusters of orthologous groups (COGs) constructed based on STRING (9) shows that it contains more COGs for persistence, improved metabolic versatility, and machinery for genomic evolution compared to *M. stadtmanae* **(Supporting Figure 5** and **Supporting Table 9**).

Cell surface variation **–** The ability to vary capsular polysaccharide surface structures *in vivo* by altering expression of glycosyltransferases (GTs) is a feature shared among sequenced bacterial species that are prominent in the distal human gut microbiota (4, 10–12). Transmission EM studies of *M. smithii* harvested from gnotobiotic mice after a 14 day colonization revealed that it too has a prominent capsule (**Figure 1A**). The proteomes of both human gut methanogens also contain an arsenal of GTs [26 in *M. smithii* and 31 in *M. stadtmanae*; see **Supporting Table 10** for a complete list organized based on the Carbohydrate Active enZyme (CAZy) classification scheme (http://www.cazy.org; 13). Unlike the sequenced Bacteroidetes, which possess large repertoires of glycoside hydrolases (GH) and carbohydrate esterases (CE) not represented in the human 'glycobiome', neither gut methanogen has any detectable GH or CE family members (**Figure 1B)**. Both *M. smithii* and *M. stadtmanae* dedicate a significantly larger proportion of their 'glycobiome' to GT2 family glycosyltransferases than any of the sequenced non-gut associated methanogens (binomial test; $P < 0.00005$; **Figure 1B**). These GT2 family enzymes have diverse predicted activities, including synthesis of hyaluronan, a component of human glycosaminoglycans in the mucosal layer.

Sialic acids are a family of nine-carbon sugars that are abundantly represented in human mucus- and epithelial cell surface-associated glycans (14). N-acetylneuraminic acid (Neu5Ac) is the predominant type of sialic acid found in our species. Unique among sequenced archaea, *M. smithii* has a cluster of genes (*MSM1535-1540*) that encode all enzymes necessary for *de novo* synthesis of sialic acid from UDP-N-acetylglucosamine **(Supporting Figure 6A**). qRT-PCR assays of RNAs prepared from the cecal contents of 12-week-old gnotobiotic mice that had been colonized for 14d with the archaeon alone, or with *B. thetaiotaomicron* for 14d followed by addition of *M. smithii* for 14d (n = 5-6 mice/treatment group) revealed that this cluster of genes is expressed *in vivo* at equivalent levels in mono- and co-colonized mice **(Supporting Table 11**). Biochemical analysis of extracts prepared from cultured *M. smithii*, plus histochemical staining of the microbe with the sialic-acid specific lectin, *Sambucus nigra* agglutinin (SNA), confirmed the presence of Neu5Ac (**Supporting Figure 6B,C**). Taken together, our findings indicate that *M. smithii* has developed mechanisms to decorate its surface with carbohydrate moieties that mimic those encountered in the glycan landscape of its intestinal habitat.

The genomes of both human gut methanogens encode a novel class of predicted surface proteins that have features similar to bacterial adhesins (48 members in *M. smithii* and 37 in *M. stadtmanae*). A phylogenetic analysis **(Supporting** *Methods*) indicated that each methanogen has a specific clade of these Adhesin-Like Proteins (ALPs; **Supporting Figure 7A**). A subset of *M. smithii* ALPs has homology to pectin esterases (GO:0030599): this GO family, which is significantly enriched in this compared to other Archaea based on the binomial test ($P < 0.0005$; **Supporting Table 6**), is associated with binding of chondroitin, a major component of mucosal glycosaminoglycans. Several other *M. smithii* ALPs have domains predicted to bind other sugar moieties (e.g., galactose-containing glycans; **Supporting Figure 7A**). Both methanogens also have ALPs with peptidase-like domains (see **Supporting Table 12** for a complete list of InterPro domains).

We conducted qRT-PCR assays of cecal RNAs from the mono- and co-colonized gnotobiotic mice described above. The results revealed one 'sugar-binding' ALP (*MSM1305*) that was significantly upregulated in the presence of *B. thetaiotaomicron*, four that were suppressed (including one with a GAG binding domain), and two that exhibited no statistically significant alterations **(Supporting Figure 7B**). Regulated expression of distinct subsets of ALPs may direct this methanogen to specific intestinal microhabitats where close association with saccharolytic bacterial partners could promote establishment and maintenance of syntrophic relationships.

Methanogenic and non-methanogenic removal of bacterial end-products of fermentation – Compared to other sequenced non-gut associated methanogens, *M. smithii* has significant enrichment of genes involved in utilization of CO_2, H_2 and formate for methanogenesis (GO:0015948; **Supporting Table 6**). They include genes that encode proteins involved in synthesis of vitamin cofactors used by enzymes in the methanogenesis pathway [methyl group carriers (F_{430} and corrinoids); riboflavin (precursor for F_{430} biosynthesis); and coenzyme M synthase (involved in the terminal step of methanogenesis)] (see **Supporting Table 7** for a gene list and **Figure 2A** for the metabolic pathways). *M. smithii* also has an intact pathway for molybdopterin biosynthesis to allow for CO_2 utilization **(Supporting Figure 8**). qRT-PCR assays demonstrated that while key central methanogenesis enzymes are constitutively expressed in the presence or absence of *B. thetaiotaomicron* [see *Fwd* (tungsten formylmethanofuran dehydrogenase), *Hmd* (methylene-H_4MPT dehydrogenase) and *Mcr* (methyl-CoM reductase)], ribofuranosylaminobenzene 5'-phosphate (RFA-P)-synthase (*RfaS, MSM0848*), an essential gene involved in methanopterin biosynthesis is significantly upregulated with co-colonization (see **Figure 2A** and **Supporting Table 11** for qRT-PCR results). *M. smithii* also upregulates a formate utilization gene cluster (*FdhCAB, MSM1403-5*) for methanogenic consumption of this *B. thetaiotaomicron*-produced metabolite (6).

Our previous qRT-PCR and mass spectrometry studies revealed that co-colonization increased *B. thetaiotaomicron* acetate production [acetate kinase (*BT3963*) is 9-fold upregulated vs. *B. thetaiotaomicron* mono-associated controls; $P < 0.0005$; n = 4-5 animals/group; ref. 6)]. Although acetate is not converted to methane by *M. smithii* (15), we found that its proteome contains an 'incomplete reductive TCA cycle' that would allow it to assimilate acetate [*Acs* (acetyl-CoA synthetase, *MSM0330*), *Por* (pyruvate:ferredoxin oxidoreductase, *MSM0557-60*), *Pyc* (pyruvate carboxylase, *MSM0765*), *Mdh* (malate dehydrogenase, *MSM1040*), *Fum* (fumarate hydratase, *MSM0447, MSM0563, MSM0769, MSM0929*), *Sdh* (succinate dehydrogenase, *MSM1258*), *Suc* (succinyl-CoA synthetase, *MSM0228, MSM0924*), and *Kor* (2-oxoglutarate synthase, *MSM0925-8*) in **Figure 2A**]. qRT-PCR assays disclosed that co-colonization upregulated two important *M. smithii* genes associated with this pathway that participate in acetate assimilation: *Por* as well as *Cab* (carbonic anhydrase, *MSM0654, MSM1223*), which converts CO_2 to bicarbonate, a substrate for Por **(Figure 2B**).

M. smithii also possesses enzymes that in other methanogens facilitate utilization of two other products of bacterial fermentation, methanol and ethanol (16, 17). qRT-PCR assays showed that co-colonization significantly increased expression of a methanol:cobalamin methyltransferase (*MtaB, MSM0515*), an NADP-dependent alcohol dehydrogenase (*Adh, MSM1381*), and an F_{420}-dependent NADP oxidoreductase (*Fno, MSM0049*) [2.4±0.3, 2.3±0.4 and 3.7±0.4 fold vs. mono-associated controls, respectively; $P < 0.01$; **Figure 2A** for pathway information and **Figure 2C** for qRT-PCR results]. Follow-up biochemical studies confirmed a significant decrease in ethanol levels in the ceca of co-colonized mice [11±2 μmol/g total protein in cecal contents versus 35±6 μmol/g in mice with *B. thetaiotaomicron* alone; n = 5-7/group; $P < 0.05$; **Figure 2D]**. Moreover, levels of ethanol in *M. smithii* mono-associated controls were not significantly different from background levels defined in germ-free controls (n = 5-7 animals/group; $P > 0.05$; data not shown). Expression of *B. thetaiotaomicron*'s alcohol dehydrogenases (*BT4512* and *BT0535*) was not al-

tered by co-colonization (6), indicating that the reduction in cecal ethanol levels observed in co-colonized mice is not due to diminished bacterial production but rather to increased archaeal consumption.

Collectively, these findings indicate that *M. smithii* supports methanogenic and non-methanogenic removal of diverse bacterial end-products of fermentation: this capacity may endow it with greater flexibility to form syntrophic relationships with a broad range of bacterial members of the distal human gut microbiota.

M. smithii utilization of ammonium as a primary nitrogen source – Host metabolism of amino acids by glutaminases associated with the intestinal mucosa (18), or deamination of amino acids during bacterial degradation of dietary proteins yields ammonium (19). The *M. smithii* proteome contains a transporter for ammonium (*AmtB*; *MSM0234*) plus two routes for its assimilation: (i) the ATP-dependent glutamine synthetase–glutamate synthase pathway which has a high affinity for ammonium and thus is advantageous under nitrogen-limited conditions; and (ii) the ATP-independent glutamate dehydrogenase pathway which has a lower affinity for ammonium (20).

Microanalytic biochemical assays revealed that the ratio of glutamine to 2-oxoglutarate was 32-fold lower in the ceca of co-colonized gnotobiotic mice compared to animals colonized with *M. smithii* alone, and 5-fold lower compared to *B. thetaiotaomicron* mono-associated hosts ($P < 0.0001$; n = 5/group; **Figure 2E**). In addition, levels of several polar amino acids were also significantly reduced in mice containing the saccharolytic bacterium and methanogen (n = 5/group; **Figure 2F**), providing additional evidence for a more nitrogen-limited gut environment. qRT-PCR analysis established that many of the key *M. smithii* genes involved in ammonium assimilation are upregulated with co-colonization, particularly those in the high affinity glutamine synthetase-glutamate synthase pathway [*GlnA* (glutamine synthetase, *MSM1418*); *GltA*/*GltB* (two subunits of glutamate synthase, *MSM0027, MSM0368*); **Figure 2A,G**]. GeneChip analysis of the transcriptional responses

of *B. thetaiotaomicron* to co-colonization with *M. smithii* indicated that it also upregulates a high affinity glutamine synthetase (*BT4339*; 2.4-fold vs. *B. thetaiotaomicron* monoassociated mice; n = 4-5 mice/group; $P < 0.001$; ref. 6). This prioritization of ammonium assimilation by *B. thetaiotaomicron* and *M. smithii* is accompanied by a decrease in cecal ammonium levels in co-colonized hosts (11.1±1.3 μmol/g dry weight of cecal contents vs. 14.4±0.6 and 14.3±0.9 in *M. smithii*- and *B. thetaiotaomicron*-monoassociated animals, respectively; n = 5-15/group; $P < 0.05$; **Figure 2H**). Together, these studies indicate that ammonium provides a key source of nitrogen for *M. smithii* when it exists in isolation in the gut of gnotobiotic mice, and that it must compete with *B. thetaiotaomicron* for this nutrient resource.

Considering targets for development of anti-*M. smithii* agents

As noted in the *Introduction*, manipulation of the representation of *M. smithii* in our gut microbiota could provide a novel means for treating obesity. When considering how to manipulate the representation of *M. smithii*, several obvious questions arise: (i) is the targeted *M. smithii* gene/pathway expressed *in vivo* and is its expression affected by the presence of actively fermenting bacteria; and (ii) are the therapeutic targets being considered conserved among different *M. smithii* strains? We addressed these questions in a final set of experiments.

Functional genomics studies in gnotobiotic mice illustrate one way to approach issue (i). For example, inhibitors exist for several *M. smithii* enzymes. A class of N-substituted derivatives of para-aminobenzoic acid (pABA) interfere with methanogenesis by competitively inhibiting ribofuranosylaminobenzene 5'-phosphate synthase (RfaS; *MSM0848*; ref. 21). As noted above, this enzyme, which participates in the first committed step in synthesis of methanopterin, is upregulated with co-colonization (4.6±0.9 fold versus monoassociated controls; $P < 0.01$; **Figure 2A**).

Archaeal membrane lipids, unlike bacterial lipids, contain ether-linkages. A key enzyme in the biosynthesis of archaeal lipids is hydroxymethylglutaryl (HMG)-CoA reductase (*MSM0227*), which catalyzes the formation of mevalonate, a precursor for membrane (isoprenoid) biosynthesis (22). Some HMG-CoA reductase inhibitors (statins) have been reported to inhibit growth of *Methanobrevibacter* species *in vitro* (22). qRT-PCR revealed that *MSM0227* is expressed at high levels *in vivo* in the presence or absence of *B. thetaiotaomicron* ($P > 0.05$; **Supporting Table 11**). Although statins are commonly used for treating hypercholesterolemia in humans, we are not aware of any studies that report a causal association between their consumption and weight loss. However, the efficacy of statins as anti-archaeal agents will depend upon factors such as their activity against the *M. smithii* enzyme, their capacity to enter the archaeal cell, and their concentration and stability within the distal human gut.

To address issue (ii) we designed a custom GeneChip containing probesets directed against 99.1% of *M. smithii*'s 1795 predicted protein-coding genes (see **Supporting Table 13** for details). This GeneChip was used to perform whole genome genotyping of *M. smithii* PS (control) plus three other strains recovered from the feces of healthy humans: F1 (DSMZ 2374), ALI (DSMZ 2375) and B181 (DSMZ 11975). Replicate hybridizations indicated that 100% of the open reading frames (ORFs) represented on the GeneChip were detected in *M. smithii* PS, while 90-94% were detected in the other strains, including the potential drug targets mentioned above **(Supporting Table 2** and **Figure 3**). Approximately 50% of the undetectable ORFs in each strain encode hypothetical proteins. The other undetectable genes are involved in genome evolution [e.g., recombinases, transposases, insertion sequence (IS) elements, and type II restriction modification (R-M) systems], are components of a putative archaeal prophage in strain PS **(Supporting Table 15**), or are related to surface variation, including several ALPs (e.g., *MSM0057* and *MSM1585-7, MSM1590*; **Supporting Figure 7**). Strains F1 and ALI also appear to lack redundant gene clusters encoding subunits of formate dehydrogenase (*MSM1462-3*) and methyl-CoM re-

ductase (*MSM0902-4*) that are found in the PS strain (the latter cluster is also undetectable in strain B181). In addition, the only methanol utilization cluster present in the PS strain (*MSM0515-8*) was not detectable in strain F1 **(Supporting Table 2**).

To further assess the degree of nucleotide sequence divergence among *M. smithii* strains, we compared the sequenced PS type strain to a 78 Mb metagenomic dataset generated from the aggregate fecal microbiome of two healthy humans (3). Their sequenced microbiomes contained 92% of the ORFs in the type strain **(Supporting Table 2**), including the potential drug targets described above. Several R-M system gene clusters (*MSM0157-8, MSM1742-8*), a number of transposases, a DNA repair gene cluster (*MSM0690-95*), and all ORFs in the prophage were not evident in the two microbiomes. Sequence divergence was also observed in 33 of the 48 ALP genes plus two 'surface variation' gene clusters (*MSM1288-1313* and *MSM1590-1616*) that encode 11 glycosyltransferases and 9 proteins involved in pseudomurein cell wall biosynthesis **(Supporting Figure 9**). A redundant methyl-CoM reductase cluster (*MSM0902-5*), an F_{420}-dependent NADP oxidoreductase (*MSM0049*) involved in consumption of bacteria-derived ethanol, and two subunits of the bicarbonate ABC transporter (*MSM0990-1*; carbon utilization) exhibited heterogeneity in the *M. smithii* populations present in the gut microbiota of these two adults **(Supporting Table 2** and **Supporting Figure 9**).

Prospectus

Ongoing comparisons of the genomes of multiple strains of a bacterial 'species' have revealed remarkable variations in gene content. This has lead to the concept that understanding the nature of a microbial 'species' requires that we consider the sum of all genes present in all strains (i.e., its pan-genome; ref. 23). Our results provide an initial glimpse of the genetic diversity of a gut archaeon and the operations of one strain's transcriptome and metabolome *in vivo*. While there is overall conservation of gene content among different strains of *M. smithii*, there are differences in metabolic capacities and

surface properties represented in the organism's pan-genome. These differences may influence its partitioning within, and adaptations to, its gut habitat, as well as its relationships with members of the bacterial community.

Tractable genetic systems for manipulating the *M. smithii* genome are not available at present. However, further characterization of the *M. smithii* pan-genome by sequencing isolates obtained from related and unrelated individuals represents an opportunity to evaluate genome conservation and evolution in an archaeon that co-exists with bacteria in an incredibly dense microbial community. In addition, gnotobiotic mice colonized with different strains of *M. smithii* together with sequenced representatives of human gut-derived Bacteroidetes and Firmicutes, should help guide drug discovery programs aimed at identifying anti-archaeal therapeutic agents with broad efficacy.

Materials and Methods

Genome sequencing and annotation – *Methanobrevibacter smithii* strain PS (ATCC 35061) was grown as described in *Supporting Methods* for 6d at 37°C. DNA was recovered from harvested cell pellets using the QIAGEN Genomic DNA Isolation kit with mutanolysin (1 unit/mg wet weight cell pellet; Sigma) added to facilitate lysis of the microbe. An ABI 3730xl instrument was used for paired end-sequencing of inserts in a plasmid library (average insert size 5 Kb; 42,823 reads; 11.6X-fold coverage), and a fosmid library (average insert size of 40 Kb; 7,913 reads; 0.6X-fold coverage). Phrap and PCAP (24) were used to assemble the reads. A primer-walking approach was used to fill-in sequence gaps. Physical gaps and regions of poor quality (as defined by Consed; ref. (25) were resolved by PCR-based re-sequencing. The assembly's integrity and accuracy was verified by clone constraints. Regions containing insufficient coverage or ambiguous assemblies were resolved by sequencing spanning fosmids. Sequence inversions were identified based on inconsistency of constraints for a fraction of read pairs in those regions. The final assembly consisted of 12.6X sequence coverage with a Phred base quality value ≥40.

Open-reading frames (ORFs) were identified and annotated as described in **Supporting** *Materials and Methods*.

Quantitative RT-PCR analyses – All mouse experiments were performed using protocols approved by the animal studies committee of Washington University. Gnotobiotic male mice belonging to the NMRI inbred strain (n = 5-6/group/experiment) were colonized with either *M. smithii* (14 d) or *B. thetaiotaomicron* (28 d) alone, or first with *B. thetaiotaomicron* for 14 d followed by co-colonization with *M. smithii* for 14 d. All mice were sacrificed at 12 weeks of age. Cecal contents from each mouse were flash frozen, and stored at -80°C. RNA was extracted from an aliquot of the harvested cecal contents (100-300 mg) and used to generate cDNA for qRT-PCR assays (see **Supporting** *Materials and Methods*). qRT-PCR data were normalized to 16S rRNA ($\Delta\Delta C_T$ method) prior to comparing treatment groups. PCR primers are listed in **Supporting Table 14**. All amplicons were 100-150 bp.

Biochemical assays – Perchloric and hydrochloric acid extracts, and alkali extracts of freeze dried cecal contents were prepared, and established pyridine nucleotide-linked microanalytic assays (26) used to measure metabolites (see Supporting *Methods* for details).

Acknowledgements

We thank W.B. Whitman (University of Georgia), F. Rohwer and R. Edwards (SDSU), plus our colleagues at Washington University (J. Amend, L. Angenent, M. Mahowald, R. Ley, E. Martens, J. Sonnenburg, J. McCutcheon, T. Jones, M. Bjursell, M. Giannakis, J. Oh) for many useful suggestions. We also thank L. Fulton, S. Clifton, and J. Xu for their help and guidance with genome sequencing, S. Wagoner, M. Karlsson, and D. O'Donnell for assistance with husbandry of gnotobiotic mice, H. Wynder for transmission EM studies, and the UCSD Glycotechnology Core Facility for sialic acid analysis. This

work was funded by the National Institutes of Health Grants DK30292 and DK70932, and by the W. M. Keck Foundation. B.S.S. is a recipient of an NSF graduate research fellowship (DGE-0202737).

References

1. **Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A.** (2005) Diversity of the human intestinal microbial flora. *Science* 308**,** 1635-1638.

2. **Miller, T. L. & Wolin, M. J.** (1986) Methanogens in human and animal intestinal tracts. *Syst Appl Microbiol* 7**,** 223-229.

3. **Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. & Nelson, K. E.** (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312**,** 1355-1359.

4. **Sonnenburg, J. L., Xu, J., Leip, D. D., Chen, C. H., Westover, B. P., Weatherford, J., Buhler, J. D. & Gordon, J. I.** (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307**,** 1955-1959.

5. **McNeil, N. I.** (1984) The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* 39**,** 338-342.

6. **Samuel, B. S. & Gordon, J. I.** (2006) A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism. *Proc Natl Acad Sci U S A* 103**,** 10011-10016.

7. **Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I.** (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444**,** 1027-1031.

8. **Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. & Gordon, J. I.** (2003) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science* 299**,** 2074-2076.

9. **von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. & Snel, B.** (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 31**,** 258-261.

10. **Sonnenburg, E. D., Sonnenburg, J. L., Manchester, J. K., Hansen, E. E., Chiang, H. C. & Gordon, J. I.** (2006) A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A* 103**,** 8834-8839.

11. **Mazmanian, S. K., Liu, C. H., Tzianabos, A. O. & Kasper, D. L.** (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122**,** 107-118.

12. **Coyne, M. J., Reinap, B., Lee, M. M. & Comstock, L. E.** (2005) Human symbionts use a host-like pathway for surface fucosylation. *Science* 307**,** 1778-1781.

13. **Coutinho, P. M. & Henrissat, B.** (1999) in *Recent Advances in Carbohydrate Bioengineering*, eds. Gilbert, H. J., Davies, G., Henrissat, B. & Svensson, B. (The Royal Society of Chemistry, Cambridge), pp. 3-12.

14. **Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M.** (2004) Diversity of microbial sialic acid metabolism. *Microbiol Mol Biol Rev* 68**,** 132-153.

15. **Miller, T. L., Wolin, M. J., de Macario, E. C. & Macario, A. J.** (1982) Isolation of Methanobrevibacter smithii from human feces. *Appl Environ Microbiol* 43**,** 227-232.

16. **Fricke, W. F., Seedorf, H., Henne, A., Kruer, M., Liesegang, H., Hedderich, R., Gottschalk, G. & Thauer, R. K.** (2006) The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and H2 for methane formation and ATP synthesis. *J Bacteriol* 188**,** 642-658.

17. **Berk, H. & Thauer, R. K.** (1997) Function of coenzyme F420-dependent NADP

reductase in methanogenic archaea containing an NADP-dependent alcohol dehydrogenase. *Arch Microbiol* 168**,** 396-402.

18. **McCauley, R., Kong, S. E., Heel, K. & Hall, J. C.** (1999) The role of glutaminase in the small intestine. *Int J Biochem Cell Biol* 31**,** 405-413.

19. **Wallace, R. J.** (1996) Ruminal microbial metabolism of peptides and amino acids. *J Nutr* 126**,** 1326S-1334S.

20. **Cabello, P., Roldan, M. D. & Moreno-Vivian, C.** (2004) Nitrate reduction and the nitrogen cycle in archaea. *Microbiology* 150**,** 3527-3546.

21. **Dumitru, R., Palencia, H., Schroeder, S. D., DeMontigny, B. A., Takacs, J. M., Rasche, M. E., Miner, J. L. & Ragsdale, S. W.** (2003) Targeting Methanopterin Biosynthesis To Inhibit Methanogenesis. *Appl. Environ. Microbiol.* 69**,** 7236-7241.

22. **Wolin, M. J. & Miller, T. L.** (2006) Control of rumen methanogenesis by inhibiting the growth and activity of methanogens with hydroxymethylglutaryl-SCoA inhibitors. *International Congress Series* 1293**,** 131-137.

23. **Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R.** (2005) The microbial pan-genome. *Curr Opin Genet Dev* 15**,** 589-594.

24. **Huang, X., Wang, J., Aluru, S., Yang, S. P. & Hillier, L.** (2003) PCAP: a whole-genome assembly program. *Genome Res* 13**,** 2164-2170.

25. **Gordon, D., Abajian, C. & Green, P.** (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8**,** 195-202.

26. **Passonneau, J. V. & Lowry, O. H.** (1993) *Enzymatic Analysis: A Practical Guide* (Humana Press, Totawa, NJ).

27. **Prangishvili, D., Garrett, R. A. & Koonin, E. V.** (2006) Evolutionary genomics of archaeal viruses: Unique viral genomes in the third domain of life. *Virus Research* 117**,** 52-67.

28. **Calendar, R. L.** (2005) *The Bacteriophages* (Oxford University Press.

29. Bertani, G. & Baresi, L. (1985) in *EMBO Workshop on Molecular Genetics of Archaebacteria and the International Workshop on Biology and Biochemistry of Archaebacteria*, eds. Kandler, O. & Zillig, W. (Gustav Fischer Verlag, Munich, Germany), pp. 398.

30. **Luo, Y., Pfister, P., Leisinger, T. & Wasserfallen, A.** (2002) Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases produced in Methanothermobacter strains. *FEMS Microbiology Letters* 208**,** 47-51.

31. **Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M.** (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* 34**,** D32-36.

32. **Hamilton, P. T. & Reeve, J. N.** (1985) Structure of genes and an insertion element in the methane producing archaebacterium Methanobrevibacter smithii. *Molecular Genetics and Genomics* 200**,** 47-59.

33. **Brugger, K., Redder, P., She, Q., Confalonieri, F., Zivanovic, Y. & Garrett, R. A.** (2002) Mobile elements in archaeal genomes. *FEMS Microbiol Lett* 206**,** 131-141.

34. **Hylemon, P. B. & Harder, J.** (1998) Biotransformation of monoterpenes, bile acids, and other isoprenoids in anaerobic ecosystems. *FEMS Microbiol Rev* 22**,** 475-488.

35. **Inagaki, T., Moschetta, A., Lee, Y. K., Peng, L., Zhao, G., Downes, M., Yu, R. T., Shelton, J. M., Richardson, J. A., Repa, J. J., *et al*.** (2006) Regulation of antibacterial defense in the small intestine by the nuclear bile acid receptor. *Proc Natl Acad Sci U S A* 103**,** 3920-3925.

36. **Bry, L., Falk, P. G., Midtvedt, T. & Gordon, J. I.** (1996) A model of host-microbial interactions in an open mammalian ecosystem. *Science* 273**,** 1380-1383.

37. **Hooper, L. V., Mills, J. C., Roth, K. A., Stappenbeck, T. S., Wong, M. H. & Gordon, J. I.** (2002) in *Molecular Cellular Microbiology*, eds. Sansonetti, P. & Zychlinsky, A. (Academic, San Diego, CA), Vol. 31, pp. 559-589.

38. **Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L.** (1999) Improved microbial gene identification with GLIMMER. *Nucleic Acids Res* 27**,** 4636-4641.

39. **Badger, J. H. & Olsen, G. J.** (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol Biol Evol* 16**,** 512-524.

40. **Besemer, J., Lomsadze, A. & Borodovsky, M.** (2001) GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29**,** 2607-2618.

41. **Lowe, T. M. & Eddy, S. R.** (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25**,** 955-964.

42. **Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., Bucher, P., Cerutti, L., Corpet, F., Croning, M. D., et al.** (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 29**,** 37-40.

43. **Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., Kiryutin, B., Galperin, M. Y., Fedorova, N. D. & Koonin, E. V.** (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29**,** 22-28.

44. **Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M.** (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32**,** D277-280.

45. **Caspi, R., Foerster, H., Fulcher, C. A., Hopkinson, R., Ingraham, J., Kaipa, P., Krummenacker, M., Paley, S., Pick, J., Rhee, S. Y., et al.** (2006) MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res* 34**,** D511-516.

46. **Fouts, D. E.** (2006) Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res* 34**,** 5839-5851.

47. **Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al.** (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25**,** 25-29.

48. **Benjamini, Y. & Hochberg, Y.** (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* 57**,** 289-300.

49. **Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S. L.** (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5**,** R12.

50. **Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J.** (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21**,** 3422-3423.

51. **Hooper, S. D. & Bork, P.** (2005) Medusa: a simple tool for interaction graph analysis. *Bioinformatics* 21**,** 4432-4433.

52. **Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D.** (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31**,** 3497-3500.

53. **Suyama, M., Torrents, D. & Bork, P.** (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34**,** W609-612.

54. **Stamatakis, A.** (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22**,** 2688-2690.

55. **Page, R. D.** (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12**,** 357-358.

56. **Manzi, A. E.** (1995) in *Current Protocols in Molecular Biology*, ed. Ausubel, F. A. e. a. (Greene Publishing and Wiley Interscience, New York).

57. **Egan, W., Liu, T. Y., Dorow, D., Cohen, J. S., Robbins, J. D., Gotschlich, E. C. & Robbins, J. B.** (1977) Structural studies on the sialic acid polysaccharide antigen of Escherichia coli strain Bos-12. *Biochemistry* 16**,** 3687-3692.

Figure Legends

Figure 1. *M. smithii* decorates its cell surface to mimic the host glycan landscape. (**A**) TEM of *M. smithii* harvested from the ceca of adult GF mice after a 14d colonization. The inset shows a comparable study of stationary phase *M. smithii* recovered from a batch fermentor containing *Methanobrevibacter* complex medium (MBC). Note that the size of the capsule is greater in cells recovered from the cecum (open vs. closed arrow). (**B**) Comparison of glycosyltransferase (GT), glycosylhydrolase (GH) and carbohydrate esterase (CE) families (defined in CAZy; **Supporting Table 10**) represented in the genomes of the following sequenced methanogens (see **Supporting Table 5**): Msm, *Methanobrevibacter smithii*; Msp, *Methanosphaera stadtmanae*; Mth, *Methanothermobacter thermoautotrophicus*; *Methanosarcina acetivorans* (Mac); *Methanosarcina barkeri* (Mba); *Methanosarcina mazei* (Mma); *Methanococcus maripaludis* (Mmr); *Methanococcus jannaschii* (Mja); Mhu, *Methanospirillum hungatei*; Mbu, *Methanococcoides burtonii*; and Mka, *Methanopyrus kandleri*. Gut methanogens (highlighted in orange) have no GH or CE family members, but have a larger proportion of family 2 GTs ($^{\Psi}$, $P < 0.00005$ based on binomial test for enrichment vs. non-gut associated methanogens). Scale bar, 100 nm in panel A.

Figure 2. Functional genomic and biochemical assays of *M. smithii* metabolism in the ceca of gnotobiotic mice. (A) *In silico* metabolic reconstructions of *M. smithii* pathways involved in (i) methanogenesis from formate, H_2/CO_2, and alcohols, (ii) carbon assimilation from acetate and bicarbonate, and (iii) nitrogen assimilation from ammonium. Abbreviations: Acs, acetyl-CoA synthase; Adh, alcohol dehydrogenase; Ags, 18 α-ketoglutarate synthase; AmtB, ammonium transporter; BtcA/B, bicarbonate (HCO_3) ABC transporter; Cab, carbonic anhydrase; CH_3, methyl; CoA, coenzyme A; CoB, coenzyme B; CoM, coenzyme M; COR, corrinoid; F_{420}, cofactor F_{420}; F_{430}, cofactor F_{430}; Fd, ferredoxin (ox-oxidized, red-reduced); FdhAB, formate dehydrogenase subunits; FdhC, formate transporter; Fno, F_{420}-dependent NADP reductase; Ftr, formylmethanofuran:tetrahydromethanopterin (H_4MPT) formyltransferase; Fum, fumarate hydratase; Fwd, tungsten formylmethanofuran dehydro-

genase; GdhA, glutamate dehydrogenase; GlnA, glutamine synthetase; GltA/B, glutamate synthase subunits A and B; Hmd, H_2-forming methylene- H_4MPT dehydrogenase; Kor, 2-oxoglutarate synthase; Mch, methenyl-H_4MPT cyclohydrolase; Mcr, methyl-CoM reductase; Mdh, malate dehydrogenase; MeOH, methanol; Mer, methylene-H_4MPT reductase; MFN, methanofuran; MtaB, methanol:cobalamin methyltransferase; Mtd, F_{420}-dependent methylene-H_4MPT dehydrogenase; Mtr, methyl-H_4MPT:CoM methyltransferase; NH4, ammonium; OA, oxaloacetate; PEP, phosphoenol pyruvate; Por, pyruvate:ferredoxin oxidoreductase; Pps, phosphoenolpyruvate synthase; PRPP, 5-phospho-a-D-ribosyl-1-pyrophosphate; Pyc, pyruvate carboxylase; RfaS, ribofuranosylaminobenzene 5'-phosphate (RFA-P) synthase; Sdh, succinate dehydrogenase; Suc, succinyl-CoA synthetase. Potential drug targets are noted (Rx). (B) Ethanol (EtOH) levels in the ceca of mice colonized with *B. thetaiotaomicron* ± *M. smithii* (n=10-15 animals/group representing 3 independent experiments; each sample assayed in duplicate; mean values ± SEM plotted). (C) Ratio of cecal concentrations of glutamine (Gln) and 2- oxoglutarate (2-OG) (n=5 animals/group; samples assayed in duplicate; mean values ± SEM). (D) Cecal levels of free Gln (glutamine), Glu (glutamate) and Asn (asparagine) (n=5 animals/group; samples assayed in duplicate; mean values ± SEM). (E) Cecal ammonium and urea levels measured in samples used for the assays shown in panels C and D. *, $p<0.05$; **, $p<0.01$; ***, $p<0.005$, according to Student's t-test.

Figure 3. Analysis of the *M. smithii* pan-genome. Schematic depiction of the conservation of *M. smithii* PS genes [depicted in the outermost circle where the color code is orange for forward strand ORFs (F) and blue for reverse strand ORFs (R)] in (i) other *M. smithii* strains (GeneChip-based genotyping of strains F1, ALI, and B181; circles in increasingly lighter shades of green, respectively; see **Supporting Table 2** for details), (ii) the fecal microbiomes of two healthy individuals [human gut microbiome (HGM), shown as the red plot in the fifth innermost circle with nucleotide identity plotted from 80% (closest to the purple circle) to 100% (closest to lightest green ring); see also **Supporting Figure 9** for

details], and (iii) two other members of the Methanobacteriales division [*M. stadtmanae* (Msp; purple circle), another human gut methanogen, and *M. thermoautotrophicus* (Mth; yellow circle), an environmental thermophile; mutual best BLASTP hits (E-value $<10^{-20}$)]. Tick marks indicate nucleotide number in Kb. Asterisks denote the positions of ribosomal rRNA operons. Letters highlight distinguishing features among *M. smithii* genomes: the table below the figure summarizes differences in *M. smithii* gene content between strains F1, ALI, and B181 as well as the two human fecal metagenomic datasets.

Figures

Figure 1.

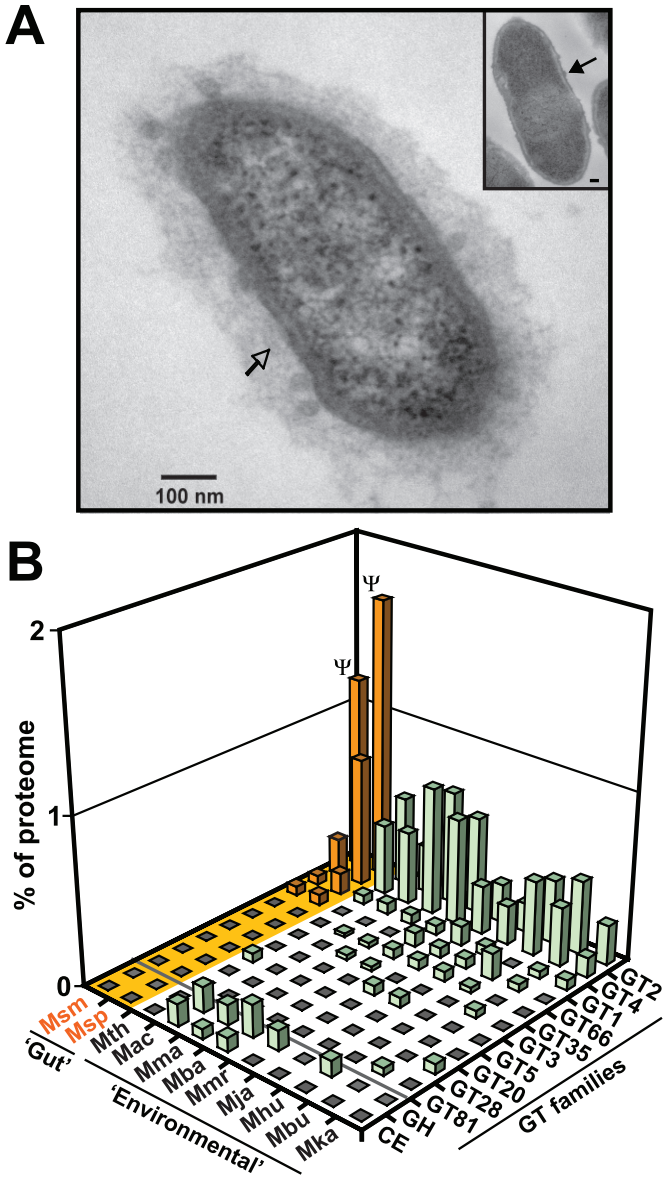


Figure 2.

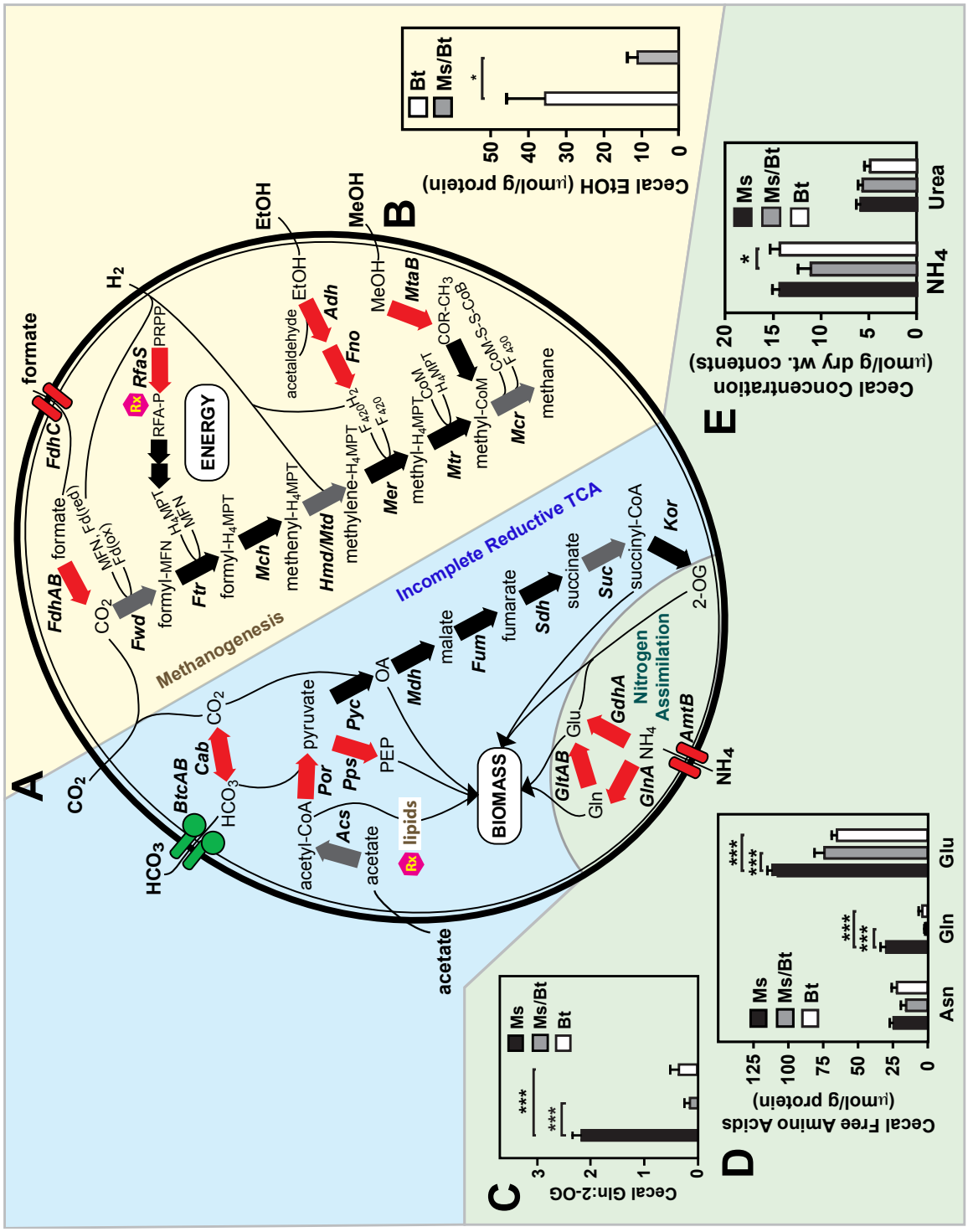
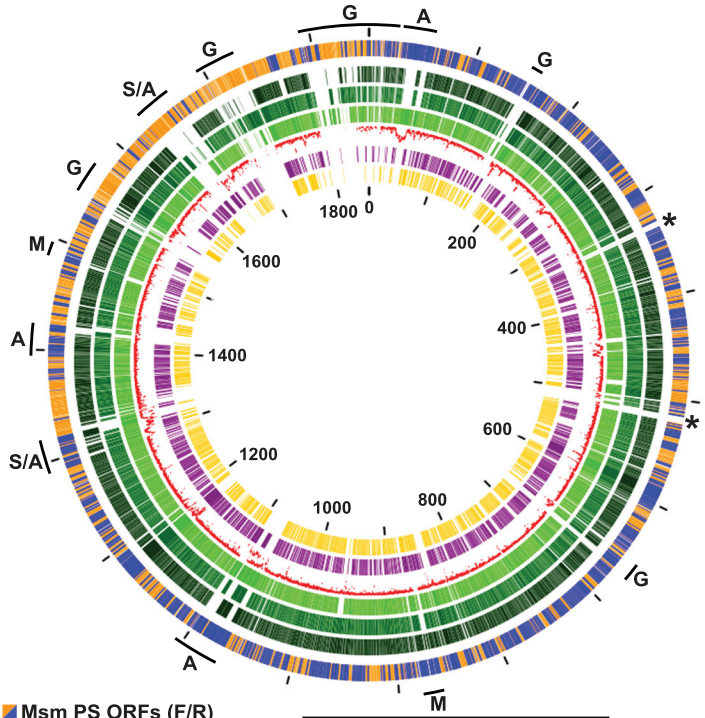

A
B
C
D
E
formate
FdhAB
FdhC
H2
RfaS
PRPP
RFA-P
ENERGY
EtOH
MeOH
Adh
MtaB
acetaldehyde
Fno
COR-CH3
CoM-S-S-CoB
Fdred
Fdox
MFN
H4MPT
Fwd
Ftr
Mch
Hmd/Mtd
Mer
Mtr
Mcr
methane
methyl-CoM
methyl-H4MPT
methylene-H4MPT
methenyl-H4MPT
formyl-H4MPT
formyl-MFN
CO2
HCO3
BtcAB
Cab
Methanogenesis
Incomplete Reductive TCA
pyruvate
Por
Pps
Pyc
Mdh
OA
malate
Fum
fumarate
Sdh
succinate
Suc
succinyl-CoA
Kor
2-OG
acetyl-CoA
Acs
acetate
lipids
BIOMASS
PEP
Glu
Gln
GltAB
GdhA
GlnA
NH4
AmtB
Nitrogen Assimilation
Cecal EtOH (μmol/g protein)
Bt
Ms/Bt
Cecal Concentration (μmol/g dry wt. contents)
Ms
Urea
Cecal Free Amino Acids (μmol/g protein)
Asn
Gln
Glu
Cecal Gln:2-OG

Figure 3.



Msm PS ORFs (F/R)
Msp homologs
Mth homologs

	M. smithii strain genotyping			
	F1	ALI	B181	HGM
PS ORFs detected (%)	91%	90%	94%	92%
Undetected PS ORFs (No.):	160	184	116	143
hypothetical proteins	83	92	50	79
genomic evolution (G)	21	26	19	26
adhesin-like proteins (A)	10	13	12	5
surface variation (S)	7	12	8	5
metabolism (M)	11	6	4	5

Supporting Results

Elements that affect genome evolution – The *M. smithii* PS genome contains multiple elements that can influence genome evolution, including 30 transposases, an integrated prophage (~38 Kb; *MSM1640-92*), eight insertion sequence (IS) elements, 16 genes involved in DNA repair, 9 restriction-modification (R-M) system subunits, and four predicted integrases (see **Supporting Table 4** for a list).

Several lytic phages have been reported to infect *M. smithii*, including a 69 Kb linear phage known as PG that belongs to the ΨM1-like viruses (27), and another 35 Kb phage (PMS11; ref. 28). The PG phage is AT-rich, heavily nicked, and lytic (burst size, 30-90), with a latent period of 3-4 h (ref. 29, plus personal communication from the authors). BLAST comparisons of the 52 predicted genes in the integrated prophage of *M. smithii* PS against known phage genes revealed only a few homologs **(Supporting Table 15**). One of the prophage genes (*MSM1691*) encodes a pseudomurein endoisopeptidase (PeiW): this enzyme may function to cleave *M. smithii*'s cell wall and contribute to autolysis, as related enzymes in a defective *Methanothermobacter wolfeii* prophage have been shown to do (30). The specific ends of the prophage genome could not be identified, and further studies are needed to determine whether the prophage is active and lytic.

The eight insertion sequence (IS) elements in *M. smithii*'s genome **(Supporting Table 4**) range in length from 137 bp (*MSM1519*) to 1013 bp (*MSM0527*) and all are ISM1 (family ISNCY) according to ISfinder (http://www-is.biotoul.fr/; ref. 31). ISM1 is a mobile IS element (32). IS elements promote genome evolution and plasticity through recombination, gene loss and, potentially, lateral gene transfer (33).

Transcriptional regulation – *M. smithii* PS contains 60 predicted transcriptional regulators, including homologs of known nutrient sensors [e.g., a HypF family member (maturation of hydrogenases), a PhoU family member (phosphate metabolism), and a NikR family member (nickel)], plus five regulators of amino acid metabolism **(Support-**

ing Table 3). However, several GO categories related to environmental sensing and regulation (e.g., two-component systems; GO:0000160) are significantly depleted in its proteome compared to the proteomes of methanogens that live in terrestrial or aquatic environments **(Supporting Table 6**). In contrast, *B. thetaiotaomicron,* which uses complex, structurally diversified glycans as its principal nutrient source, possesses a large and diverse arsenal of nutrient sensors including 32 hybrid two-component systems plus 50 ECF-type sigma factors and 25 anti-sigma factors (8, 10). This relative paucity of nutrient sensors may reflect the fact that *M. smithii*'s niche is restricted, and its nutrient substrates are relatively small, readily diffusible molecules that may not require extensive machinery for their recognition.

Bile acid detoxification **–** In humans, cholic and chenodeoxycholic acids are synthesized in the liver and during their enterohepatic circulation undergo transformation by the intestinal microbiota to an array of metabolites (34). Bile acids and their metabolites have microbicidal activity and a genetically engineered deficiency of the bile acid-activated nuclear receptor FXR leads to reduced bile acid pools and bacterial overgrowth (35). Both *M. smithii* and *M. stadtmanae* encode a sodium:bile acid symporter (*MSM1078*), a conjugated bile acid hydrolase (CBAH; *MSM0986*), a short chain dehydrogenase with homology to a 7α-hydroxysteroid dehydrogenase (*MSM0021*). This is consistent with *in vitro* studies of *M. smithii* that demonstrate it is not inhibited by 0.1% deoxycholic acid (15).

Supporting Materials and Methods

Microbes and culturing – All *M. smithii* strains [PS (ATCC 35061), F1 (DSMZ 2374), ALI (DSMZ 2375), and B181 (DSMZ 11975)] were cultivated in 125 ml serum bottles containing 15 ml MBC medium supplemented with 3 g/L formate, 3 g/L acetate, and 0.3 ml of a freshly prepared anaerobic solution of filter-sterilized 2.5% Na_2S (6). The remaining volume in the bottle (headspace) contained a 4:1 mixture of H_2 and CO_2: the headspace was replenished every 1-2 d for a 6 d growth at 37°C.

M. smithii PS was also cultured in a BioFlor-110 batch fermentor with dual 1.5 L fermentation vessels (New Brunswick Scientific). Each vessel contained 750 ml of supplemented MBC medium. One hour prior to inoculation, 7.5 ml of sterile 2.5% Na_2S solution was added to the vessel, followed by 7.5 ml of a serum bottle culture that had been harvested on day 5 of growth. Microbes were then incubated at 37°C under a constant flow of H_2/CO_2 (4:1) (agitation setting, 250 rpm). One milliliter of a sterile solution of 2.5% Na_2S was added daily.

Colonization of germ-free mice with *M. smithii* PS with and without *B. thetaiotaomicron* VPI-5482 – Mice belonging to the NMRI/KI inbred strain (36) were housed in gnotobiotic isolators (37) where they were maintained under a strict 12 h light cycle (lights on at 0600h) and fed a standard, autoclaved, polysaccharide-rich chow diet (B&K Universal, East Yorkshire, UK) *ad libitum*. Each mouse was inoculated at age 8 weeks with a single gavage of 10^8 microbes/strain [*B. thetaiotaomicron* was harvested from an overnight culture in TYG medium (4); *M. smithii* from serum bottles containing MBC medium after a 5 d incubation at 37°C (6)]. For a given experiment, the same preparation of cultured microbes was used for mono-association (single species added) and co-colonization (both species added; see main text for details concerning the length and order of colonization).

Immediately after animals were sacrificed, cecal contents were recovered for preparation of DNA, RNA and biochemical studies (n = 5 mice/treatment group/experiment; n = 3 independent experiments). Colonization density was assessed using a qPCR-based assay employing species-specific primers as described in ref. (6).

Genome Annotation – *M. smithii* genes were identified by comparing outputs from GLIMMER (v.3.01; ref. 38), CRITICA (v.1.05b; ref. 39), and GeneMarkS (v.2.1; ref. 40). WUBLAST (http://blast.wustl.edu/) was then used to identify all ORFs with significant hits to the NR database (as of December 1, 2006). ORFs containing <30 codons and without significant homology (e-value threshold of 10^{-5}) to other proteins were eliminated.

rRNA and tRNA genes were identified using BLASTN and tRNA-Scan (41). Annotation of the predicted proteome of *M. smithii* was completed by using BLAST homology searches against public databases, and domain analysis with Pfam (http://pfam.janelia.org/) and InterProScan (release 12.1; database as of July 16, 2006; ref. 42). Functional classifications were made based on GO terms assigned by InterProScan and homology searches against COGs (43) followed by manual curation. Metabolic pathways were constructed based on KEGG (44) and MetaCyc (http://metacyc.org/; ref. 45). Glycosyltransferases (GT) were categorized according to CAZy (http://www.cazy.org; ref. 13). Putative prophage genes were identified using two independent approaches: (i) BLASTN of predicted *M. smithii* ORFs against a database of all known phage sequences (http://phage.sdsu.edu/phage); and (ii) Hidden Markov Model (HMM)-based analysis using Phage_Finder (46).

Comparative Genomic Analyses

GO term assignments – The number of genes in each archaeal genome that were assigned to each GO term, or to its parents in the GO hierarchy (version available on June 6, 2006; ref. (47)) were totaled. All terms assigned to at least five genes in a given genome were then subjected to statistical tests for overrepresentation, and all terms with a total of five genes across all tested genomes for under-representation, using a binomial comparison reference set (see **Supporting Table 6**). Genes that could not be assigned to a GO category were excluded from the reference sets. A false discovery rate of <0.05 was set for each comparison (48). All tests were implemented using the Math::CDF Perl module (E. Callahan, Environmental Statistics, Fountain City, WI; available at http://www.cpan.org/), and scripts written in Perl.

Percent identity comparisons – The *M. smithii* PS genome sequence was compared to the *M. stadtmanae* genome (16) and a 78 Mb metagenomic dataset of the human fecal microbiome (3) using NUCmer (part of MUMmer v.3.19 package; ref. 49), and a percent identity plot was generated using Mummerplot.

Comparisons of synteny between *M. smithii* and *M. stadtmanae* were completed using the Artemis Comparison Tool (50) set to tBLASTX and the most stringent confidence level.

***M. smithii* interaction network analyses** – All *M. smithii* COGs were submitted to the STRING database (http://string.embl.de/; ref. 9) to create predicted interaction networks (0.95 confidence interval). The program Medusa (51) was then used to organize the networks and color the nodes based on their conservation in *M. smithii*'s proteome (mutual best BLASTP hits with E-values $<10^{-20}$ to the other Methanobacteriales genomes).

Clustering of adhesin-like proteins – *M. smithii* and *M. stadtmanae* ALPs were first aligned using CLUSTALW (v.1.83; ref. 52). To retain the highest level of discrimination between the proteins, the alignment was subsequently converted into a nucleotide alignment using PAL2NAL (53). The resulting alignment was used to create a maximum likelihood tree with RAxML [Randomized accelerated maximum likelihood for high performance computing (v2.2.1; ref. 54) first using the GTR+CAT approximation method for rapid generation of tree topology, followed by the GTR+gamma evolutionary model for determination of likelihood values]. ModelTest (v3.7; http://darwin.uvigo.es/software/modeltest.html) also identified GTR+gamma as the most appropriate evolutionary model for the dataset. Bootstrap values were determined from 100 neighbor-joining trees in Paup (v. 4.0b10; http://paup.csit.fsu.edu/). Tree visualization was completed with TreeView (55).

Functional genomic analysis of *M. smithii* gene expression in gnotobiotic mice - Approximately 100-300 mg aliquots of frozen cecal contents from each gnotobiotic mouse was added to 2 ml tubes containing 250 μl of 212-300 μm-diameter acid-washed glass beads (Sigma), 500 μl of buffer (200 mM NaCl, 20 mM EDTA), 210 μl of 20% SDS, and 500 μl of a mixture of phenol:chloroform:isoamyl alcohol (125:24:1; pH 4.5; Ambion). Samples were lysed using a bead beater (BioSpec; 'high' setting for 5 min at room temperature) and cellular debris was pelleted by centrifugation (10,000 x g at 4˚C for 3 min).

The extraction was repeated by adding another 500 μL of phenol:chloroform:isoamyl alcohol to the aqueous supernatant. RNA was precipitated from the pooled aqueous phases, resuspended in 100 μl nuclease-free water (Ambion), 350 μl Buffer RLT (QIAGEN) was added, and RNA further purified using the RNeasy mini kit (QIAGEN) according to manufacturer's protocols.

Analysis of the sialic acid production by *M. smithii*

Reverse-phase HPLC analysis of cellular extracts – *M. smithii* was cultured in MBC medium, in a batch fermenter, to stationary phase (6 d incubation). Cells were collected by centrifugation, washed three times in PBS, snap frozen in liquid nitrogen, and stored at -80^{o}C. Sialic acid content was assayed using established protocols (56). Briefly, sialic acids were liberated by homogenization of the cell pellet (~30-50 mg wet weight) in 0.5 ml of 2 M acetic acid with subsequent incubation of the homogenate for 3h at 80^{o}C. Samples were filtered through Microcon 10 filters (Millipore) and the filtrate, containing free sialic acid, was dried (speed-vacuum). The released sialic acid was derivatized with DMB (1,2-diamino-4,5-methylene-dioxybenzene) to yield a fluorescent adduct, which was analyzed by C18 reverse phase high-pressure liquid chromatography (RP-HPLC; Dionex DX-600 workstation). Sialic acid was quantified by comparison to known amounts of derivatized standards [N-acetylneuraminic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc)], and blanks (buffer alone).

Histochemical studies – *M. smithii* strains PS and F1 were grown in MBC as above. *B. thetaiotaomicron* VPI-5482, and *Bifidobacterium longum* NCC2705 were grown under anaerobic conditions in TYG medium to stationary phase and used as negative controls. *Escherichia coli* strain K92 (ATCC 35860), which is known to produce sialic acid (57), was incubated in 1419 medium (ATCC) to stationary phase and used as a positive control. All strains were fixed in 1.5 ml conical plastic tubes in either 4% paraformaldehyde or 100% ethanol for at least 8 h at 4^{o}C. Samples were then washed with PBS and

stored at -20°C in 50% ethanol, 20 mM Tris and 0.1% IGEPAL CA-630 [Sigma; prepared in double distilled water (ddH_2O)] until assayed. Samples were diluted in deionized water, placed on coated glass slides (Cel-Line/Erie Scientific Co.), air-dried, dehydrated in graded ethanols (50%, 80%, 100%), treated with blocking buffer (0.3% Triton X-100, 1% BSA in PBS; 30 min at room temperature), and then incubated with 10 μg/ml fluorescein-labeled *Sambucus nigra* lectin (SNA; Vector Laboratories; specificity, Neu5Acα2,6Gal/GalNAc epitopes) for 1 h at room temperature. Slides were subsequently washed with PBS, stained with 4',6-diamidino-2-phenylindole (DAPI, 2 μg/ml; 5 min at room temperature), washed with ddH_2O, and mounted in PBS/glycerol. All solutions were filtered using a 0.22 μm filter (Fisher Scientific). Slides were visualized with an Olympus BX41 microscope and photographed using a Q Imaging QICAM camera and OpenLab software (Improvision, Inc., v.3.1.5).

Transmission electron microscopy (TEM) of *M. smithii* – Cells were harvested at day six of growth in the batch fermentor, and cellular morphology was defined by TEM using methods identical to those described previously for *B. thetaiotaomicron* (4). TEM studies of *M. smithii* present in the ceca of gnotobiotic mice that had been colonized for 14 d with the archaeon were conducted using the same protocol.

Microanalytic biochemical analyses of cecal samples recovered from gnotobiotic mice

Extraction of metabolites from cecal contents – For measurement of ammonium and urea levels, perchloric acid extracts were prepared from 2mg of freeze-dried cecal contents. [Contents were collected with a 10 μl inoculation loop, quick frozen in liquid nitrogen, and lyophilized at – 35°C.] The lyophilized sample was homogenized in 0.2 ml of 0.3 M perchloric acid at 1°C.

For the remaining metabolites, alkali and acid extracts were prepared from 4mg of dried cecal samples that were homogenized in 0.4 ml 0.2 M NaOH at 1°C. For the alkali

extract, an 80 μl aliquot was removed, heated for 20 min at 80°C and then neutralized with 80 μl of 0.25 M HCl and 100 mM Tris base. For the acid extract, a 60 μl aliquot was removed and added to 20 μl 0.7 M HCl, heated for 20 min at 80°C, and then neutralized with 40 μl 100 mM Tris base. Protein content was determined in the alkali extracts using the Bradford method (BioRad).

Metabolite assays – The sample concentrations for ammonium and urea were high enough so that direct fluorometric measurements could be used for detection. However, to measure the low sample concentrations for asparagine, glutamate, glutamine, α–ketoglutarate and ethanol, protocols were adapted from previously established pyridine nucleotide-linked assays, an "oil well" technique, and enzymatic cycling amplification (26). All chemicals and enzymes were from Sigma unless otherwise noted.

Ammonium and Urea: For measurement of ammonium, a 20 μl aliquot of a perchloric acid extract of a given sample of cecal contents was added to 1 ml of a solution containing 50 mM imidazole HCl (pH 7.0), 0.2 mM α–ketoglutarate, 0.5 mM EDTA, 0.02% BSA, 10 μM NADH, and 10 μg/ml beef liver glutamate dehydrogenase (in glycerol; specific activity, 40 units/mg protein). Following a 40 min incubation at 24°C, fluorescence was measured using a Ratio-3 system filter fluorometer (Farrand Optical Components and Instruments, Valhalla, NY**;** excitation at 360 nm; emission at 460 nm). Sample blanks were run that lacked added glutamate dehydrogenase. Ammonium acetate standards were carried throughout all steps.

To measure urea concentrations, 2 μl of a 50 mg/ml solution of Jack bean urease (50 units/mg) was added to the same sample used to determine ammonium levels. Following a 40 min incubation at 24°C, urea levels were defined based on a further reduction in fluorescence. Control sample blanks lacked added urease. Reference urea standards were carried throughout all steps.

Asparagine: A 0.5 μl aliquot of the alkali extract of a given sample of cecal contents was added to 0.5 μl of a solution containing 50 mM Trizma HCl (pH 8.7), 0.04% BSA, and 4 μg/ml *E. coli* asparaginase (160 units/mg protein). Sample blanks lacked added asparaginase. After a 30 min incubation at 24°C, 2 μl of a solution containing 50 mM Trizma HCl (pH 8.1), 10 μM α–ketoglutarate, 10 μM NADH, 4 mM freshly prepared ascorbic acid, 10 μg/ml of pig heart glutamic-oxalacetic transaminase (220 units/mg protein), plus 5 μg/ml beef heart malic dehydrogenase (2800 units/mg protein) was added, and the resulting mixture was incubated for 30 min at 24°C. One microliter of 0.25 M HCl was then introduced. After a 10 min incubation at 24°C, a 2 μl aliquot of the reaction mixture was transferred to 0.1 ml of NAD cycling reagent for 20,000 cycles of amplification and the amplified product measured according to methods described by Passonneau and Lowry (26). Sample blanks lacked added asparaginase. Reference asparagine standards were carried throughout all steps.

Glutamate and Glutamine: A 0.1 μl aliquot from an acid extract of a given sample of cecal contents was added to 0.1 μl of reagent containing 100 mM Na acetate (pH 4.9), 20 mM HCl, 0.4 mM EDTA and 50 μg/ml *E. coli* glutaminase (780 units/mg protein). Another 0.1 μl aliquot of the cecal contents was added to the same reagent in a parallel reaction that lacked added glutaminase (to measure glutamate alone). Following a 60 min incubation at 24°C, 2 μl of a solution containing 50mM Tris acetate (pH 8.5), 0.1mM NAD+, 0.1 mM ADP and 50 μg/ml beef liver glutamate dehydrogenase (120 units/mg protein; Roche) was added to both reaction mixtures, which were subsequently incubated for 30 min at 24°C. The reactions were terminated by addition of 1 μl of 0.2 M NaOH and then heated for 20 min at 80°C. A 2 μl aliquot was subsequently transferred to 0.1 ml NAD cycling reagent and subjected to 20,000 cycles of amplification. Reference glutamine and glutamate standards were carried throughout all steps.

α-*Ketoglutarate:* A 0.5 μl aliquot from an given alkali extract was added to 0.5 μl of reagent containing 100 mM imidazole acetate (pH 6.5), 0.04% BSA, 50 mM ammonium

acetate, 0.2 mM ADP, 4 mM ascorbic acid (freshly prepared), 40 μM NADH and 20 μg/ml beef liver glutamate dehydrogenase (120 units/mg protein; Roche). Following a 30 min incubation at 24°C, the reaction was terminated by adding 0.5 μl of 0.2 M HCl. A 1μl aliquot was transferred to 0.1 ml NAD cycling reagent and subjected to 30,000 cycles of amplification. α–ketoglutarate standards were carried throughout all steps.

Ethanol: A 0.5 μl aliquot of an acid extract from cecal contents was added to 0.5 μl of a solution consisting of 5 mM Tris HCl (pH 8.1), 0.04% BSA, 0.1 mM NAD+, and 20 μg/ml yeast alcohol dehydrogenase (350 units/mg protein). Following a 60 min incubation at 24°C, 1 μl of 0.15 M NaOH was added and the mixture heated for 20 min at 80°C. A 0.5 μl aliquot of this reaction mixture was transferred to 0.1 ml of NAD cycling reagent and amplified 5000-fold. Ethanol standards were carried throughout all steps.

Whole genome genotyping with custom *M. smithii* GeneChips – GeneChips were manufactured by Affymetrix (http://www.affymetrix.com) based on the sequence of the PS strain genome (see **Table S13** for details of the GeneChip design). Duplicate cultures of M. smithii strains PS (ATCC 35061), F1 (DSMZ 2374), ALI (DSMZ 2375) and B181 (DSMZ 11975) were grown in 125 ml serum bottles as described above. Genomic DNA (gDNA) was prepared from each strain using the QIAGEN Genomic DNA Isolation kit; mutanolysin (Sigma; 1 unit/mg wet wt. cell pellet) was added to facilitate lysis of the microbes. gDNA (5-7 µg) was further purified by phenol-chloroform extraction and then sheared by sonication to <200 bp, labeled with biotin (Enzo BioArray Terminal Labeling Kit), denatured at 95˚C for 5 min, and hybridized to replicate GeneChips using standard Affymetrix protocols (http://www.affymetrix.com). M. smithii genes represented on the GeneChip were called ìPresentî or ìAbsentî by DNA-Chip Analyzer v1.3 (dChip; www.biostat.harvard.edu/complab/dchip/) using modeled (PM/MM ratio) data.

Statistical analysis – Pairwise comparisons were made using unpaired Student's t-test. One-way ANOVA, followed by Tukey's *post hoc* multiple comparison test, was used to determine the statistical significance of differences observed between three groups.

Supporting Figure Legends

Figure 4. Analysis of synteny between *M. smithii* and *M. stadtmanae* genomes. (**A**) Dot plot comparison. (**B**) Results obtained with the Artemis Comparison Tool (50) set to tBLASTX and the most stringent confidence level (blue, forward strand; orange, reverse strand). The gut methanogens exhibit limited synteny.

Figure 5. Predicted interaction network of *M. smithii* clusters of orthologous groups (COGs) based on STRING. Individual *M. smithii* COGs are represented by nodes (circles; 622 of the 1352 COGs in *M. smithii*'s genome). Predicted interactions are represented by black lines (0.95 confidence interval; summary of 9,765 total predicted interactions are shown). COG conservation among the Methanobacteriales is denoted by node color: red, *M. smithii* alone; yellow, gut methanogens; green, *M. smithii* and *M. thermoautotrophicus*; and gray, all three genomes. Several clusters are highlighted: (**A**) molybdopterin biosynthesis (methanogenesis from CO_2); (**B**) ion transport; (**C**) DNA repair/recombination; (**D**) antimicrobial transport; (**E**) sialic acid synthesis; (**F**) amino acid transport system; (**G**) HMG-CoA reductase cluster; and (**H**) conserved archaeal membrane protein cluster. See **Supporting Table 9** for lists of genes assigned to COGs.

Figure 6. Sialic acid production by *M. smithii in vitro*. (A) *M. smithii* gene cluster (*MSM1535-40*) encoding enzymes needed to synthesize sialic acid (N-acetylneuraminic acid; Neu5Ac): CapD, polysaccharide biosynthesis protein/sugar epimerase; DegT, pleiotropic regulatory protein/amidotransferase; NeuS, Neu5Ac cytidylyltransferase; NeuA, CMP-Neu5Ac synthetase; NeuB, Neu5Ac synthase; Gpd, glycerol-3-phosphate dehydrogenase. **(B)** Reverse phase-HPLC of derivatized *M. smithii* cell wall extracts. The position of elution of Neu5Ac and N-glycolylneuraminic acid (Neu5Gc) standards are shown. The concentration of Neu5Ac species of sialic acid in *M. smithii* cell walls, when the organism has been cultured in a batch fermentor for 6 d in supplemented MBC medium (does not contain any sialic acid sources; see Supporting *Methods*), is 410±76 pmol/g wet wt. of

cells (n = 3; each sample assayed in duplicate; mean ± SEM). **(C)** Lectin staining with SNA (*Sambucus nigra* agglutinin) shows that *M. smithii* F1 is decorated with Neu5Ac epitopes (counter stained with DAPI; X100 magnification). The specificity of lectin staining was assessed using *E. coli* K92 (positive control; sialic acid-producing), *B. longum* NCC2705 (negative control) and *M. smithii* cells with no lectin added (background autofluorescence control) (data not shown; see Supporting *Methods* for details).

Figure 7. Distinct complements of adhesin-like proteins in gut methanogens. (**A**) A maximum likelihood tree of a CLUSTALW alignment of all adhesin-like proteins (ALPs) in *M. smithii* (47; red branches) and in *M. stadtmanae* (38; black branches). Each methanogen possesses specific clades of ALPs. Branches that are supported by bootstrap values >70% are noted. InterPro-based analysis reveals that many of these proteins contain common adhesin domains [i.e., invasin/intimin domains (IPR008964) and pectate lyase folds (IPR011050)]. They also have domains associated with additional functionality (basis for branch highlighting): (i) sugar binding [e.g., galactose-binding-like (IPR008979) and Concanavalin A-like lectin (IPR013320)]; (ii) glycosaminoglycan (GAG)-binding (IPR012333); or (iii) peptidase activity [e.g., carboxypeptidase regulatory region (IPR008969) and beta-lactamase/transpeptidase-like fold (IPR012338)]; (iv) trans-glycosidase activity [e.g., glycosidase superfamily domains (SSF51445)]; and/or (v) general adhesin/porin activity [e.g., *Bacillus anthracis* OMP repeats/DUF11 (IPR001434)]. See **Supporting Table 12** for a complete list of ALPs and domains identified by InterProScan.

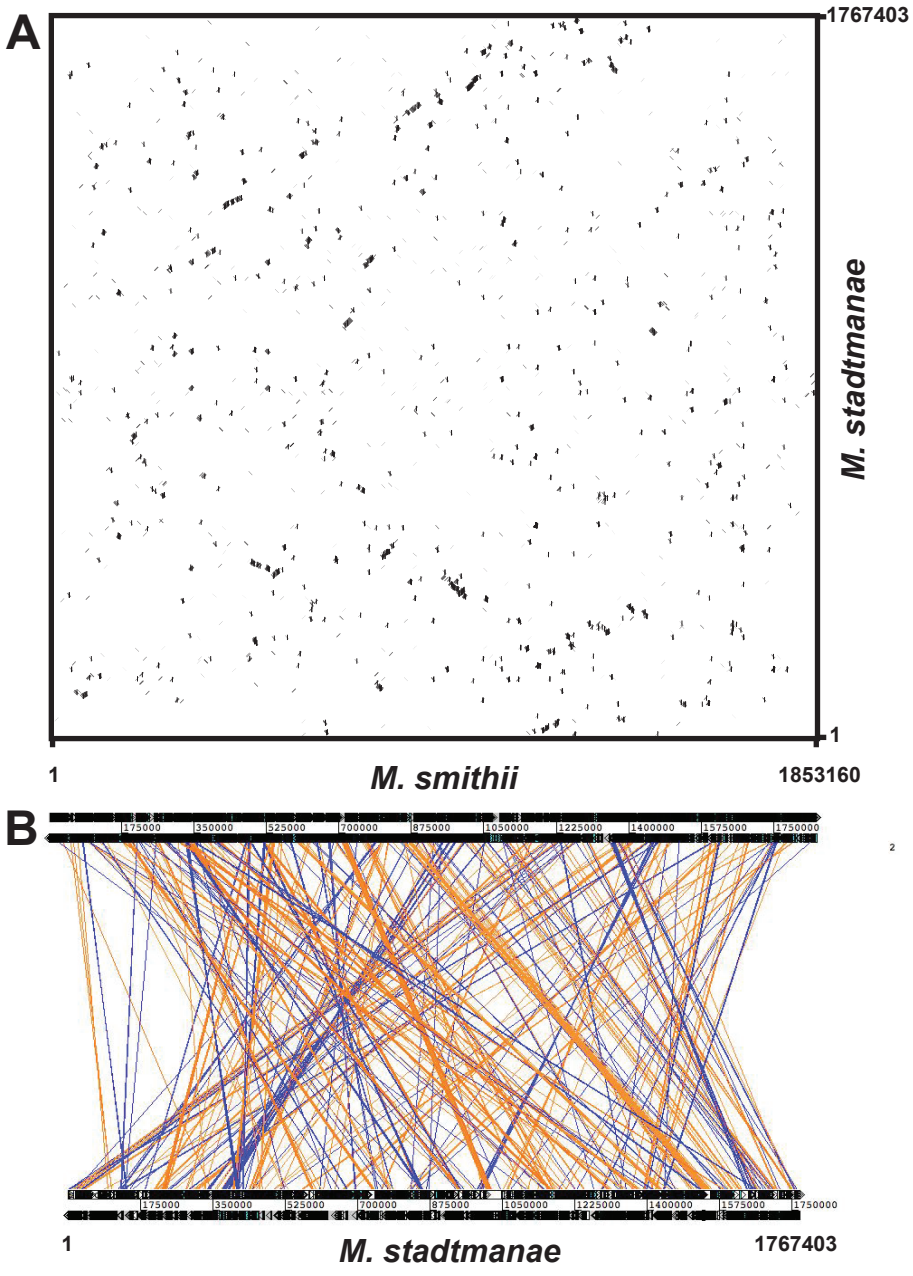
Figure 8. Importance of the molybdopterin biosynthesis pathway for methanogenesis from carbon dioxide in *M. smithii*. (**A**) *In silico* metabolic reconstruction of the predicted molybdopterin biosynthesis pathway encoded by the *M. smithii* genome. Molybdopterin can chelate molybdate (MoO_4^-) or tungstate (WO_4^{2-}) ions. Abbreviations: MoaABCE, molybdenum cofactor biosynthesis proteins A (*MSM0849, MSM1406*), B (*MSM0820*), C (*MSM1362*), and E (*MSM0130*); MoeAB, molybdopterin biosynthesis proteins A (*MSM1343*) and B (*MSM0729*); ModABC, molybdate ABC transport system (*MSM1609-*

11); MobAB, molybdopterin-guanine dinucleotide (MGD) biosynthesis proteins A (*MSM0240*) and B (*MSM1407*); PP, pyrophosphate. Note that the molybdate transporter may also be used for WO_4^{2-}, as no dedicated complex has been identified for its transport. (**B**) Schematic of the first step in the methanogenesis pathway from carbon dioxide (CO_2) catalyzed by tungsten-containing formylmethanofuran dehydrogenase (FwdEFGDBAC; *MSM1408-14*). Essential cofactors for this reaction include tungsten delivered by MGD, methanofuran (MFN), and ferridoxin [Fd; converted from a reduced (red) to oxidized (ox) form during the reaction].

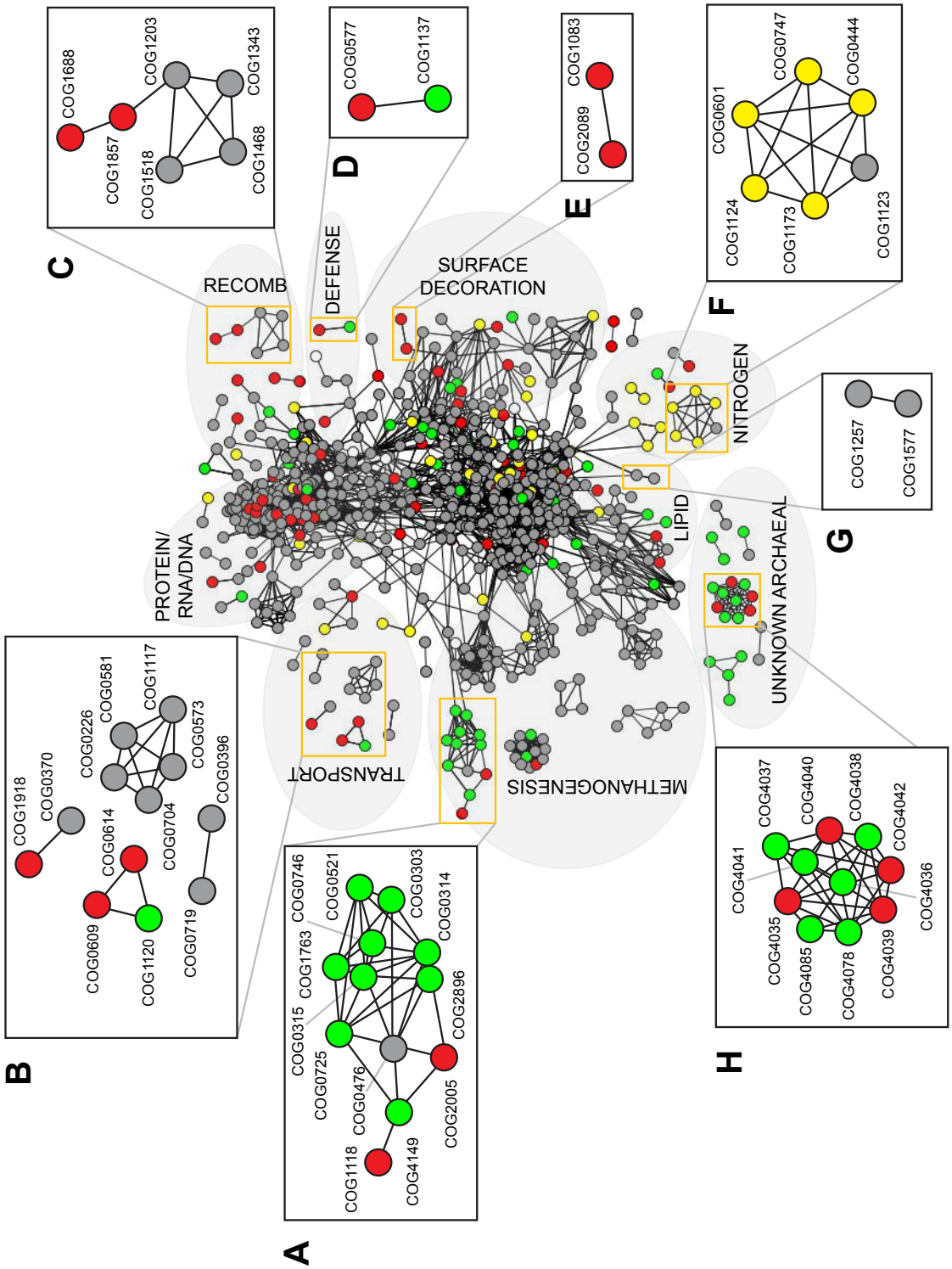
Figure 9. Divergence in genes involved in surface variation, genome evolution, and metabolism among *M. smithii* strains and in the human gut microbiomes of two lean healthy adults. Each of the 139,521 unidirectional reads in the metagenomic dataset (3) were compared to the *M. smithii* PS genome using NUCmer (Supporting *Methods*). Reads with nucleotide sequence identity ≥80% (present) are plotted. A summary of representation of *M. smithii* PS genes present in the metagenomic dataset is displayed at the bottom of the graph (92% of the total ORFs). [Note that the gaps are indications of genome plasticity in the dataset, and include transposases, restriction-modification systems and prophage genes.] Selected regions of heterogeneity (divergence) are highlighted; genes in these regions are involved in the metabolism of bacterial products (Metabolism), recombination/repair machinery (Recomb), anti-microbial resistance (AntiMicrob), surface variation (Surface), and adhesion (ALPs). See **Supporting Table 2** for details.

Supporting Figures

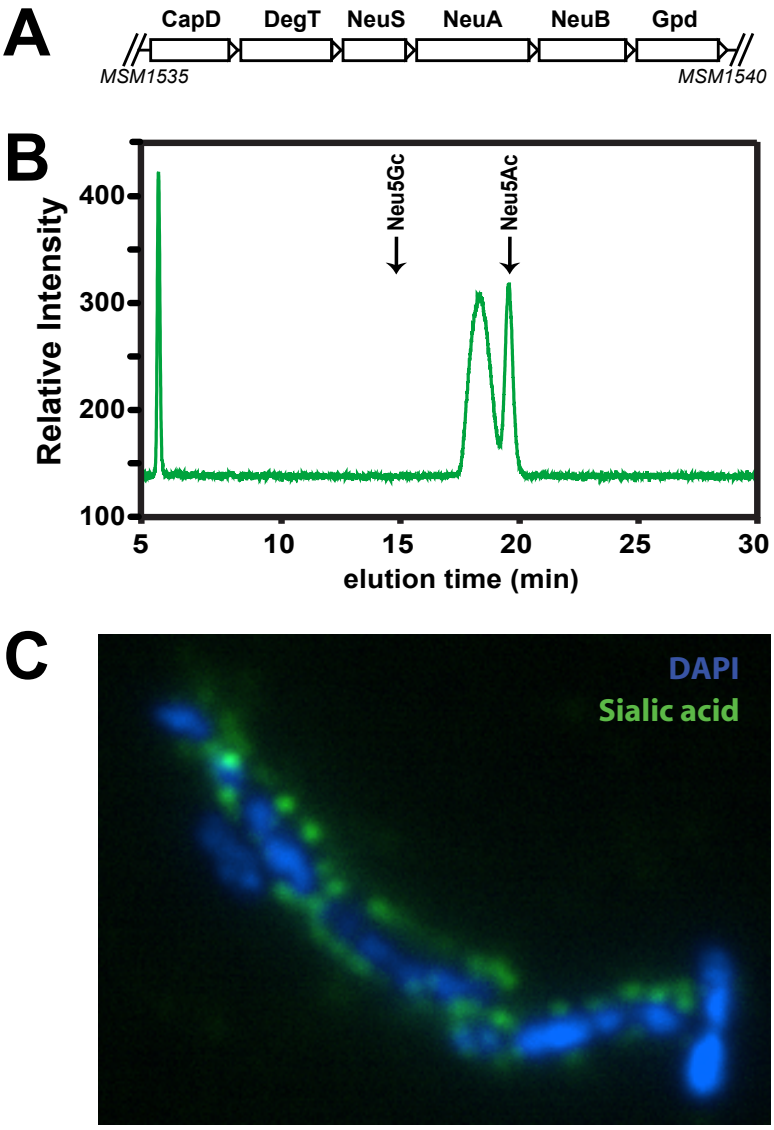
Supporting Figure 4.



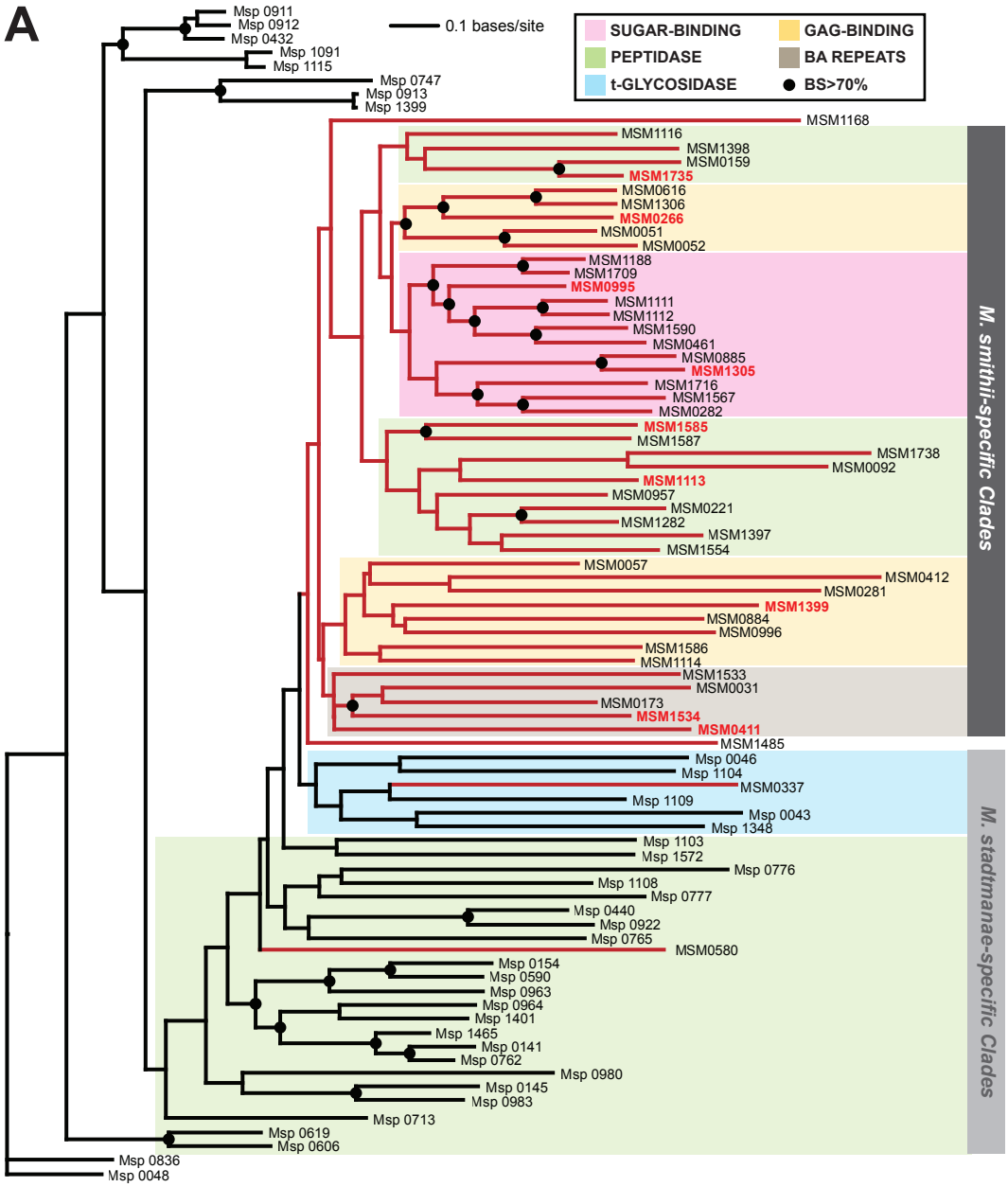
Supporting Figure 5.



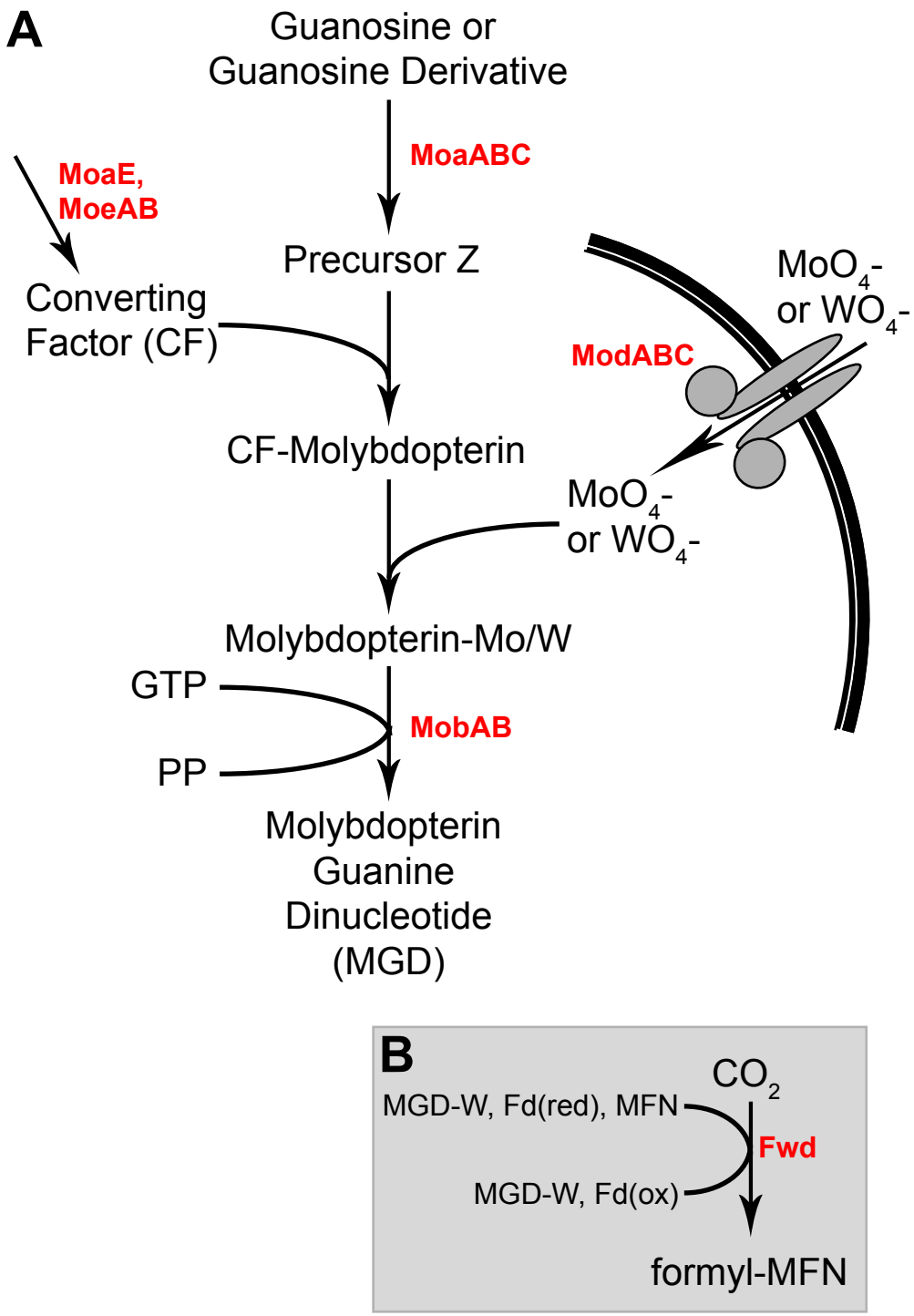
Supporting Figure 6.



Supporting Figure 7.



Supporting Figure 8.



Supporting Figure 9.

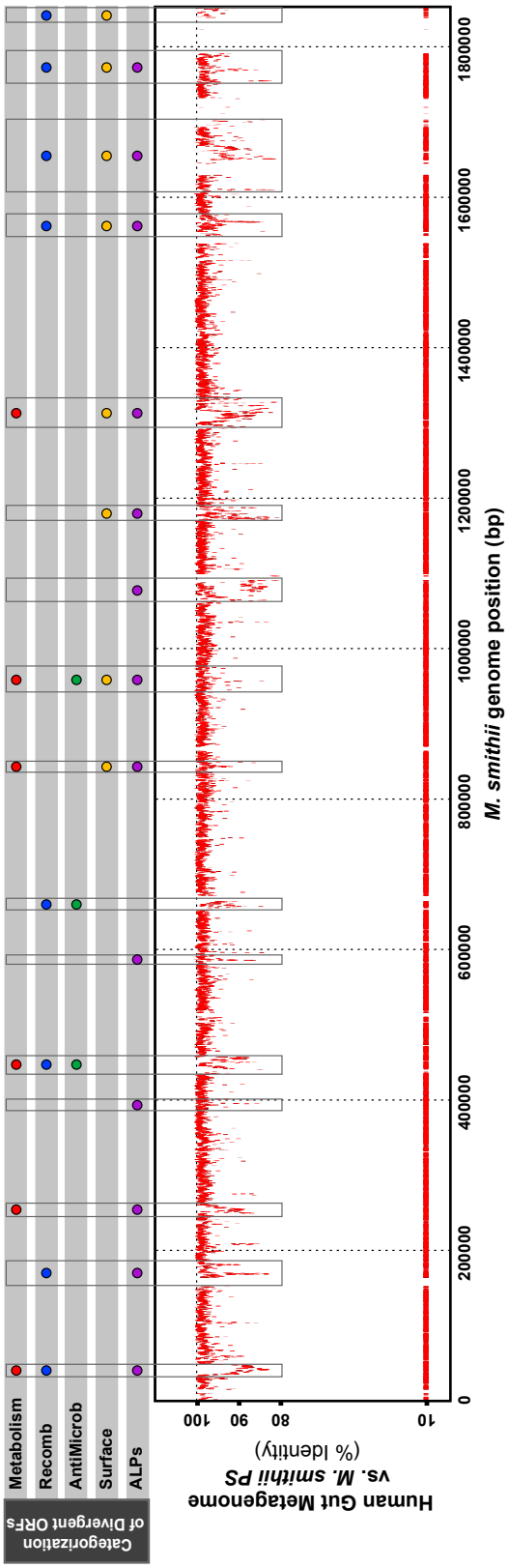


Table 1. General features of the *M. smithii* genome compared to other sequenced Methanobacteriales

	Methanobrevibacter smithii	*Methanosphaera stadtmanae*	*Methanothermobacter thermoautotrophicus*
Genome Size (bp)	1,853,160	1,767,403	1,751,377
G+C content (%)	31	28	50
Coding Regions (%)	90	84	90
Number of ORFs	1795	1534	1869
rRNA operons	2	4	2
tRNA genes	34	40	39
tRNA genes with intron	1	1	3
Transposases (remnants)	2 (20)	1 (2)	0
Insertion Sequences	8	4	0
Restriction Modification System Subunits (Type I/II/III)	2/6/1	3/2/1	3/0/0
Putative Prophage	Yes	No	No

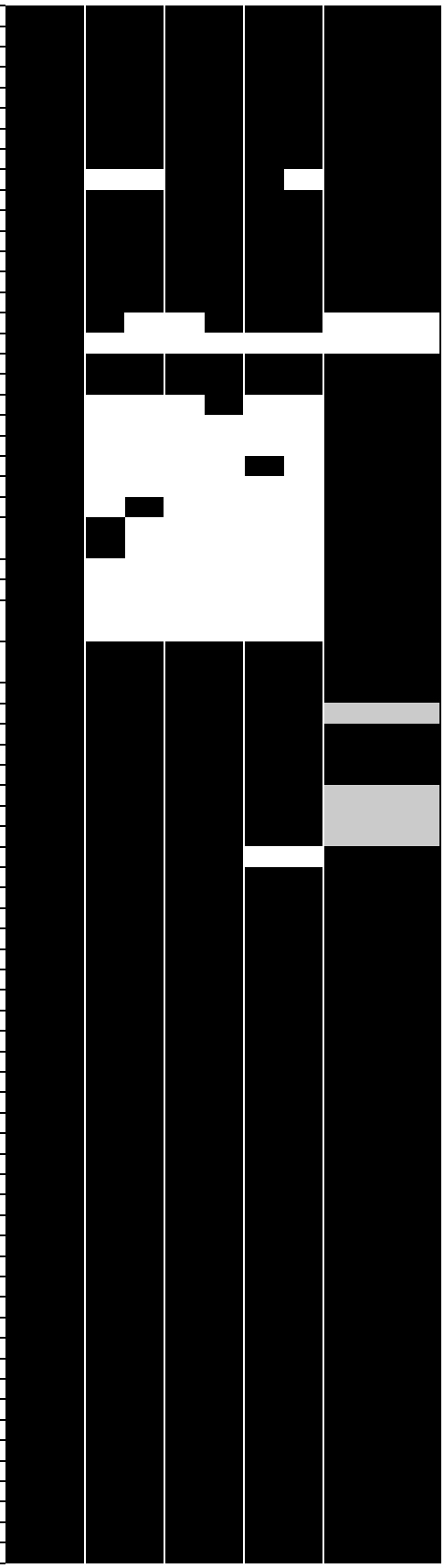
Table 2. Predicted proteome of *M. smithii* strain PS and conservation among other strains and in the fecal microbiome of two healthy adults.

KEY: - Present | - Absent | - Divergent

		***M. smithii* strain genotyping[1]**				**Human Gut Microbiome[2]**
Gene	**Annotation**	**PS**	**F1**	**ALI**	**B181**	
MSM0001	exoribonuclease VII, large subunit, XseA					
MSM0002	integrase-recombinase protein					
MSM0003	conserved hypothetical membrane protein (putative heme utilization/adhesion related)					
MSM0004	predicted lysine decarboxylase					
MSM0005	conserved hypothetical protein					
MSM0006	conserved hypothetical protein					
MSM0007	SAM-dependent methyltransferase					
MSM0008	putative transposase		ND[3]			
MSM0009	conserved hypothetical protein					
MSM0010	N-acetyltransferase, GNAT family					
MSM0011	hypothetical protein					
MSM0012	conserved hypothetical protein					
MSM0013	hypothetical protein					
MSM0014	putative heat shock related protein					
MSM0015	hypothetical protein					
MSM0016	hypothetical protein					
MSM0017	hypothetical protein					
MSM0018	hypothetical protein					
MSM0019	hypothetical protein					
MSM0020	predicted O-linked GlnNAc transferase					
MSM0021	short chain dehydrogenase (7-alpha-hydroxysteroid dehydrogenase)					
MSM0022	hypothetical protein					
MSM0023	uncharacterized protein predicted to be involved in DNA repair					
MSM0024	hypothetical protein					
MSM0025	long-chain-fatty-acid-CoA ligase					
MSM0026	predicted transcriptional regulator					
MSM0027	glutamate synthase, domain 2 with rubredoxin					
MSM0028	SAM-dependent methyltransferase					
MSM0029	putative calcium-binding protein					
MSM0030	conserved hypothetical membrane protein					
MSM0031	adhesin-like protein					
MSM0032	hypothetical protein					
MSM0033	ketopantoate reductase, ApbA					
MSM0034	conserved hypothetical protein					
MSM0035	hypothetical protein					
MSM0036	hypothetical protein					
MSM0037	hypothetical protein					
MSM0038	hypothetical protein					
MSM0039	hypothetical protein					
MSM0040	conserved hypothetical protein					
MSM0041	hypothetical protein					
MSM0042	hypothetical protein					
MSM0043	peptide methionine sulfoxide reductase, PMSR					
MSM0044	PLP dependent aminotransferase (aspartate)					
MSM0045	nucleotide-binding protein (putative ATPase involved in chromosome partitioning)					
MSM0046	NADH oxidase					
MSM0047	Chloramphenicol O-acetyltransferase					
MSM0048	conserved hypothetical protein					
MSM0049	F420-dependent NADP oxidoreductase, fno					
MSM0050	predicted metal-binding protein					
MSM0051	adhesin-like protein					
MSM0052	adhesin-like protein					
MSM0053	tRNA nucleotidyltransferase (CCA-adding enzyme)					
MSM0054	2'-5' RNA ligase, LigT					
MSM0055	predicted alternative 3-dehydroquinate synthase					
MSM0056	archaeal fructose-1,6-bisphosphate aldolase					
MSM0057	adhesin-like protein					
MSM0058	DNA helicase II					
MSM0059	SAM-dependent methyltransferase					
MSM0060	predicted archaeal kinase (GHMP kinase family)					
MSM0061	predicted ATPase (AAA+ superfamily)					
MSM0062	flavodoxin					
MSM0063	amidohydrolase (PHP family)					
MSM0064	conserved hypothetical protein					
MSM0065	riboflavin-specific deaminase					
MSM0066	N-acetylglucosamine-1-phosphate transferase, GT4 family					
MSM0067	conserved hypothetical protein					

MSM0068	hypothetical protein	
MSM0069	conserved hypothetical protein	
MSM0070	conserved hypothetical protein	
MSM0071	methionyl-tRNA synthetase, MetG	
MSM0072	putative exonuclease SBCC	
MSM0073	DNA primase, large subunit (eukaryotic-type)	
MSM0074	hypothetical protein	
MSM0075	DNA primase, small subunit	
MSM0076	conserved hypothetical protein	
MSM0077	thymidylate kinase	
MSM0078	dolichol kinase (cytidylyltransferase family)	
MSM0079	CofH protein (7,8-didemethyl-8-hydroxy-5-deazariboflavin (FO)/F420 biosynthesis)	
MSM0080	sulfopyruvate decarboxylase, comD	
MSM0081	sulfopyruvate decarboxylase, ComE	
MSM0082	heterodisulfide reductase, subunit A, HdrA	
MSM0083	heterodisulfide reductase, subunit B, HdrB	
MSM0084	heterodisulfide reductase, subunit C, HdrC	
MSM0085	putative ferredoxin	
MSM0086	(2R)-phospho-3-sulfolactate synthase, ComA	
MSM0087	putative transposase	ND
MSM0088	conserved hypothetical protein	
MSM0089	pyrroline-5-carboxylate reductase (NADP oxidoreductase, coenzyme F420-dependent), ProC	
MSM0090	conserved hypothetical protein (UPF0058)	
MSM0091	2,3-diphosphoglycerate synthase (putative GTPase)	
MSM0092	putative adhesin-like protein	
MSM0093	conserved hypothetical membrane-spanning protein (phage infection)	
MSM0094	predicted transcription regulator (TetR family)	
MSM0095	predicted phosphotransacetylase	
MSM0096	undecaprenyl pyrophosphate synthase, UppS	
MSM0097	Mg-dependent DNase, TatD	
MSM0098	hypothetical protein	
MSM0099	conserved hypothetical membrane protein	
MSM0100	conserved hypothetical protein	
MSM0101	precorrin-3 methylase, CbiF	
MSM0102	cobalamin-independent methionine synthase, MetE	
MSM0103	conserved hypothetical protein	
MSM0104	conserved hypothetical protein	
MSM0105	conserved hypothetical protein	
MSM0106	conserved hypothetical protein	
MSM0107	hydrogenase expression/formation protein, HypB	
MSM0108	hydrogenase nickel incorporation protein, HypA	
MSM0109	conserved hypothetical membrane-spanning protein	
MSM0110	predicted transposase	
MSM0111	hypothetical protein	
MSM0112	ATP-dependent RNA helicase, eIF-4A family	
MSM0113	DNA helicase	
MSM0114	hypothetical protein	
MSM0115	conserved hypothetical protein	
MSM0116	MobA-related protein	
MSM0117	conserved hypothetical membrane protein	
MSM0118	cell wall biosynthesis protein, MurD-like peptide ligase family	
MSM0119	predicted nuclease	
MSM0120	purine NTPase involved in DNA repair, Rad50	
MSM0121	DNA repair exonuclease (SbcD/Mre11-family), Rad32	
MSM0122	predicted ATPase	
MSM0123	uncharacterized protein conserved in archaea	
MSM0124	predicted phosphate-binding protein (PcrB family)	
MSM0125	ribosomal protein L40e	
MSM0126	conserved hypothetical protein	
MSM0127	hypothetical protein	
MSM0128	conserved hypothetical protein	
MSM0129	nicotinamide mononucleotide adenylyltransferase, NadR	
MSM0130	molybdenum cofactor biosynthesis protein, MoaE	
MSM0131	molybdenum-binding protein, MopI	
MSM0132	conserved hypothetical protein	
MSM0133	predicted thioesterase, FcbC	
MSM0134	M42 glutamyl aminopeptidase/endo-glucanase	
MSM0135	coenzyme F420-reducing hydrogenase, beta subunit	
MSM0136	putative ferredoxin	
MSM0137	putative archaeal flagellar protein D/E	
MSM0138	predicted exonuclease	
MSM0139	hypothetical protein	
MSM0140	conserved hypothetical protein	
MSM0141	dephospho-CoA kinase, CoaE	

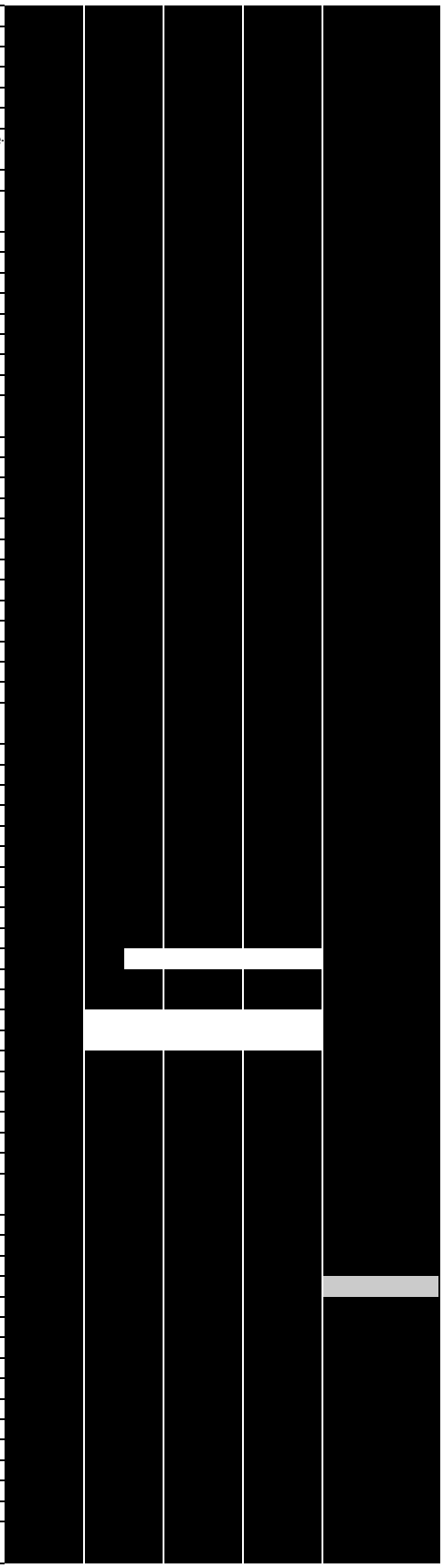
MSM0142	predicted ATPase (PP-loop superfamily)
MSM0143	conserved hypothetical membrane protein
MSM0144	hypothetical protein (putative ADP-ribosylation domain)
MSM0145	conserved hypothetical protein
MSM0146	type IV leader peptidase
MSM0147	CTP synthase (UTP-ammonia lyase), PyrG
MSM0148	predicted oxidoreductase, aldo/keto reductase family
MSM0149	predicted acetylesterase
MSM0150	hypothetical protein
MSM0151	hypothetical protein
MSM0152	Na+-driven multidrug efflux pump (MATE family), NorM
MSM0153	predicted phosphoglycerate mutase
MSM0154	homoserine dehydrogenase, ThrA
MSM0155	predicted allosteric regulator of homoserine dehydrogenase
MSM0156	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase, C subunit
MSM0157	predicted type I restriction-modification enzyme, subunit S
MSM0158	type I restriction-modification system methylase, subunit S
MSM0159	adhesin-like protein
MSM0160	asparagine synthetase, AsnB
MSM0161	hypothetical protein
MSM0162	hypothetical protein
MSM0163	conserved hypothetical proetin predicted to be involved in DNA repair
MSM0164	conserved hypothetical protein predicted to be involved in DNA repair
MSM0165	predicted exonuclease
MSM0166	predicted helicase
MSM0167	conserved hypothetical protein predicted to be involved in DNA repair (RAMP superfamily)
MSM0168	conserved hypothetical protein predicted to be involved in DNA repair
MSM0169	predicted CRISPR-associated protein
MSM0170	conserved hypothetical protein predicted to be involved in DNA repair (RAMP superfamily)
MSM0171	conserved hypothetical membrane protein (invasin/intimin cell-adhesion domain)
MSM0172	hypothetical protein
MSM0173	adhesin-like protein
MSM0174	O-acetylhomoserine sulfhydrylase (PLP-dependent), MET17
MSM0175	homoserine O-acetyltransferase, MetX
MSM0176	ribonuclease III (dsRNA-specific), Rnc
MSM0177	hypothetical protein
MSM0178	conserved hypothetical protein
MSM0179	hypothetical protein
MSM0180	hypothetical protein
MSM0181	ribosomal protein L37e
MSM0182	snRNP Sm-like protein
MSM0183	RNA-binding protein, PUA domain family
MSM0184	creatinine amidohydrolase
MSM0185	conserved hypothetical membrane protein
MSM0186	conserved hypothetical protein
MSM0187	rubredoxin
MSM0188	rubredoxin
MSM0189	acetyl/acyl transferase related protein
MSM0190	predicted ATPase
MSM0191	conserved hypothetical protein
MSM0192	argininosuccinate lyase, ArgH
MSM0193	ribosomal protein S27ae
MSM0194	ribosomal protein S24ae
MSM0195	uncharacterized protein conserved in archaea
MSM0196	archaeal DNA-dependent RNA polymerase, subunit E, RpoE
MSM0197	archaeal DNA-dependent RNA polymerase, subunit E, RpoE
MSM0198	inorganic pyrophosphatase
MSM0199	conserved hypothetical protein (PilT N-term./Vapc superfamily)
MSM0200	translation initiation factor aIF-2, gamma subunit
MSM0201	ribosomal protein S6e
MSM0202	translation initiation factor aIF-2, InfB
MSM0203	nucleoside diphosphate kinase, Ndk
MSM0204	ribosomal protein L24e
MSM0205	ribosomal protein S28e
MSM0206	ribosomal protein L7ae
MSM0207	predicted DNA-binding protein
MSM0208	predicted DNA-binding protein
MSM0209	ferredoxin
MSM0210	hypothetical protein
MSM0211	hypothetical protein
MSM0212	conserved hypothetical protein
MSM0213	archaeal histone, HMtA
MSM0214	threonine synthase (pyridoxal-phosphate dependent), ThrC



MSM0215	conserved hypothetical integral membrane protein	
MSM0216	tryptophanyl-tRNA synthetase, TrpS	
MSM0217	tRNA intron endonuclease, EndA	
MSM0218	iron dependent transcriptional regulator (Fe2+-binding)	
MSM0219	putative cysteine protease (transglutaminase-like superfamily)	
MSM0220	chaperonin (TCP-1/cpn60 family), alpha subunit	
MSM0221	adhesin-like protein	
MSM0222	flavoprotein (Metallo-beta-lactamase superfamily), FpaA	
MSM0223	conserved hypothetical protein	
MSM0224	conserved hypothetical protein	
MSM0225	conserved hypothetical protein	
MSM0226	hypothetical protein	
MSM0227	hydroxymethylglutaryl-CoA (HMG-CoA) reductase, HmgA	
MSM0228	succinyl-CoA synthetase, alpha subunit, SucD	
MSM0229	conserved hypothetical protein	
MSM0230	putative transposase	ND
MSM0231	3-dehydroquinate dehydratase	
MSM0232	signal peptidase I	
MSM0233	nitrogen regulatory protein P-II, GlnK	
MSM0234	ammonium transporter	
MSM0235	hypothetical protein	
MSM0236	phosphohydrolase (HD superfamily)	
MSM0237	3-polyprenyl-4-hydroxybenzoate decarboxylase, UbiX	
MSM0238	precorrin-6B methylase, CbiT	
MSM0239	conserved hypothetical protein	
MSM0240	molybdopterin-guanine dinucleotide biosynthesis protein A, MobA	
MSM0241	ribonuclease PH-related protein	
MSM0242	ribonuclease PH, Rph	
MSM0243	RNA-binding protein Rrp4	
MSM0244	predicted exosome subunit	
MSM0245	proteasome, alpha subunit, PsmA	
MSM0246	ribonuclease P, subunit Rpp14	
MSM0247	ribonuclease P, subunit p30	
MSM0248	hypothetical protein	
MSM0249	conserved hypothetical protein	
MSM0250	conserved hypothetical membrane protein (putative zinc-finger domain, Znf265)	
MSM0251	hypothetical protein	
MSM0252	Na+-driven multidrug efflux pump, NorM	
MSM0253	conserved hypothetical protein	
MSM0254	hypothetical protein	
MSM0255	putative transcription regulator (winged helix DNA-binding domain)	
MSM0256	putative transposase	
MSM0257	conserved hypothetical membrane protein	
MSM0258	hypothetical protein (putative zinc-finger domain, Znf265)	
MSM0259	hypothetical protein (putative zinc beta-ribbon superfamily)	
MSM0260	archaea-specific RecJ-like exonuclease	
MSM0261	conserved hypothetical protein	
MSM0262	desulfoferrodoxin (dfx)	
MSM0263	nitrogen fixation protein, NifU	
MSM0264	cysteine desulfurase, NifS	
MSM0265	O-acetylhomoserine sulfhydrylase	
MSM0266	adhesin-like protein	
MSM0267	NAD(P)H-dependent FMN reductase (multimeric flavodoxin)	
MSM0268	cysteinyl-tRNA synthetase, CysS	
MSM0269	predicted transcriptional regulator (lambda repressor-like)	
MSM0270	serine acetyltransferase, CysE	
MSM0271	cysteine synthase, CysK	
MSM0272	endonuclease III	
MSM0273	EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase)	
MSM0274	SAM-dependent methyltransferase (cyclopropane fatty acid synthase-related)	
MSM0275	valyl-tRNA synthetase, ValS	
MSM0276	conserved hypothetical protein	
MSM0277	phenylalanyl-tRNA synthetase, beta subunit, PheT	
MSM0278	hypothetical protein	
MSM0279	conserved hypothetical protein (UPF0047 family)	
MSM0280	predicted archaeal ATPase (AAA+ superfamily)	
MSM0281	putative adhesin-like protein	
MSM0282	adhesin-like protein	
MSM0283	hypothetical protein	
MSM0284	ribose 5-phosphate isomerase, RpiA	
MSM0285	conserved hypothetical protein (UPF0179 family)	
MSM0286	glycerol 1-phosphate dehydrogenase (Dehydroquinate synthase-like family)	
MSM0287	prolyl-tRNA synthetase, ProS	

MSM0288	conserved hypothetical protein (DUF121 daomain)					
MSM0289	phosphomethylpyrimidine kinase (HMPP-kinase), ThiD					
MSM0290	nitrate/sulfonate/bicarbonate ABC transporter, ATPase component, TauB					
MSM0291	nitrate/sulfonate/bicarbonate ABC transporter, permease component, TauC					
MSM0292	predicted metal-dependent membrane protease					
MSM0293	cation transport ATPase, HAD family					
MSM0294	conserved hypothetical protein					
MSM0295	formate dehydrogenase accessory protein, FdhD					
MSM0296	putative carboxymuconolactone decarboxylase					
MSM0297	predicted exosome subunit					
MSM0298	ribosomal protein L15e					
MSM0299	conserved hypothetical protein					
MSM0300	peptide/nickel ABC transporter, solute-binding component					
MSM0301	peptide/nickel ABC transporter, permease component, DppB					
MSM0302	peptide/nickel ABC transporter, permease component, DppC					
MSM0303	peptide/nickel ABC transporter, ATP-binding component, DppD					
MSM0304	peptide/nickel ABC transporter, ATP-binding component, DppF					
MSM0305	conserved hypothetical membrane protein (IMP dehydrogenase related)					
MSM0306	polyferredoxin, iron-sulfur binding					
MSM0307	sugar kinase (ribokinase/pfkB superfamily)					
MSM0308	formylmethanofuran:tetrahydromethanopterin formyltransferase, FtrC					
MSM0309	conserved hypothetical membrane protein					
MSM0310	polyferredoxin, iron-sulfur binding					
MSM0311	polyferredoxin, iron-sulfur binding					
MSM0312	[NiFe]-hydrogenase-3-type complex, large subunit/NADH:quinone oxidoreductase (complex I), subunit 49K/NdhH/NuoD					
MSM0313	[NiFe]-hydrogenase-3-type complex, small subunit/NADH:quinone oxidoreductase (complex I), subunit PSST/NdhK/NuoB					
MSM0314	conserved hypothetical protein					
MSM0315	predicted [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaL					
MSM0316	hypothetical protein					
MSM0317	NADH dehydrogenase (ubiquinone), subunit 1					
MSM0318	conserved hypothetical membrane protein					
MSM0319	NADH dehydrogenase I, subunit N related					
MSM0320	predicted [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaG					
MSM0321	conserved hypothetical membrane protein					
MSM0322	predicted [NiFe]-hydrogenase-3-type complex Eha, membrane protein EhaE					
MSM0323	conserved hypothetical membrane protein					
MSM0324	conserved hypothetical membrane protein					
MSM0325	conserved hypothetical membrane protein					
MSM0326	conserved hypothetical membrane protein					
MSM0327	UDP-glucose 4-epimerase (NAD dependent)					
MSM0328	conserved hypothetical protein					
MSM0329	DNA binding protein (regulator), xenobiotic response element family					
MSM0330	acetyl-CoA synthetase, AMP-forming-related, Acs					
MSM0331	2-oxoisovalerate ferredoxin oxidoreductase, delta subunit					
MSM0332	2-oxoisovalerate ferredoxin oxidoreductase, alpha subunit					
MSM0333	2-oxoisovalerate ferredoxin oxidoreductase, beta subunit					
MSM0334	L-asparaginase, GatD,					
MSM0335	archaeal glutamyl-tRNA(Gln) amidotransferase, subunit E, GatE					
MSM0336	hypothetical protein					
MSM0337	putative adhesin-like protein					
MSM0338	hypothetical protein					
MSM0339	hypothetical protein					
MSM0340	thioredoxin reductase (NADPH), TrxB					
MSM0341	hypothetical protein					
MSM0342	putative transposase	ND				
MSM0343	GMP synthase (glutamine-hydrolysing), subunit A, GuaA					
MSM0344	hypothetical protein					
MSM0345	GMP synthase (glutamine-hydrolysing), PP-ATPase domain/subunit, GuaA					
MSM0346	conserved hypothetical protein					
MSM0347	putative pyridoxal phosphate-dependent enzyme					
MSM0348	conserved hypothetical protein					
MSM0349	hypothetical protein					
MSM0350	2-isopropylmalate synthase, LeuA					
MSM0351	conserved hypothetical protein					
MSM0352	predicted DNA modification methylase					
MSM0353	conserved hypothetical protein					

MSM0354	ATP-dependent 26S proteasome regulatory subunit, RPT1
MSM0355	predicted transcription factor (eukaryotic MBF1 related)
MSM0356	conserved hypothetical protein
MSM0357	conserved hypothetical membrane protein (possible Zinc-binding)
MSM0358	conserved hypothetical membrane protein
MSM0359	cell wall biosynthesis protein, MurD-like peptide ligase family
MSM0360	cell wall biosynthesis protein, phospho-N-acetylmuramoyl-pentapeptide-transferase family
MSM0361	carbamoyl-phosphate synthase, large subunit, CarB
MSM0362	coenzyme F420-reducing hydrogenase (Ni,Fe-hydrogenase maturation protease), delta subunit
MSM0363	predicted RNA methylase
MSM0364	transcriptional regulator (nickel-responsive), NikR
MSM0365	conserved hypothetical protein
MSM0366	hypothetical protein
MSM0367	conserved hypothetical protein
MSM0368	glutamate synthase (NADPH), subunit 2
MSM0369	glutamate synthase, subunit 3
MSM0370	glutamate synthase, subunit 1
MSM0371	predicted glutamine amidotransferase involved in pyridoxine biosynthesis, Pdx2
MSM0372	phycobiliprotein (PBS) lyase (HEAT repeat)
MSM0373	isocitrate/isopropylmalate dehydrogenase, LeuB
MSM0374	long-chain fatty-acid-CoA ligase (AMP-forming), CaiC
MSM0375	acetylglutamate kinase, ArgB
MSM0376	alcohol dehydrogenase (zinc-binding), GroES-like
MSM0377	4-diphosphocytidyl-2-methyl-D-erithritol synthase, IspD
MSM0378	SAM-dependent methyltransferase
MSM0379	glutamate N-acetyltransferase, ArgJ
MSM0380	hypothetical protein
MSM0381	conserved hypothetical protein
MSM0382	conserved hypothetical protein (PIN domain-like)
MSM0383	predicted phosphohydrolase, calcineurin-like superfamily
MSM0384	transcription factor, NACalpha-BTF3 related
MSM0385	anaerobic magnesium-protoporphyrin IX monomethyl ester cyclase, Elongator protein 3/MiaB/NifB family
MSM0386	sodium/proline symporter (proline permease), PutP
MSM0387	coenzyme F390 synthetase, PaaK
MSM0388	amino acid regulator
MSM0389	hypothetical protein
MSM0390	hypothetical protein
MSM0391	indolepyruvate ferredoxin oxidoreductase, beta subunit
MSM0392	indolepyruvate ferredoxin oxidoreductase, alpha subunit
MSM0393	fumarate reductase, iron-sulfur protein
MSM0394	rRNA methylase, SpoU family
MSM0395	ferredoxin, iron-sulfur binding
MSM0396	putative transposase
MSM0397	xanthine/uracil permease, UraA
MSM0398	uracil phosphoribosyltransferase, Upp
MSM0399	hypothetical protein
MSM0400	hypothetical protein
MSM0401	predicted surface protease
MSM0402	dCTP deaminase, dUTPase family
MSM0403	glycyl-tRNA synthetase
MSM0404	predicted transcriptional regulator
MSM0405	predicted metal-dependent DNase, TatD-related family
MSM0406	conserved hypothetical protein
MSM0407	P-loop containing nucleoside triphosphate hydrolase (NAD(P)-binding)
MSM0408	2-phosphoglycerate kinase/small-molecule binding protein
MSM0409	C4-type Zinc-finger protein
MSM0410	conserved hypothetical protein, histone-fold superfamily
MSM0411	adhesin-like protein
MSM0412	putative adhesin-like protein
MSM0413	transcriptional regulator, MarR family
MSM0414	Na+-driven multidrug efflux pump, NorM
MSM0415	uridylate kinase, PyrH
MSM0416	Mg-dependent DNase, TatD-related
MSM0417	predicted transmembrane protein with a zinc ribbon
MSM0418	conserved hypothetical protein
MSM0419	conserved hypothetical protein
MSM0420	predicted permease
MSM0421	hypothetical protein
MSM0422	conserved hypothetical membrane protein
MSM0423	glycosyltransferase (modular protein with two domains distantly related to glycosyltransferases), GT2/GT1 families [CAZy]

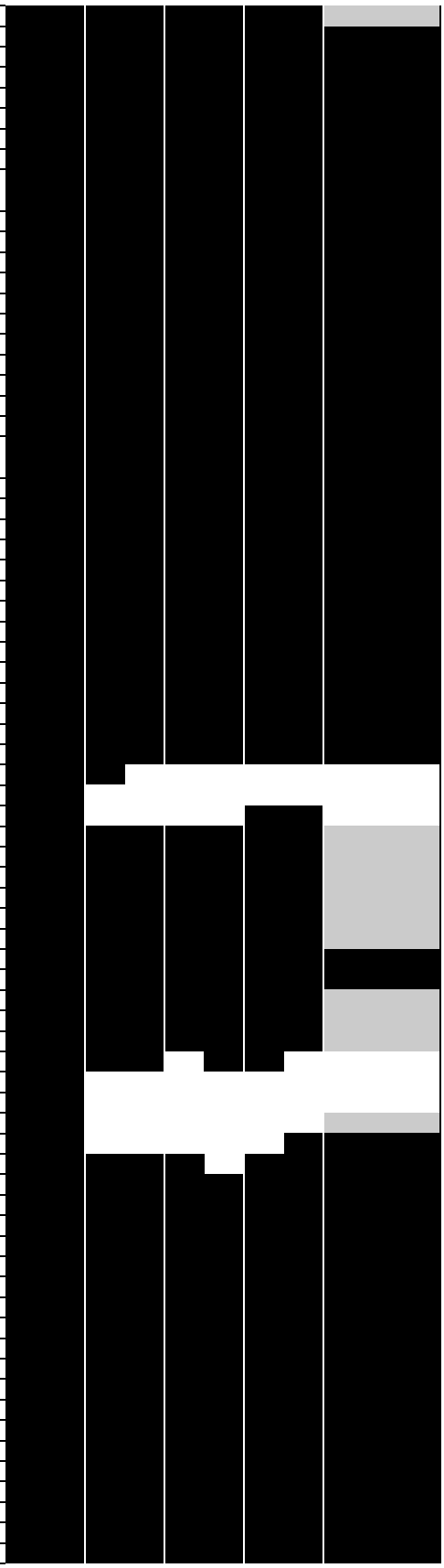


MSM0424	transcription initiation factor TFIIB (zinc-binding)	
MSM0425	predicted RNA-binding protein involved in rRNA processing	
MSM0426	demethylmenaquinone methyltransferase	
MSM0427	DNA primase (bacterial type), DnaG	
MSM0428	integrase-recombinase protein, phage integrase family	
MSM0429	biotin biosynthesis protein, BioY	
MSM0430	conserved hypothetical protein, predicted metal-binding	
MSM0431	predicted ATP-dependent carboligase, biotin carboxylase-related	
MSM0432	conserved hypothetical protein	
MSM0433	archaeal/vacuolar-type H+-transporting ATP synthase, subunit D	
MSM0434	archaeal/vacuolar-type H+-transporting ATP synthase, subunit B	
MSM0435	archaeal/vacuolar-type H+-transporting ATP synthase, subunit A	
MSM0436	archaeal/vacuolar-type H+-transporting ATP synthase, subunit F	
MSM0437	archaeal/vacuolar-type H+-transporting ATP synthase, subunit C	
MSM0438	archaeal/vacuolar-type H+-transporting ATP synthase, subunit E	
MSM0439	archaeal/vacuolar-type H+-transporting ATP synthase, subunit K	
MSM0440	archaeal/vacuolar-type H+-transporting ATP synthase, subunit I	
MSM0441	archaeal/vacuolar-type H+-transporting ATP synthase, subunit H	
MSM0442	hypothetical protein	
MSM0443	hypothetical protein	
MSM0444	hypothetical protein	
MSM0445	NADH dehydrogenase/NAD(P)H nitroreductase	
MSM0446	citrate synthase, GltA	
MSM0447	fumarate hydratase, alpha subunit	
MSM0448	conserved hypothetical protein	
MSM0449	2-methylcitrate dehydratase, MmgE/PrpD family	
MSM0450	conserved hypothetical membrane protein	
MSM0451	conserved hypothetical membrane protein	
MSM0452	predicted DNA-binding protein	
MSM0453	predicted transcriptional regulator	
MSM0454	hypothetical protein	
MSM0455	conserved hypothetical protein	
MSM0456	conserved hypothetical protein	
MSM0457	D-3-phosphoglycerate dehydrogenase, SerA	
MSM0458	transposase, homeodomain-like superfamily	ND
MSM0459	hypothetical protein	
MSM0460	predicted transposase	
MSM0461	adhesin-like protein	
MSM0462	predicted metal-dependent protease, PAD1/JAB1 superfamily	
MSM0463	predicted tRNA(His) guanylyltransferase	
MSM0464	homoserine/aspartate dehydrogenase (NAD binding), glyceraldehyde-3-phosphate dehydrogenase-like superfamily	
MSM0465	conserved hypothetical protein	
MSM0466	predicted tRNA-binding protein	
MSM0467	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	
MSM0468	conserved hypothetical membrane protein	
MSM0469	conserved hypothetical membrane protein	
MSM0470	conserved hypothetical membrane protein	
MSM0471	type II secretion system protein F, GspF	
MSM0472	Xaa-Pro aminopeptidase	
MSM0473	hypothetical protein	
MSM0474	hypothetical protein	
MSM0475	hypothetical protein	
MSM0476	hypothetical protein	
MSM0477	hypothetical protein	
MSM0478	hypothetical protein	
MSM0479	Zn-dependent protease, peptidase M50 family	
MSM0480	YcaO-like protein	
MSM0481	TfuA-like protein	
MSM0482	ATP-utilizing enzymes, PP-loop superfamily	
MSM0483	conserved hypothetical protein	
MSM0484	inosine-5'-monophosphate dehydrogenase related protein	
MSM0485	universal stress protein, UspA	
MSM0486	N-ethylammeline chlorohydrolase, metallo-dependent amidohydrolase family	
MSM0487	hypothetical protein	
MSM0488	carbamoylphosphate synthase, large subunit, CarB	
MSM0489	carbamoylphosphate synthase, small subunit, CarA	
MSM0490	SAM-dependent methyltransferase, UbiE/CobQ family	
MSM0491	nicotinate-nucleotide pyrophosphorylase (carboxylating), NadC	
MSM0492	ribonuclease Z (zinc-dependent), beta-lactamase superfamily, ElaC	
MSM0493	mechanosensitive ion channel protein, Sm-like ribonucleoprotein superfamily, MscS	
MSM0494	quinolinate synthetase, subunit A, NadA	
MSM0495	conserved hypothetical protein	
MSM0496	homoserine O-acetyltransferase	

MSM0497	predicted nuclease, RecB family	
MSM0498	hypothetical protein	
MSM0499	conserved hypothetical protein	
MSM0500	N-carbamoyl-D-amino acid amidohydrolase	
MSM0501	phycocyanin alpha phycocyanobilin lyase, CpcE	
MSM0502	ATP-dependent helicase, Lhr-like	
MSM0503	flavodoxin, FldA	
MSM0504	conserved hypothetical protein	
MSM0505	hypothetical protein	
MSM0506	ATP-utilizing enzyme, ATP-grasp superfamily	
MSM0507	predicted phosphoesterase, YfcE	
MSM0508	cell division protein J (23S rRNA methylase), FtsJ	
MSM0509	conserved hypothetical protein	
MSM0510	predicted ATPase involved in DNA replication control, MCM2/3/5 family	
MSM0511	translation initiation factor 2, beta subunit (aIF-2beta)	
MSM0512	NMD3-related protein (nonsense mediated mRNA decay)	
MSM0513	tyrosyl-tRNA synthetase, TyrS	
MSM0514	conserved hypothetical protein	
MSM0515	methanol:cobalamin methyltransferase, MtaB	
MSM0516	corrinoid protein (methionine synthase-related), MtaC	
MSM0517	methyltransferase activation protein, MapA	
MSM0518	methylcobalamin:coenzyme M methyltransferase, MtaA	
MSM0519	conserved hypothetical protein	
MSM0520	thymidylate kinase, Tmk	
MSM0521	conserved hypothetical membrane protein	
MSM0522	collagenase, peptidase family U32	
MSM0523	collagenase, peptidase family U32	
MSM0524	DNA mismatch repair ATPase, MutS	
MSM0525	predicted unusual protein kinase, ubiquinone biosynthesis protein-related, AarF	
MSM0526	conserved hypothetical protein	
MSM0527	IS element ISM1 (ICSNY family)	
MSM0528	IS element ISM1 (ICSNY family)	
MSM0529	hypothetical protein	
MSM0530	predicted O-linked GlcNAc transferase	
MSM0531	adenine/cytosine DNA methyltransferase	
MSM0532	IS element ISM1 (ICSNY family)	
MSM0533	IS element ISM1 (ICSNY family)	
MSM0534	IS element ISM1 (ICSNY family)	
MSM0535	hypothetical protein	
MSM0536	hypothetical protein	
MSM0537	TPR-repeat protein	
MSM0538	pyruvate formate-lyase activating enzyme, PflA	
MSM0539	putative DNA-directed DNA polymerase	
MSM0540	predicted transcriptional regulator	
MSM0541	hypothetical protein	ND
MSM0542	coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase, Mer	
MSM0543	DNA repair photolyase, SplB	
MSM0544	predicted Fe-S oxidoreductase	
MSM0545	conserved hypothetical protein	
MSM0546	conserved hypothetical protein	
MSM0547	predicted nucleotidyltransferase, cytidyltransferase-related	
MSM0548	6-phosphogluconate dehydrogenase, beta-hydroxyacid dehydrogenase related, MmsB	
MSM0549	cytochrome C-type biogenesis protein, DsbD	
MSM0550	protein disulfide-isomerase, thioredoxin-related	
MSM0551	conserved hypothetical protein	
MSM0552	sulfur transfer protein involved in thiamine biosynthesis	
MSM0553	ATPase, PP-loop superfamily	
MSM0554	protein containing von Willebrand factor type A (vWA) domain, CoxE	
MSM0555	MoxR-like ATPase	
MSM0556	dihydropteroate synthase	
MSM0557	pyruvate:ferredoxin oxidoreductase, gamma subunit, PorG	
MSM0558	pyruvate:ferredoxin oxidoreductase, delta subunit, PorD	
MSM0559	pyruvate:ferredoxin oxidoreductase, alpha subunit, PorA	
MSM0560	pyruvate:ferredoxin oxidoreductase, beta subunit, PorB	
MSM0561	formate dehydrogenase, iron-sulfur subunit	
MSM0562	formate dehydrogenase, iron-sulfur subunit	
MSM0563	fumarate hydratase, alpha subunit	
MSM0564	phosphate uptake regulator, PhoU	
MSM0565	phosphate ABC transporter, ATPase component, PstB	
MSM0566	phosphate ABC transporter, permease component, PstA	
MSM0567	phosphate ABC transporter, permease component, PstC	
MSM0568	phosphate ABC transporter, phosphate-binding component, PstS	

MSM0569	phosphate transport system regulator related protein, PhoU					
MSM0570	conserved hypothetical protein					
MSM0571	conserved hypothetical protein					
MSM0572	H2-forming N5,N10-methylenetetrahydromethanopterin dehydrogenase (coenzyme F420-dependent), Mth					
MSM0573	biotin synthetase, BioB					
MSM0574	conserved hypothetical protein					
MSM0575	conserved hypothetical protein					
MSM0576	NIF3-related protein (NGG1p interacting factor 3)					
MSM0577	predicted dinucleotide-utilizing enzyme, ThiF/HesA family					
MSM0578	conserved hypothetical protein					
MSM0579	polyferredoxin, iron-sulfur binding					
MSM0580	putative adhesin-like protein					
MSM0581	conserved hypothetical membrane protein					
MSM0582	peptide methionine sulfoxide reductase, PMSR					
MSM0583	cobalt ABC transporter, permease component, CbiM					
MSM0584	cobalt ABC transporter, permease component					
MSM0585	cobalt ABC transporter, permease component, CbiQ					
MSM0586	cobalt ABC transporter, ATPase component, CbiO					
MSM0587	conserved hypothetical protein					
MSM0588	ferrous iron transport protein A, FeoA					
MSM0589	ferrous iron transport protein B, FeoB					
MSM0590	hypothetical protein					
MSM0591	hypothetical protein					
MSM0592	conserved hypothetical protein					
MSM0593	multidrug ABC transporter, ATPase component, CcmA					
MSM0594	multidrug ABC transporter, permease component					
MSM0595	multidrug ABC transporter, permease component					
MSM0596	bacterial type II secretion system protein, GspF					
MSM0597	bacterial type II/IV secretion system protein kinase, GspE					
MSM0598	SAM-dependent methyltransferase					
MSM0599	conserved hypothetical membrane protein					
MSM0600	transcriptional regulator, MarR family					
MSM0601	putative transposase	ND				
MSM0602	translation elongation factor EF-1, beta subunit					
MSM0603	predicted Zn-ribbon RNA-binding protein involved in translation					
MSM0604	predicted archaeal aspartate/glutamate/uridylate kinase					
MSM0605	peptidyl-tRNA hydrolase, PTH2 family					
MSM0606	hypothetical protein					
MSM0607	predicted ATPase, RNase L inhibitor family					
MSM0608	putative metal-binding protein					
MSM0609	ferredoxin, iron-sulfur binding					
MSM0610	aspartate aminotransferase					
MSM0611	DNA repair protein, RadB					
MSM0612	putative translation factor, Sua5/YciO/YrdC/YwlC family					
MSM0613	phosphatidylglycerophosphate synthase, PgsA					
MSM0614	conserved hypothetical protein					
MSM0615	archaeal fructose 1,6-bisphosphatase					
MSM0616	adhesin-like protein					
MSM0617	thiamine biosynthesis ATP pyrophosphatase, ThiI					
MSM0618	pH regulator (monovalent cation:H+ antiporter)					
MSM0619	alanyl-tRNA synthetase, AlaS					
MSM0620	ribosomal protein L12p					
MSM0621	ribosomal protein L10p					
MSM0622	ribosomal protein L1p					
MSM0623	ribosomal protein L11					
MSM0624	transcription antiterminator, NusG					
MSM0625	protein translocation complex sec61, gamma subunit					
MSM0626	cell division protein, FtsZ					
MSM0627	tetrahydromethanopterin S-methyltransferase, subunit H, MtrH					
MSM0628	conserved hypothetical protein					
MSM0629	putative transposase	ND				
MSM0630	conserved hypothetical protein					
MSM0631	transcription initiation factor IIE, alpha subunit					
MSM0632	predicted hydrolase, HD superfamily					
MSM0633	archaeosine tRNA-ribosyltransferase					
MSM0634	predicted metal-sulfur cluster biosynthetic enzyme					
MSM0635	predicted regulator of amino acid metabolism					
MSM0636	hydrogenase expression/formation protein, HypC					
MSM0637	dihydrolipoamide dehydrogenase					
MSM0638	pfam match to MurG; not predicted to be a carbohydrate active enzyme by CAZy					
MSM0639	putative cell wall biosynthesis protein					
MSM0640	cell division protein (RNA-binding), PelA					
MSM0641	prephenate dehydrogenase (NADP+)					
MSM0642	cell division control protein Cdc48, AAA+ ATPase family					

MSM0643	conserved hypothetical protein
MSM0644	thiamine biosynthesis protein, ThiC
MSM0645	ATP-dependent DNA ligase, Cdc9
MSM0646	conserved hypothetical protein
MSM0647	predicted RNA-binding protein, contains TRAM domain
MSM0648	phosphomannomutase, ManB
MSM0649	conserved hypothetical protein
MSM0650	transcriptional regulator, TetR/AcrR family
MSM0651	best blast hit to MTH_1585; not predicted to be a carbohydrate active enzyme by CAZy
MSM0652	pyruvate formate-lyase activating enzyme, PflA
MSM0653	histidinol-phosphate aminotransferase, HisC
MSM0654	carbonic anhydrase, Cab
MSM0655	glucose-1-phosphate thymidylyltransferase
MSM0656	phosphomannomutase, ManB
MSM0657	phosphoglycerate mutase, AP superfamily
MSM0658	hypothetical protein
MSM0659	conserved hypothetical membrane protein
MSM0660	LemA protein
MSM0661	small subunit ribosomal protein S3Ae
MSM0662	putative flagellar protein, FliL
MSM0663	dinitrogenase iron-molybdenum cofactor biosynthesis protein, NifX_NifB family
MSM0664	multimeric flavodoxin, NADPH-dependent FMN reductase family
MSM0665	5'-methylthioadenosine phosphorylase
MSM0666	conserved hypothetical protein
MSM0667	conserved hypothetical protein
MSM0668	conserved hypothetical protein
MSM0669	hypothetical protein
MSM0670	conserved hypothetical protein
MSM0671	cell division control protein Cdc6-related, AAA+ ATPase superfamily
MSM0672	thiamine pyrophosphokinase
MSM0673	conserved hypothetical membrane protein
MSM0674	hypothetical protein
MSM0675	hypothetical protein
MSM0676	conserved hypothetical membrane protein
MSM0677	archaeal aspartate aminotransferase
MSM0678	conserved hypothetical membrane protein
MSM0679	conserved hypothetical membrane protein
MSM0680	predicted ATPase, AAA+ superfamily
MSM0681	conserved hypothetical protein
MSM0682	hypothetical protein
MSM0683	conserved hypothetical protein
MSM0684	conserved hypothetical protein
MSM0685	hypothetical protein
MSM0686	acetolactate synthase (TPP-requiring), large subunit, IlvB
MSM0687	deoxycytidine-triphosphate deaminase, Dcd
MSM0688	4-oxalocrotonate tautomerase
MSM0689	hypothetical protein
MSM0690	helicase
MSM0691	mutator mutT protein (NUDIX domain)
MSM0692	conserved hypothetical protein
MSM0693	ATPase involved in DNA repair, SbcC
MSM0694	hypothetical protein
MSM0695	DNA repair helicase
MSM0696	Fe-S oxidoreductase
MSM0697	hypothetical protein
MSM0698	hypothetical protein
MSM0699	Na+-dependent transporter, SNF family
MSM0700	putative poly-gamma-glutamate synthesis protein, PgsA
MSM0701	signal recognition particle GTPase SRP54
MSM0702	predicted prefoldin, alpha subunit
MSM0703	ribosomal protein LX
MSM0704	translation initiation factor 6 (aIF-6)
MSM0705	ribosomal protein L31a
MSM0706	ribosomal protein L39a
MSM0707	predicted subunit of tRNA methyltransferase
MSM0708	dsDNA-binding protein
MSM0709	ribosomal protein S16a
MSM0710	RNA-binding protein, CRS1/YhbY family
MSM0711	ribonuclease P, subunit RPR2
MSM0712	conserved hypothetical protein (DUF1696 domain)
MSM0713	predicted nucleotide kinase
MSM0714	predicted GTPase
MSM0715	predicted GTPase
MSM0716	oligosaccharyl transferase, STT3 subunit

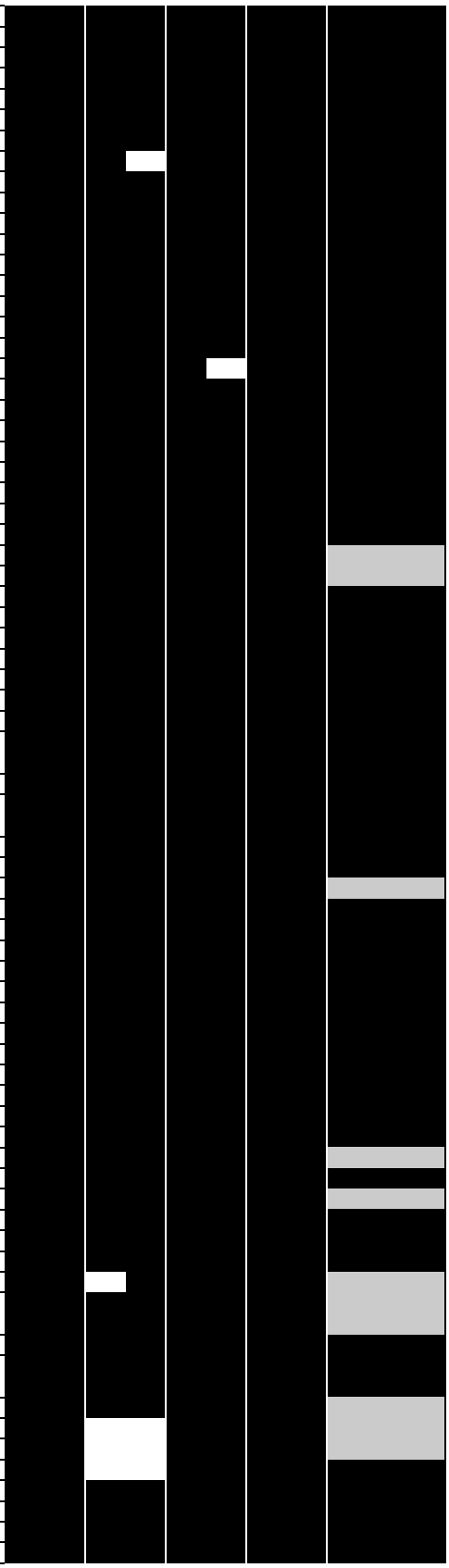


MSM0717	DNA topoisomerase I, TopA					
MSM0718	conserved hypothetical protein					
MSM0719	phosphoserine phosphatase, HAD family, SerB					
MSM0720	transcription initiation factor TFIID TATA binding protein					
MSM0721	adenylate cyclase, class 2					
MSM0722	2-isopropylmalate synthase, LeuA					
MSM0723	3-isopropylmalate dehydratase, LeuC					
MSM0724	4-hydroxybenzoate synthetase (chorismate lyase)					
MSM0725	DNA repair flap structure-specific 5'-3' endonuclease					
MSM0726	conserved hypothetical protein					
MSM0727	S-adenosylhomocysteine hydrolase (adenosylhomocysteinase), AhcY					
MSM0728	predicted oxidoreductase, aldo/keto reductase family					
MSM0729	molybdopterin biosynthesis protein, MoeB					
MSM0730	putative transposase	ND				
MSM0731	putative DNA helicase II, UvrD					
MSM0732	tRNA pseudouridine synthase B, TruB					
MSM0733	ribosomal protein L14e					
MSM0734	cytidylate kinase, Cmk					
MSM0735	ribosomal protein L34e					
MSM0736	conserved hypothetical membrane protein					
MSM0737	archaeal adenylate kinase, AdkA					
MSM0738	preprotein translocase, SecY subunit, SecY					
MSM0739	ribosomal protein L15p					
MSM0740	ribosomal protein L30p					
MSM0741	ribosomal protein S5p, RpsE					
MSM0742	ribosomal protein L18p, RplR					
MSM0743	ribosomal protein L19e					
MSM0744	ribosomal protein L32e					
MSM0745	ribosomal protein L6p, RplF					
MSM0746	ribosomal protein S8p					
MSM0747	ribosomal protein S14p					
MSM0748	ribosomal protein L5p					
MSM0749	ribosomal protein S4e					
MSM0750	ribosomal protein L24p					
MSM0751	ribosomal protein L14p					
MSM0752	ribosomal protein S17p					
MSM0753	ribonuclease P, subunit P29					
MSM0754	translation initiation factor SUI1					
MSM0755	ribosomal protein L29p					
MSM0756	ribosomal protein S3p					
MSM0757	ribosomal protein L22p					
MSM0758	ribosomal protein S19p					
MSM0759	ribosomal protein L2p					
MSM0760	ribosomal protein L23p					
MSM0761	ribosomal protein L1e					
MSM0762	ribosomal protein L3p					
MSM0763	conserved hypothetical protein					
MSM0764	ribosomal L11 RNA methyltransferase (SAM-dependent)					
MSM0765	pyruvate carboxylase (acetyl-CoA/biotin carboxylase), subunit A, PycA					
MSM0766	biotin-[acetyl-CoA-carboxylase] ligase/biotin operon regulator bifunctional protein, BirA					
MSM0767	selenocysteine synthase, SelA					
MSM0768	conserved hypothetical protein					
MSM0769	fumarate hydratase, class I					
MSM0770	cobalt ABC transporter, ATPase component, CbiO					
MSM0771	cobalt ABC transporter, permease component, CbiQ					
MSM0772	predicted permease, major facilitator superfamily					
MSM0773	multidrug ABC transporter, ATPase component					
MSM0774	multidrug ABC transporter, ATPase component					
MSM0775	transcriptional regulator, AraC family					
MSM0776	conserved hypothetical membrane protein					
MSM0777	conserved hypothetical protein					
MSM0778	predicted RNA-binding protein, eukaryotic snRNP-like					
MSM0779	predicted Zn-dependent hydrolase, metallo-beta-lactamase superfamily					
MSM0780	conserved hypothetical protein					
MSM0781	conserved hypothetical protein					
MSM0782	hypothetical protein					
MSM0783	tungsten formylmethanofuran dehydrogenase, subunit F, FwdF					
MSM0784	ferredoxin					
MSM0785	predicted phosphopantetheine adenylyltransferase (PPAT)					
MSM0786	transglutaminase-like protein, putative cysteine protease					
MSM0787	Fe-S oxidoreductase					
MSM0788	aspartate aminotransferase					

MSM0789	cation efflux system protein (zinc/cadmium/cobalt)
MSM0790	CBS-domain-containing protein
MSM0791	2-phosphoglycerate kinase
MSM0792	predicted calcineurin-like phosphoesterase
MSM0793	conserved hypothetical protein
MSM0794	conserved hypothetical protein
MSM0795	heterodisulfide reductase, subunit B, HdrB
MSM0796	heterodisulfide reductase, subunit C, HdrC
MSM0797	archaeosine tRNA-ribosyltransferase
MSM0798	hypothetical protein
MSM0799	conserved hypothetical protein
MSM0800	hypothetical protein
MSM0801	diphthine synthase, DphB
MSM0802	methyltransferase
MSM0803	predicted metal-dependent membrane protease, CAAX amino terminal protease family
MSM0804	translation initiation factor aIF-2B, alpha subunit
MSM0805	polar amino acid ABC transporter, ATPase component
MSM0806	polar amino acid ABC transporter, permease component
MSM0807	polar amino acid ABC transporter, substrate-binding component
MSM0808	nitrogenase iron-molybdenum cofactor biosynthesis protein, NifB
MSM0809	conserved hypothetical protein
MSM0810	activator of (R)-2-hydroxyglutaryl-CoA dehydratase
MSM0811	conserved hypothetical protein
MSM0812	conserved hypothetical protein
MSM0813	predicted peptidyl-prolyl cis-trans isomerase
MSM0814	phosphoribosylformylglycinamidine synthase-related protein (selenophosphate synthetase)
MSM0815	conserved hypothetical protein
MSM0816	predicted nucleic acid-binding protein, PIN domain-like family
MSM0817	predicted transcriptional regulator
MSM0818	predicted transcriptional regulator
MSM0819	putative transcription regulator, ArsR family
MSM0820	molybdenum cofactor biosynthesis protein, MoaB
MSM0821	orotate phosphoribosyltransferase, PyrE
MSM0822	photosynthetic reaction centre cytoplasmic domain-containing protein
MSM0823	phosphoenolpyruvate synthase/pyruvate phosphate dikinase, PpsA
MSM0824	putative N-acetyltransferase, GNAT family
MSM0825	adenosylcobinamide amidohydrolase, CbiZ
MSM0826	chaperonin, Cpn60/TCP-1/thermosome family, GroL
MSM0827	predicted metal-dependent hydrolase, cyclase family
MSM0828	best blast hit to Msp_0220; not predicted to be a carbohydrate active enzyme by CAZy
MSM0829	aspartate-semialdehyde dehydrogenase, Asd
MSM0830	dihydrodipicolinate reductas, DapB
MSM0831	dihydrodipicolinate synthase, DapA
MSM0832	aspartokinase, alpha subunit
MSM0833	ribosomal protein S17a
MSM0834	chorismate mutase
MSM0835	archaeal shikimate kinase
MSM0836	related to alpha-glycosyltransferases, GT4 family
MSM0837	cobalamin biosynthesis protein D, CbiD
MSM0838	putative thioredoxin/glutaredoxin
MSM0839	ATP-dependent helicase
MSM0840	conserved hypothetical protein
MSM0841	photosynthetic reaction centre cytoplasmic domain containing protein
MSM0842	histone acetyltransferase, radical SAM superfamily
MSM0843	2-deoxyribose-5-phosphate aldolase (DERA), DeoC
MSM0844	archaeal histone, HmtA
MSM0845	2-methylthioadenine synthetase, MiaB
MSM0846	uncharacterized archaeal Zn-finger protein
MSM0847	archaeal 3-isopropylmalate dehydratase, small subunit, LeuD
MSM0848	ribofuranosylaminobenzene 5'-phosphate synthase, RfaS
MSM0849	molybdenum cofactor biosynthesis-related protein, MoaA
MSM0850	predicted CDP-diglyceride synthetase
MSM0851	predicted transcriptional regulator
MSM0852	predicted ATP-utilizing enzyme
MSM0853	UDP-N-acetylglucosamine 2-epimerase, WecB
MSM0854	hypothetical protein
MSM0855	archaeal tRNA pseudouridine synthase A, TruA
MSM0856	antimicrobial peptide ABC transporter, permease component
MSM0857	antimicrobial peptide ABC transporter, ATPase component
MSM0858	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide (ProFAR) isomerase, HisA

MSM0859	glycerol-3-phosphate cytidylyltransferase	
MSM0860	aspartate-semialdehyde dehydrogenase, ArgC	
MSM0861	flavodoxin	
MSM0862	aspartate carbamoyltransferase regulatory chain, PyrI	
MSM0863	pyridoxamine-phosphate oxidase (FMN-binding)	
MSM0864	predicted transcriptional regulator	
MSM0865	putative glucose-methanol-choline oxidoreductase (FAD-dependent)	
MSM0866	Zn metalloprotease, TldD	
MSM0867	AMMECR1-related protein	
MSM0868	hypothetical protein	
MSM0869	GTPase, GTP1/OBG family	
MSM0870	molecular chaperone (small heat shock protein), HSP20/alpha crystallin family	
MSM0871	putative transposase	ND
MSM0872	glucosamine:fructose-6-phosphate aminotransferase (isomerizing), AgaS	
MSM0873	conserved hypothetical protein	
MSM0874	adenine deaminase, AdeC	
MSM0875	lysine-oxoglutarate reductase/Saccharopine dehydrogenase (LOR/SDH) bifunctional enzyme	
MSM0876	arginase/agmatinase/formimionoglutamate hydrolase, SpeB	
MSM0877	translation initiation factor 5A (aIF-5A)	
MSM0878	pyruvoyl-dependent arginine decarboxylase, PdaD	
MSM0879	Poly(P)/ATP NAD kinase, inositol monophosphatase family, PpnK	
MSM0880	UDP-N-acetylmuramyl tripeptide synthetase (Mur ligase)	
MSM0881	porphobilinogen deaminase	
MSM0882	3-chlorobenzoate-3,4-dioxygenase dyhydrogenase	
MSM0883	orotate phosphoribosyltransferase	
MSM0884	adhesin-like protein	
MSM0885	adhesin-like protein	
MSM0886	hypothetical protein	
MSM0887	universal stress protein, adenine nucleotide alpha hydrolase-like family	
MSM0888	glutamate dehydrogenase (NADP+), GdhA	
MSM0889	hypothetical protein	
MSM0890	hypothetical protein	
MSM0891	peptide chain release factor eRF, subunit 1	
MSM0892	putative zinc-binding protein	
MSM0893	acetyltransferase	
MSM0894	conserved hypothetical protein	
MSM0895	cation transport ATPase, HAD family	
MSM0896	precorrin-6X reductase, CbiJ	
MSM0897	ribosomal protein S10p	
MSM0898	translation elongation factor 1-alpha (EF-Tu)	
MSM0899	translation elongation factor EF-2, FusA	
MSM0900	ribosomal protein S7p	
MSM0901	ribosomal protein S12p	
MSM0902	methyl-coenzyme M reductase, alpha subunit, McrA	
MSM0903	methyl-coenzyme M reductase, gamma subunit, McrG	
MSM0904	methyl-coenzyme M reductase, D subunit, McrD	
MSM0905	methyl-coenzyme M reductase, beta subunit, McrB	
MSM0906	transcription termination factor, NusA	
MSM0907	ribosomal protein L17Ae	
MSM0908	DNA-dependent RNA polymerase, subunit A, RpoA	
MSM0909	DNA-dependent RNA polymerase, subunit A', RpoA	
MSM0910	DNA-dependent RNA polymerase, subunit B', RpoB	
MSM0911	DNA-dependent RNA polymerase, subunit B, RpoB	
MSM0912	DNA-dependent RNA polymerase, subunit H, RpoH	
MSM0913	hypothetical protein	
MSM0914	predicted O-linked GlcNAc transferase	
MSM0915	hypothetical protein	
MSM0916	hydroxyethylthiazole kinase, ThiM	
MSM0917	thiamine monophosphate synthase, ThiE	
MSM0918	3-phosphoglycerate kinase, Pgk	
MSM0919	triosephosphate isomerase, TpiA	
MSM0920	conserved hypothetical protein	
MSM0921	predicted surface protein	
MSM0922	Fe-S oxidoreductase	
MSM0923	multimeric flavodoxin	
MSM0924	succinyl-CoA synthetase, beta subunit, SucC	
MSM0925	2-oxoglutarate ferredoxin oxidoreductase, gamma subunit, KorC	
MSM0926	2-oxoglutarate ferredoxin oxidoreductase, beta subunit, KorB	
MSM0927	2-oxoglutarate ferredoxin oxidoreductase, alpha subunit, KorA	
MSM0928	2-oxoglutarate ferredoxin oxidoreductase, delta subunit, KorD	
MSM0929	fumarate hydratase, FumA	
MSM0930	peptidyl-prolyl cis-trans isomerase, FKBP-type	

MSM0931	conserved hypothetical protein
MSM0932	conserved hypothetical protein
MSM0933	cobalamin-5-phosphate synthase, CobS
MSM0934	predicted phosphatidylglycerophosphatase A-related protein
MSM0935	conserved hypothetical protein
MSM0936	transcription regulator-related ATPase, ExsB
MSM0937	HD superfamily hydrolase
MSM0938	hypothetical protein
MSM0939	pyruvate carboxylase, subunit B, PycB
MSM0940	myo-inositol-1-phosphate synthase
MSM0941	prenyltransferase, UbiA
MSM0942	conserved hypothetical membrane protein
MSM0943	conserved hypothetical protein
MSM0944	CMP-N-acetylneuraminic acid synthetase, NeuA
MSM0945	hydrogenase expression/formation protein, HypD
MSM0946	archaeal sucrose-phosphate phosphatase (SPP-like), HAD family
MSM0947	predicted zinc metalloprotease, modulator of DNA gyrase
MSM0948	hypothetical protein
MSM0949	transcriptional activator
MSM0950	molybdopterin biosynthesis protein, MoeA
MSM0951	translation initiation factor aIF-1A
MSM0952	serine/threonine protein kinase, RIO1 family
MSM0953	conserved hypothetical membrane protein
MSM0954	predicted RNA-binding protein
MSM0955	type II DNA topoisomerase VI, subunit B
MSM0956	type II DNA topoisomerase VI, subunit A
MSM0957	adhesin-like protein
MSM0958	predicted 1,4-beta-cellobiosidase
MSM0959	conserved hypothetical protein
MSM0960	cation transport ATPase, HAD family
MSM0961	heavy-metal cation transporting ATPase
MSM0962	glyceraldehyde 3-phosphate dehydrogenase, GapA
MSM0963	endonuclease IV, xylose isomerase-like TIM barrel family, Nfo
MSM0964	calcineurin-like phosphoesterase
MSM0965	3-hydroxyacyl-CoA dehydrogenase, FadB
MSM0966	predicted 26S protease regulatory subunit (ATP-dependent), AAA+ family ATPase
MSM0967	glutamyl-tRNA reductase, HemA
MSM0968	bifunctional precorrin-2 oxidase/chelatase (siroheme synthase), CysG
MSM0969	predicted metal-binding transcription factor
MSM0970	conserved hypothetical protein
MSM0971	methyl-coenzyme M reductase, component A2
MSM0972	tRNA-dihydrouridine synthase
MSM0973	GTP cyclohydrolase III, GGDN family
MSM0974	LPPG:FO 2-phospho-L-lactate transferase, CofD
MSM0975	F420-0:gamma-glutamyl ligase, CofE
MSM0976	archaeal IMP cyclohydrolase, PurO
MSM0977	putative biopolymer transport protein, ExbD/TolR family
MSM0978	biopolymer transport protein, MotA/TolQ/ExbB proton channel family
MSM0979	ribonuclease HII, RnhB
MSM0980	rod shape-determining protein, MreB/Mrl family
MSM0981	conserved hypothetical protein
MSM0982	phosphatidylserine synthase, PssA
MSM0983	conserved hypothetical protein
MSM0984	sortase (surface protein transpeptidase), SrtA
MSM0985	conserved hypothetical protein
MSM0986	conjugated bile acid hydrolase (CBAH)
MSM0987	tyrosine decarboxylase, MfnA
MSM0988	phosphoenolpyruvate synthase, PpsA
MSM0989	ribosomal protein L10e
MSM0990	nitrate/sulfonate/bicarbonate ABC transporter, ATPase component
MSM0991	nitrate/sulfonate/bicarbonate ABC transporter, substrate-binding component
MSM0992	conserved hypothetical protein
MSM0993	putative ATPase, glucocorticoid receptor-like (DNA-binding domain) family
MSM0994	predicted nucleotidyltransferase
MSM0995	adhesin-like protein
MSM0996	adhesin-like protein
MSM0997	dihydroorotase, PyrC
MSM0998	polyferredoxin, MvhB
MSM0999	methyl viologen-reducing hydrogenase, alpha subunit, MvhA
MSM1000	methyl viologen-reducing hydrogenase, gamma subunit, MvhG
MSM1001	methyl viologen-reducing hydrogenase, delta subunit, MvhD

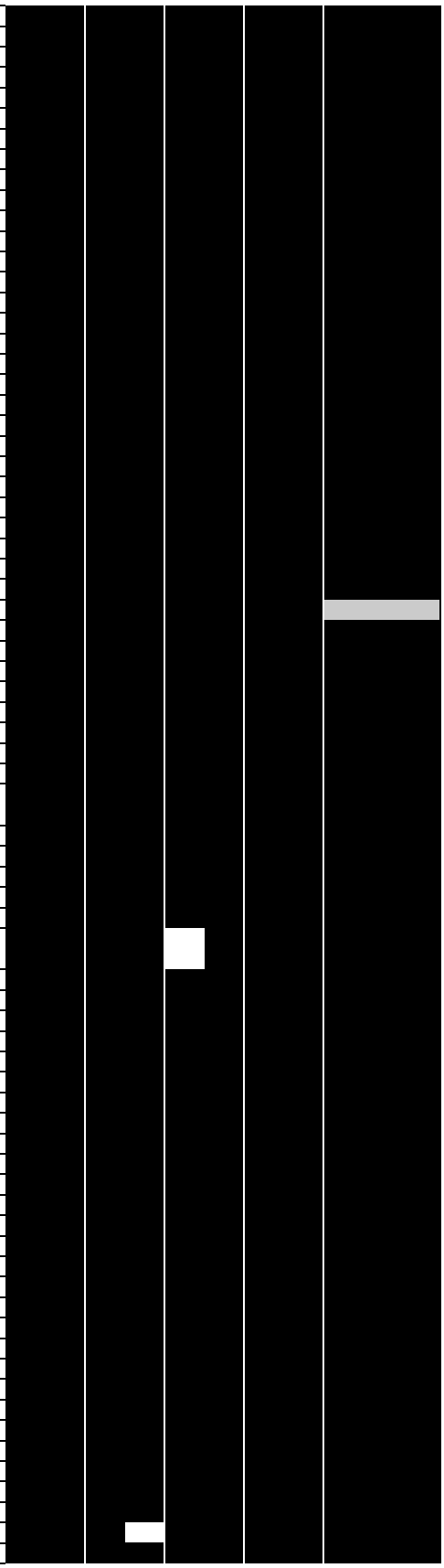


MSM1002	ABC transporter involved in Fe-S cluster assembly, permease component
MSM1003	ABC transporter involved in Fe-S cluster assembly, ATPase component
MSM1004	photosynthetic reaction centre cytoplasmic domain containing protein
MSM1005	GTP:adenosylcobinamide-phosphate guanylyltransferase
MSM1006	conserved hypothetical protein
MSM1007	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H, MtrH
MSM1008	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit G, MtrG
MSM1009	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit F, MtrF
MSM1010	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit A, MtrA
MSM1011	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit B, MtrB
MSM1012	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit C, MtrC
MSM1013	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit D, MtrD
MSM1014	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit E, MtrE
MSM1015	methyl-coenzyme M reductase, alpha subunit, McrA
MSM1016	methyl-coenzyme M reductase, gamma subunit, McrG
MSM1017	methyl-coenzyme M reductase, C subunit, McrC
MSM1018	methyl-coenzyme M reductase, D subunit, McrD
MSM1019	methyl-coenzyme M reductase, beta subunit, McrB
MSM1020	Fe-S oxidoreductase, Radical SAM family
MSM1021	uncharacterized protein related to methyl coenzyme M reductase subunit C (McrC)
MSM1022	conserved hypothetical protein
MSM1023	2-phosphosulpholactate phosphatase, ComB, (coenzyme M biosynthesis)
MSM1024	pheromone shutdown protein, TraB family
MSM1025	conserved hypothetical protein
MSM1026	hemolysin-related protein, transporter-associated family, TlyC
MSM1027	Ca2+/Na+ antiporter (K+-dependent)
MSM1028	predicted ATPase, PP-loop family
MSM1029	conserved hypothetical protein
MSM1030	predicted pyridoxal phosphate-dependent enzyme
MSM1031	N2,N2-dimethylguanosine tRNA methyltransferase, Trm1
MSM1032	transcriptional regulator, Lrp family
MSM1033	conserved hypothetical protein
MSM1034	conserved hypothetical protein
MSM1035	FO synthase subunit 1 (SAM-dependent), CofG (F420 biosynthesis)
MSM1036	predicted methyltransferase
MSM1037	proteasome, beta subunit
MSM1038	predicted metal-dependent RNase
MSM1039	phosphoribosylformylglycinamidine cyclo-ligase (AIRS), PurM
MSM1040	malate/L-lactate dehydrogenase
MSM1041	DNA-dependent DNA polymerase I, PolB1
MSM1042	predicted permease
MSM1043	dihydroorotate dehydrogenase electron transfer subunit, PyrK
MSM1044	dihydroorotate dehydrogenase, PyrD
MSM1045	possible glycosyltransferase
MSM1046	pre-mRNA splicing ribonucleoprotein PRP31
MSM1047	fibrillarin-like pre-rRNA processing protein, FlpA
MSM1048	phosphopantothenoylcysteine synthetase/decarboxylase
MSM1049	phosphopantothenoylcysteine synthetase/decarboxylase
MSM1050	conserved hypothetical protein
MSM1051	putative endoglucanase
MSM1052	prephenate dehydratase, PheA
MSM1053	IMP dehydrogenase related protein
MSM1054	IMP dehydrogenase related protein
MSM1055	coenzyme PQQ synthesis protein, SAM family
MSM1056	6-pyruvoyl-tetrahydropterin synthase
MSM1057	conserved hypothetical protein
MSM1058	conserved hypothetical protein
MSM1059	predicted RecB family exonuclease
MSM1060	energy-converting hydrogenase B, subunit Q, EhbQ
MSM1061	energy-converting hydrogenase B, subunit P, EhbP
MSM1062	energy-converting hydrogenase B, subunit O, EhbO
MSM1063	energy-converting hydrogenase B, subunit N, EhbN
MSM1064	energy-converting hydrogenase B, subunit M, EhbM

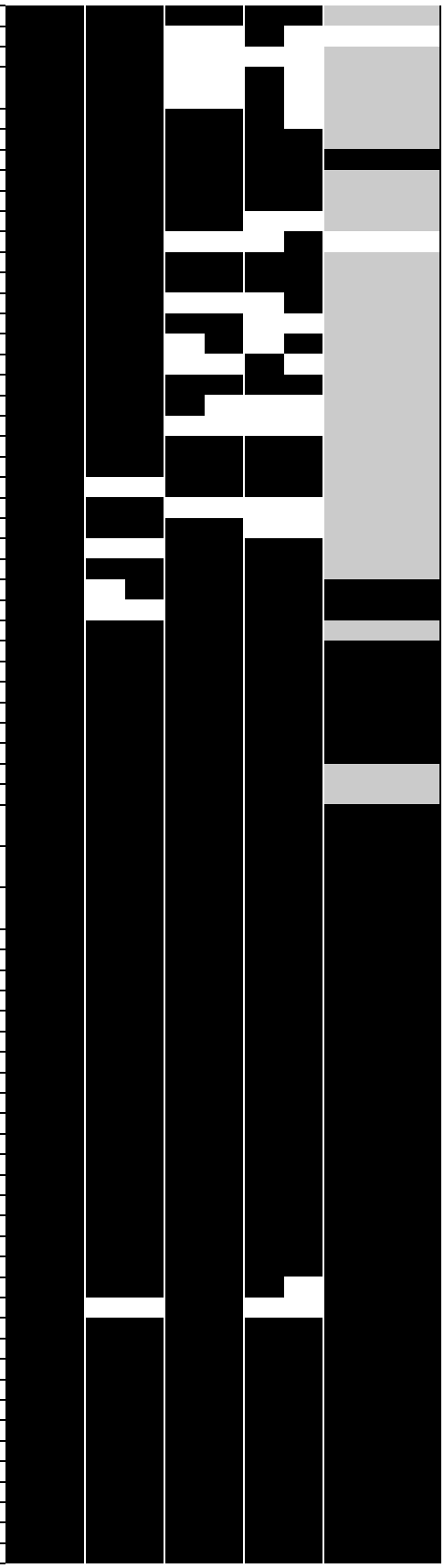
MSM1065	energy-converting hydrogenase B, subunit L, EhbL	
MSM1066	energy-converting hydrogenase B, subunit K, EhbK	
MSM1067	energy-converting hydrogenase B, subunit J, EhbJ	
MSM1068	energy-converting hydrogenase B, subunit I, EhbI	
MSM1069	energy-converting hydrogenase B, subunit H, EhbH	
MSM1070	energy-converting hydrogenase B, subunit G, EhbG	
MSM1071	energy-converting hydrogenase B, subunit F, EhbF	
MSM1072	energy-converting hydrogenase B, subunit E, EhbE	
MSM1073	energy-converting hydrogenase B, subunit D, EhbD	
MSM1074	energy-converting hydrogenase B, subunit C, EhbC	
MSM1075	energy-converting hydrogenase B, subunit B, EhbB	
MSM1076	energy-converting hydrogenase B, subunit A, EhbA	
MSM1077	putative permease (transport)	
MSM1078	predicted bile acid/sodium symporter	
MSM1079	predicted membrane-bound metal-dependent hydrolase, NCS2 family	
MSM1080	predicted deacylase	
MSM1081	transcriptional regulator (enhancer-binding protein), DNA2/NAM7 helicase family	
MSM1082	hypothetical protein	
MSM1083	conserved hypothetical membrane protein	
MSM1084	argininosuccinate synthase, ArgG	
MSM1085	aquaporin, MIP superfamily, AqpM	
MSM1086	conserved hypothetical protein	
MSM1087	NAD-dependent protein deacetylase, SIR2 family	
MSM1088	hypothetical protein	
MSM1089	hypothetical protein	
MSM1090	sugar fermentation stimulation protein, SfsA	
MSM1091	sugar kinase, YjeF-related protein family	
MSM1092	formylmethanofuran:tetrahydromethanopterin formyltransferase, Ftr	
MSM1093	putative transposase	ND
MSM1094	conserved hypothetical integral membrane protein	
MSM1095	Trk-type potassium transport system, membrane component, TrkH	
MSM1096	Trk-type potassium transport system, NAD-binding component, TrkA	
MSM1097	Zn-dependent hydrolase	
MSM1098	archaeal holliday junction resolvase	
MSM1099	biotin synthase related protein	
MSM1100	conserved hypothetical protein	
MSM1101	Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase, B subunit, GatB	
MSM1102	IMP dehydrogenase related protein	
MSM1103	phosphoribosyl-ATP pyrophosphohydrolase, HisE	
MSM1104	acetyltransferase, GNAT family	
MSM1105	NCAIR mutase related protein, PurE	
MSM1106	hydrogenase maturation factor, HypF	
MSM1107	predicted transcriptional regulator	
MSM1108	molecular chaperone GrpE	
MSM1109	molecular chaperone DnaK	
MSM1110	molecular chaperone DnaJ	
MSM1111	adhesin-like protein	
MSM1112	adhesin-like protein	
MSM1113	adhesin-like protein	
MSM1114	adhesin-like protein	
MSM1115	putative transposase	ND
MSM1116	adhesin-like protein	
MSM1117	cobalamin biosynthesis protein N, CobN	
MSM1118	conserved hypothetical protein	
MSM1119	conserved hypothetical protein	
MSM1120	methionine aminopeptidase, Map	
MSM1121	coenzyme F420-reducing hydrogenase, beta subunit, FrhB	
MSM1122	coenzyme F420-reducing hydrogenase, gamma subunit, FrhG	
MSM1123	coenzyme F420-reducing hydrogenase, delta subunit, FrhD	
MSM1124	coenzyme F420-reducing hydrogenase, alpha subunit, FrhA	
MSM1125	predicted endoglucanase (CobN-related)	
MSM1126	predicted transcriptional regulator, ArsR family	
MSM1127	cation transport ATPase, HAD family	
MSM1128	hypothetical protein	
MSM1129	conserved hypothetical protein	
MSM1130	conserved hypothetical protein	
MSM1131	conserved hypothetical protein	
MSM1132	ribosome biogenesis protein Nop10	
MSM1133	translation initiation factor aIF-2, alpha subunit	
MSM1134	ribosomal protein S27e	
MSM1135	ribosomal protein L44e	
MSM1136	conserved hypothetical protein	
MSM1137	DNA polymerase sliding clamp subunit, PCNA family, Pcn	
MSM1138	predicted glutamine amidotransferase, CobB/CobQ-like family	

MSM1139	cell wall biosynthesis protein, MurD-like peptide ligase family					
MSM1140	hypothetical protein					
MSM1141	tryptophan synthase, alpha subunit, TrpA					
MSM1142	tryptophan synthase, beta subunit, TrpB					
MSM1143	indole-3-glycerol phosphate synthase, TrpC					
MSM1144	anthranilate phosphoribosyltransferase, TrpD					
MSM1145	anthranilate/para-aminobenzoate synthase component II, TrpG					
MSM1146	anthranilate/para-aminobenzoate synthase component I, TrpE					
MSM1147	hypothetical protein					
MSM1148	predicted metal-dependent membrane protease					
MSM1149	conserved hypothetical membrane protein					
MSM1150	predicted transcriptional regulator					
MSM1151	adenylosuccinate lyase, PurB					
MSM1152	conserved hypothetical membrane protein					
MSM1153	cation transport ATPase, HAD family					
MSM1154	metal-dependent amidohydrolase					
MSM1155	conserved hypothetical protein					
MSM1156	tRNA pseudouridine synthase D, TruD					
MSM1157	hypothetical protein					
MSM1158	hydrogenase expression/formation protein, HypE					
MSM1159	glutamine amidotransferase, HisH					
MSM1160	nitrogenase molybdenum-iron protein, NifD					
MSM1161	hypothetical protein					
MSM1162	conserved hypothetical protein					
MSM1163	hypothetical protein					
MSM1164	predicted GTPase, HSR1-related family					
MSM1165	predicted phosphohydrolase (metal-dependent)					
MSM1166	conserved hypothetical membrane protein					
MSM1167	cobalt precorrin-6Y C5,15-methyltransferase, CbiE					
MSM1168	putative adhesin-like protein					
MSM1169	hypothetical protein					
MSM1170	arsenite-transporting ATPase					
MSM1171	ammonia-dependent NAD+ synthetase, NadE					
MSM1172	leucyl-tRNA synthetase, LeuS					
MSM1173	tRNA(1-methyladenosine) methyltransferase					
MSM1174	heat shock protein HtpX (Zn-dependent)					
MSM1175	conserved hypothetical membrane protein					
MSM1176	replication factor C, small subunit, RfcS					
MSM1177	replication factor C, large subunit, RfcL					
MSM1178	putative ATPase implicated in cell cycle control					
MSM1179	shikimate 5-dehydrogenase, AroE					
MSM1180	predicted metal-dependent membrane protease					
MSM1181	histidyl-tRNA synthetase, HisS					
MSM1182	phosphoribosyl-AMP cyclohydrolase, HisI					
MSM1183	ATPase, PilT family					
MSM1184	sugar phosphate isomerase/epimerase, AP endonuclease family 2					
MSM1185	methylated-DNA-[protein]-cysteine S-methyltransferase					
MSM1186	potassium transport system, membrane component, KefB					
MSM1187	ERCC4-like helicase					
MSM1188	adhesin-like protein					
MSM1189	putative transposase	ND				
MSM1190	cell wall biosynthesis protein, UDP-N-acetylmuramate-alanine ligase family					
MSM1191	cell wall biosynthesis protein, MurD-like peptide ligase family					
MSM1192	conserved hypothetical protein					
MSM1193	single-stranded DNA-specific exonuclease, DHH family					
MSM1194	ribosomal protein S15p					
MSM1195	xanthosine triphosphate pyrophosphatase, Ham1 family					
MSM1196	predicted archaeal ATPase, AAA+ superfamily					
MSM1197	predicted ATPase, AAA+ superfamily					
MSM1198	O-sialoglycoprotein endopeptidase					
MSM1199	conserved hypothetical protein					
MSM1200	phosphoribosyltransferase, CobT					
MSM1201	undecaprenyl-diphosphatase, UppP					
MSM1202	branched-chain-amino-acid aminotransferase, IlvE					
MSM1203	Zn-dependent protease, peptidase M48 family					
MSM1204	coenzyme F420-dependent methylenetetrahydromethanopterin dehydrogenase, Mtd					
MSM1205	conserved hypothetical membrane protein					
MSM1206	imidazoleglycerol-phosphate dehydrogenase, HisB					
MSM1207	molybdate transport system regulatory protein					
MSM1208	teichoic acid transporter					
MSM1209	multimeric flavodoxin					
MSM1210	efflux pump antibiotic resistance protein, MFS permease family					
MSM1211	putative phosphoserine phosphatase					
MSM1212	conserved hypothetical protein					

MSM1213	3-hexulose 6-phosphate synthase/formaldehyde activating enzyme
MSM1214	threonyl-tRNA synthetase, ThrS
MSM1215	cobyrinic acid a,c-diamide synthase, CbiA
MSM1216	conserved hypothetical membrane protein
MSM1217	type II restriction endonuclease
MSM1218	predicted acid phosphatase (survival protein), SurE
MSM1219	hypothetical protein
MSM1220	small nucleolar ribonucleoprotein, Sm-like family
MSM1221	actin-like ATPase
MSM1222	ketol-acid reductoisomerase, IlvC
MSM1223	carbonic anhydrase
MSM1224	acetolactate synthase, small subunit (regulatory), IlvH
MSM1225	acetolactate synthase, large subunit (TPP-requiring), IlvB
MSM1226	ornithine carbamoyltransferase, ArgF
MSM1227	phosphoribosylamine-glycine ligase, PurD
MSM1228	Na+-driven multidrug efflux pump
MSM1229	Na+-driven multidrug efflux pump
MSM1230	transcriptional regulator, MarR family
MSM1231	arginyl-tRNA synthetase, ArgS
MSM1232	signal peptidase I
MSM1233	glutamate-1-semialdehyde 2,1-aminomutase, HemL
MSM1234	cobalt-precorrin-8X methylmutase, CbiC
MSM1235	predicted flavoprotein
MSM1236	aspartyl-tRNA synthetase, AspS
MSM1237	dihydroxy-acid dehydratase, IlvD
MSM1238	histidinol dehydrogenase, HisD
MSM1239	predicted DNA-binding protein
MSM1240	predicted AAA ATPase
MSM1241	chromosome partitioning ATPase
MSM1242	tryptophan synthase, beta subunit, TrpB
MSM1243	putative actin-like ATPase
MSM1244	predicted metal-dependent phosphoesterases, PHP family
MSM1245	archaeal DNA-binding protein, AlbA
MSM1246	isopropylmalate synthase, LeuA
MSM1247	serine/threonine protein kinase related protein (PQQ-binding)
MSM1248	multidrug ABC transporter, permease component
MSM1249	multidrug ABC transporter, ATPase component
MSM1250	predicted transcriptional regulator, PadR-like family
MSM1251	predicted sugar phosphate isomerase/epimerase, AP endonuclease family 2
MSM1252	cation transporting P-type ATPase, HAD family
MSM1253	glutamyl-tRNA (Gln) amidotransferase subunit A, GatA
MSM1254	cobyric acid synthase
MSM1255	hypothetical protein
MSM1256	3,4-dihydroxy-2-butanone 4-phosphate synthase, RibB
MSM1257	predicted transcriptional regulator of riboflavin/FAD biosynthetic operon
MSM1258	fumarate reductase/succinate dehydrogenase flavoprotein, Sdh
MSM1259	predicted metal-dependent hydrolase, TRZ/ATZ family
MSM1260	archaeal histone
MSM1261	ATP phosphoribosyltransferase, HisG
MSM1262	flavodoxin (protoporphyrinogen oxidase)
MSM1263	aspartate carbamoyltransferase, PyrB
MSM1264	cell division control protein 6, Cdc6
MSM1265	conserved hypothetical protein
MSM1266	cobalamin biosynthesis protein D, CobD
MSM1267	cobalamin biosynthesis protein G, CbiG
MSM1268	conserved hypothetical protein
MSM1269	putative Met repressor-like protein
MSM1270	fuculose-1-phosphate aldolase, class II aldolase/adducin family
MSM1271	archaeal DNA polymerase II, small subunit
MSM1272	conserved hypothetical protein
MSM1273	cobalt precorrin-3B C17-methyltransferase, CbiH
MSM1274	predicted potassium ion transport protein
MSM1275	mgtE-like divalent cation transporter
MSM1276	conserved hypothetical protein
MSM1277	conserved hypothetical membrane protein
MSM1278	predicted archaeal ATPase, AAA+ superfamily
MSM1279	predicted nucleic-acid-binding protein containing a Zn-ribbon
MSM1280	sirohydrochlorin cobaltochelatase, CbiX
MSM1281	sirohydrochlorin cobaltochelatase-related protein
MSM1282	putative adhesin-like protein
MSM1283	thiamine monphosphate kinase, ThiL
MSM1284	pyruvate formate-lyase activating enzyme, PflA
MSM1285	conserved hypothetical protein
MSM1286	3-octaprenyl-4-hydroxybenzoate carboxy-lyase, UbiD

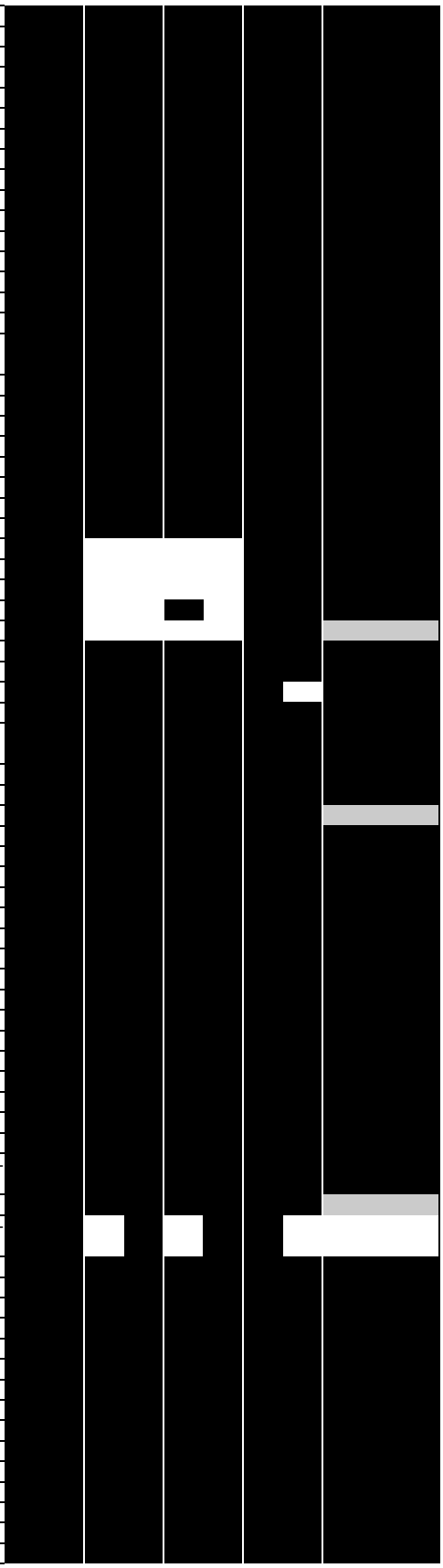


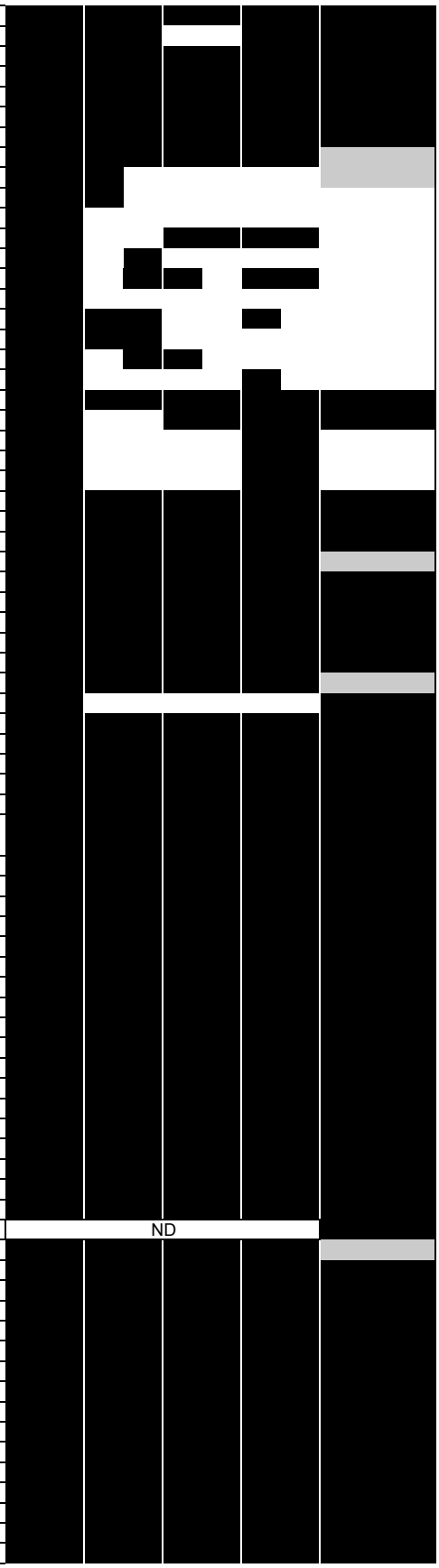
MSM1287	phosphoribosylaminoimidazole carboxylase (NCAIR mutase), PurE
MSM1288	conserved hypothetical membrane protein
MSM1289	GtrA-like surface polysaccharide biosynthesis protein, GtrA
MSM1290	glycosyltransferase (related to beta-glycosidases), GT2 family [CAZy]
MSM1291	conserved hypothetical membrane protein
MSM1292	transcriptional accessory protein, S1 RNA binding family, Tex
MSM1293	nitroreductase, NADH oxidase/flavin reductase family
MSM1294	glycosyltransferase, GT2 family
MSM1295	predicted DNA-binding protein
MSM1296	riboflavin synthase, beta subunit, RibH
MSM1297	glycosyltransferase, GT2 family
MSM1298	3-isopropylmalate dehydrogenase, LeuB
MSM1299	3-isopropylmalate dehydratase, small subunit, LeuD
MSM1300	3-isopropylmalate dehydratase, large subunit, LeuC
MSM1301	predicted Fe-S oxidoreductase
MSM1302	conserved hypothetical protein
MSM1303	UDP-N-acetyl-D-mannosaminuronate dehydrogenase
MSM1304	dTDP-4-dehydrorhamnose reductase, RfbD
MSM1305	adhesin-like protein
MSM1306	adhesin-like protein
MSM1307	dTDP-glucose pyrophosphorylase, RfbA
MSM1308	dTDP-4-dehydrorhamnose 3,5-epimerase
MSM1309	dTDP-D-glucose 4,6-dehydratase, RfbB
MSM1310	glycosyltransferase, GT2 family
MSM1311	glycosyltransferase, GT2 family
MSM1312	glycosyltransferase, GT2 family
MSM1313	distantly related to glycosyltransferases, GT4 family
MSM1314	hypothetical protein
MSM1315	predicted transcriptional regulator
MSM1316	glycosyltransferase, GT2 family
MSM1317	distantly related to glycosyltransferases, GT4 family
MSM1318	conserved hypothetical protein
MSM1319	conserved hypothetical protein
MSM1320	possible glycosyltransferase
MSM1321	predicted glycosyltransferase, GT2 family
MSM1322	distantly related to alpha-glycosyltransferases, GT4 family
MSM1323	glycosyltransferase, GT2 family
MSM1324	glycosyltransferase, GT2 family
MSM1325	predicted polysaccharide/polyol phosphate ABC transporter, permease component
MSM1326	polysaccharide/polyol phosphate ABC transporter, ATPase component
MSM1327	predicted CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase
MSM1328	glycosyltransferase, GT2 family
MSM1329	predicted glycosyltransferase, GT2 family
MSM1330	predicted glycosyltransferase, GT2 family
MSM1331	bacterial sugar transferase, WcaJ
MSM1332	ssDNA-binding protein
MSM1333	DNA repair protein RadA, RadA
MSM1334	predicted permease
MSM1335	hypothetical protein
MSM1336	heterodisulfide reductase, subunit A, HdrA
MSM1337	glycine hydroxymethyltransferase, GlyA
MSM1338	archaeal flavoprotein
MSM1339	conserved hypothetical protein
MSM1340	archaeal S-adenosylmethionine synthetase, MetK
MSM1341	isoleucyl-tRNA synthetase, IleS
MSM1342	phosphoribosylformylglycinamidine (FGAM) synthase, PurL
MSM1343	molybdenum cofactor biosynthesis protein, MoeA
MSM1344	predicted membrane-associated Zn-dependent protease
MSM1345	hypothetical protein
MSM1346	conserved hypothetical protein
MSM1347	hypothetical protein
MSM1348	rubrerythrin
MSM1349	F420H2-oxidase/flavoprotein, FprA
MSM1350	predicted transcriptional regulator, ArsR family
MSM1351	precorrin-2 C20-methyltransferase, CbiL
MSM1352	predicted ATP-dependent DNA helicase
MSM1353	putative topoisomerase IV, subunit A
MSM1354	DNA-directed RNA polymerase subunit M, RpoM
MSM1355	ADP-ribose pyrophosphatase, NUDIX hydrolase family
MSM1356	DNA-directed RNA polymerase, subunit L, RpoL
MSM1357	predicted RNA-binding protein
MSM1358	diphthamide synthase, subunit DPH2



MSM1359	adenine phosphoribosyltransferase, Apt	
MSM1360	signal recognition particle GTPase SRP54	
MSM1361	predicted pseudouridylate synthase	
MSM1362	molybdenum cofactor biosynthesis protein C, MoaC	
MSM1363	preprotein translocase, SecG subunit, SecG	
MSM1364	imidazoleglycerol-phosphate synthase, HisF	
MSM1365	3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase	
MSM1366	lactoylglutathione lyase, LglU	
MSM1367	peptidyl-prolyl cis-trans isomerase, PpiB	
MSM1368	N-acetylornithine aminotransferase, ArgD	
MSM1369	MutT-related protein, NUDIX family	
MSM1370	conserved hypothetical membrane protein	
MSM1371	diaminopimelate decarboxylase, LysA	
MSM1372	diaminopimelate epimerase, DapF	
MSM1373	methyltransferase, HemK	
MSM1374	dimethyladenosine transferase, KsgA	
MSM1375	predicted RNA-binding protein	
MSM1376	DNA-directed RNA polymerase subunit F	
MSM1377	ribosomal protein L21e	
MSM1378	putative monooxygenase, ABM family	
MSM1379	predicted NADP-dependent alcohol dehydrogenase	
MSM1380	NADP-dependent alcohol dehydrogenase	
MSM1381	putative NADP-dependent alcohol dehydrogenase	
MSM1382	conserved hypothetical membrane protein	
MSM1383	anaerobic ribonucleoside-triphosphate reductase, NrdD	
MSM1384	archaeal DNA polymerase II, large subunit, PolC	
MSM1385	predicted acyltransferase	
MSM1386	cytosine deaminase	
MSM1387	lysyl-tRNA synthetase (class I), LysS	
MSM1388	thiamine biosynthesis protein, ThiC	
MSM1389	sugar kinase, ribokinase/pfkB superfamily	
MSM1390	transcriptional regulator, LysR family	
MSM1391	predicted sugar phosphate isomerase involved in capsule formation, GutQ	
MSM1392	formate dehydrogenase accessory protein FdhD, FdhD	
MSM1393	iron(III) ABC transporter, substrate-binding component	
MSM1394	iron(III) ABC transporter, permease component	
MSM1395	iron(III) ABC transporter, ATPase component	
MSM1396	tungsten formylmethanofuran dehydrogenase, subunit E, FwdE	
MSM1397	adhesin-like protein	
MSM1398	adhesin-like protein	
MSM1399	adhesin-like protein	
MSM1400	putative antimicrobial peptide ABC transporter, permease component	
MSM1401	biopolymer transport protein	
MSM1402	conserved hypothetical protein	
MSM1403	formate/nitrite transporter, FdhC	
MSM1404	formate dehydrogenase, alpha subunit, FdhA	
MSM1405	formate dehydrogenase, beta subunit, FdhB	
MSM1406	molybdopterin cofactor biosynthesis protein A, MoaA	
MSM1407	molybdopterin-guanine dinucleotide biosynthesis protein B, MobB	
MSM1408	tungsten formylmethanofuran dehydrogenase, subunit E, FwdE	
MSM1409	tungsten formylmethanofuran dehydrogenase, subunit F, FwdF	
MSM1410	tungsten formylmethanofuran dehydrogenase, subunit G, FwdG	
MSM1411	tungsten formylmethanofuran dehydrogenase, subunit D, FwdD	
MSM1412	tungsten formylmethanofuran dehydrogenase, subunit B, FwdB	
MSM1413	tungsten formylmethanofuran dehydrogenase, subunit A, FwdA	
MSM1414	tungsten formylmethanofuran dehydrogenase, subunit C, FwdC	
MSM1415	conserved hypothetical protein	
MSM1416	conserved hypothetical protein	
MSM1417	conserved hypothetical protein	
MSM1418	glutamine synthetase, GlnA	
MSM1419	putative transposase	ND
MSM1420	helicase, UvrD/REP family	
MSM1421	conserved hypothetical membrane protein	
MSM1422	LemA protein	
MSM1423	exopolyphosphatase, GppA	
MSM1424	polyphosphate kinase, ppk	
MSM1425	ribosomal protein S13p	
MSM1426	ribosomal protein S4p	
MSM1427	ribosomal protein S11p	
MSM1428	DNA-directed RNA polymerase, subunit D, RpoD	
MSM1429	ribosomal protein L18e	
MSM1430	ribosomal protein L13p	
MSM1431	ribosomal protein S9p	
MSM1432	DNA-directed RNA polymerase, subunit N, RpoN	
MSM1433	DNA-directed RNA polymerase, subunit K, RpoK	
MSM1434	conserved hypothetical protein	

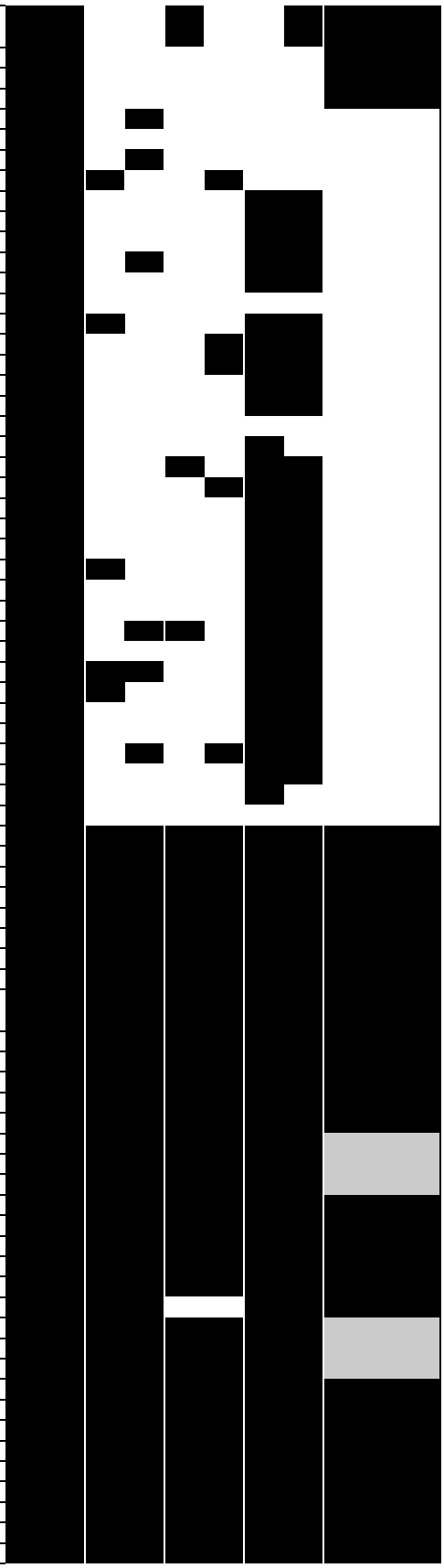
MSM1435	enolase
MSM1436	ferredoxin
MSM1437	ribosomal protein S2p
MSM1438	predicted dioxygenase
MSM1439	mevalonate kinase
MSM1440	predicted archaeal kinase
MSM1441	isopentenyl-diphosphate delta-isomerase
MSM1442	predicted RNA hydrolase, metallo-beta-lactamase superfamily
MSM1443	bifunctional short chain isoprenyl diphosphate synthase, IdsA
MSM1444	conserved hypothetical membrane protein
MSM1445	predicted transcriptional regulator
MSM1446	predicted hydroxylamine reductase, Hcp
MSM1447	conserved hypothetical protein
MSM1448	SAM-dependent methyltransferase
MSM1449	putative O-linked GlcNAc transferase
MSM1450	predicted oxidoreductase, aldo/keto reductase family
MSM1451	best blast hit to TPR repeat protein (Mba); not predicted to be a carbohydrate active enzyme by CAZy
MSM1452	glutamyl-tRNA synthetase, GltX
MSM1453	hypothetical protein
MSM1454	predicted ATPase, AAA+ family
MSM1455	aspartate/tyrosine/aromatic aminotransferase
MSM1456	conserved hypothetical protein
MSM1457	hypothetical protein
MSM1458	hypothetical protein
MSM1459	multidrug efflux permease, AraJ
MSM1460	energy-converting hydrogenase B, subunit K, EhbK
MSM1461	methyl viologen-reducing hydrogenase, delta subunit, MvhD
MSM1462	formate dehydrogenase, beta subunit, FdhB
MSM1463	formate dehydrogenase, alpha subunit, FdhA
MSM1464	FlpE-related protein
MSM1465	multidrug efflux permease, AraJ
MSM1466	hypothetical protein
MSM1467	hypothetical protein
MSM1468	adenylosuccinate synthetase, PurA
MSM1469	nitrate/sulfonate/bicarbonate ABC transporter, substrate-binding component, TauA
MSM1470	hypothetical protein
MSM1471	acyl-CoA synthetase
MSM1472	conserved hypothetical protein
MSM1473	metal-dependent hydrolase, beta-lactamase superfamily
MSM1474	chorismate synthase, AroC
MSM1475	predicted endonuclease III-related protein
MSM1476	porphobilinogen synthase, HemB
MSM1477	ATP:dephospho-CoA triphosphoribosyl transferase
MSM1478	phenylalanyl-tRNA synthetase, PheS
MSM1479	exodeoxyribonuclease, XthA
MSM1480	predicted hydrolase, HAD superfamily
MSM1481	DNA-directed DNA polymerase, family B, PolB
MSM1482	hypothetical protein
MSM1483	multidrug ABC transporter, ATPase component
MSM1484	multidrug ABC transporter, permease component
MSM1485	putative adhesin-like protein
MSM1486	ribosomal protein S8e
MSM1487	conserved hypothetical protein
MSM1488	cobalt ABC transporter, permease component, CbiM
MSM1489	protein related to formylmethanofuran dehydrogenase subunit E, metal-binding
MSM1490	conserved hypothetical protein
MSM1491	protein related to formylmethanofuran dehydrogenase subunit E, metal-binding
MSM1492	hydrogenase maturation factor, HypE
MSM1493	conserved hypothetical membrane protein, RDD family
MSM1494	hypothetical protein
MSM1495	nuclease, Staphylococcus nuclease-like family
MSM1496	conserved hypothetical protein
MSM1497	predicted coenzyme PQQ synthesis protein
MSM1498	helicase
MSM1499	predicted transcriptional regulator
MSM1500	ssDNA exonuclease, RecJ
MSM1501	signal recognition particle, subunit SRP19
MSM1502	UDP-galactopyranose mutase
MSM1503	glycosyltransferase, GT2 family
MSM1504	uroporphyrinogen III synthase, HemD
MSM1505	hypothetical protein
MSM1506	hypothetical protein

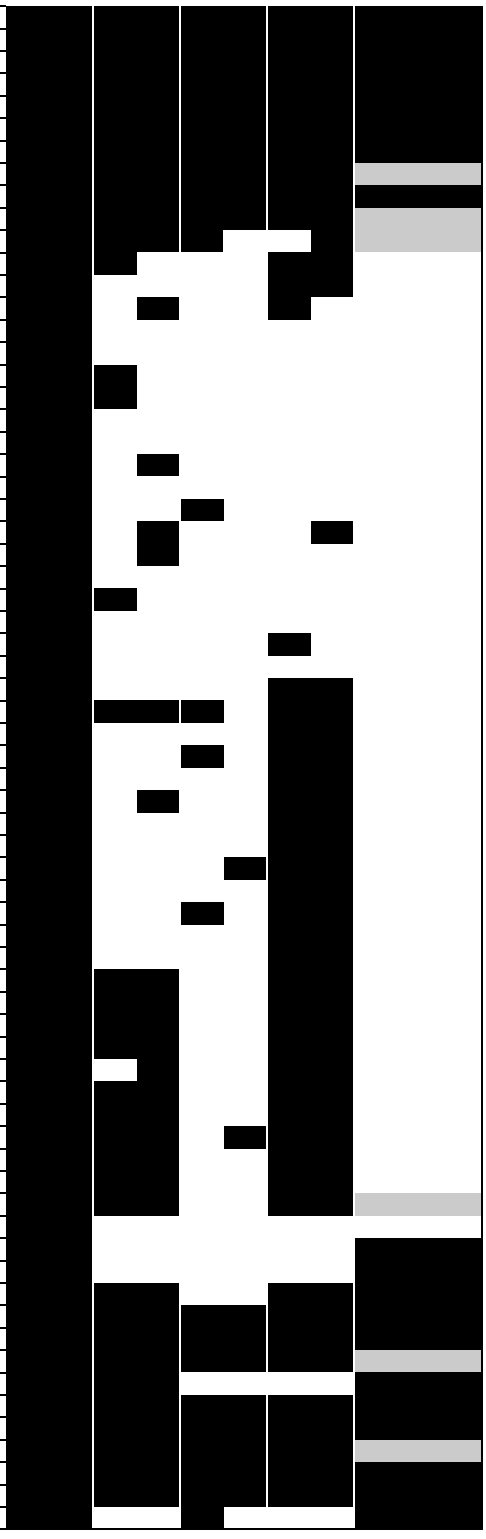


MSM1507	glycosyltransferase, GT2 family	
MSM1508	hypothetical protein	
MSM1509	hypothetical protein	
MSM1510	putative SAM-dependent methyltransferase	
MSM1511	hypothetical protein	
MSM1512	lipopolysaccharide cholinephosphotransferase, LicD family	
MSM1513	aspartate aminotransferase	
MSM1514	glycerol-3-phosphate cytidyltransferase, TagD	
MSM1515	lipopolysaccharide cholinephosphotransferase, LicD family	
MSM1516	histidinol-phosphate aminotransferase, HisC	
MSM1517	ornithine cyclodeaminase	
MSM1518	IS element ISM1 (ICSNY family)	
MSM1519	IS element ISM1 (ICSNY family)	
MSM1520	IS element ISM1 (ICSNY family)	
MSM1521	hypothetical protein	
MSM1522	hypothetical protein	
MSM1523	transposase	
MSM1524	conserved hypothetical protein	
MSM1525	conserved hypothetical protein	
MSM1526	conserved hypothetical membrane protein	
MSM1527	predicted ATPase, AAA+ superfamily	
MSM1528	predicted transcriptional regulator, HTH XRE-like family	
MSM1529	putative Zn peptidase	
MSM1530	putative nucleic acid-binding protein	
MSM1531	Na+-dependent transporter, SNF family	
MSM1532	Na+-dependent transporter, SNF family	
MSM1533	adhesin-like protein	
MSM1534	adhesin-like protein	
MSM1535	predicted dTDP-D-glucose 4,6-dehydratase	
MSM1536	pleiotropic regulatory protein DegT (PLP-dependent)	
MSM1537	predicted acylneuraminate cytidylyltransferase, NeuS	
MSM1538	CMP-sialic acid synthetase, NeuA	
MSM1539	sialic acid synthase, NeuB	
MSM1540	glycerol-3-phosphate dehydrogenase (NAD)	
MSM1541	hypothetical protein	
MSM1542	4-diphosphocytidyl-2-methyl-D-erithritol synthase, IspD	
MSM1543	hypothetical protein	
MSM1544	lipopolysaccharide cholinephosphotransferase	
MSM1545	glycosyltransferase, GT2 family	
MSM1546	hypothetical protein	
MSM1547	phosphoribosylaminoimidazole-succinocarboxamide (SAICAR) synthase, PurC	
MSM1548	phosphoribosylformylglycinamidine (FGAM) synthase, PurS	
MSM1549	phosphoribosylformylglycinamidine (FGAM) synthase, PurQ	
MSM1550	uroporphyrin-III C-methyltransferase, CobA	
MSM1551	glucosamine--fructose-6-phosphate aminotransferase, GlmS	
MSM1552	hypothetical protein	
MSM1553	hypothetical protein	
MSM1554	putative adhesin-like protein	
MSM1555	SAM-dependent methyltransferase	
MSM1556	conserved hypothetical membrane protein	
MSM1557	queuine/archaeosine tRNA-ribosyltransferase	
MSM1558	SAM-dependent methyltransferase, UbiE family	
MSM1559	polysaccharide biosynthesis protein, MviN-like family	
MSM1560	polysaccharide biosynthesis protein, MviN-like family	
MSM1561	3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase	
MSM1562	acetyl-CoA acyltransferase, SCP-type thiolase family	
MSM1563	hypothetical protein	
MSM1564	predicted SAM-dependent methyltransferase	
MSM1565	cobyric acid synthase, CobQ	
MSM1566	putative transposase	ND
MSM1567	adhesin-like protein	
MSM1568	putative transcription regulator	
MSM1569	ATP-dependent protease La, LonB	
MSM1570	cell wall biosynthesis protein, MurD-like peptide ligase family	
MSM1571	hypothetical protein	
MSM1572	ADP-ribosylglycohydrolase	
MSM1573	N-acetyltransferase, GNAT family	
MSM1574	nitroreductase, NfnB	
MSM1575	hypothetical protein	
MSM1576	hypothetical protein	
MSM1577	ribose-phosphate pyrophosphokinase, PrsA	
MSM1578	hypothetical protein	
MSM1579	excinuclease ABC, subunit B, UvrB	
MSM1580	hypothetical protein	
MSM1581	excinuclease ABC, subunit A, UvrA	
MSM1582	conserved hypothetical membrane protein	

MSM1583	archaea-specific helicase	
MSM1584	predicted excinuclease ABC, C subunit, UvrC	
MSM1585	adhesin-like protein	
MSM1586	adhesin-like protein	
MSM1587	adhesin-like protein	
MSM1588	transposase	
MSM1589	transposase, RNaseH-like family	ND
MSM1590	adhesin-like protein	
MSM1591	conserved hypothetical protein	
MSM1592	polysaccharide/polyol phosphate ABC transporter, ATPase component	
MSM1593	polysaccharide/polyol phosphate ABC transporter, permease component	
MSM1594	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	
MSM1595	SAM-dependent methyltransferase, FkbM family	
MSM1596	putative transposase	ND
MSM1597	hypothetical protein	
MSM1598	SAM-dependent methyltransferase	
MSM1599	SAM-dependent methyltransferase	
MSM1600	putative acetyltransferase, trimeric LpxA-like family	
MSM1601	conserved hypothetical protein	
MSM1602	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	
MSM1603	conserved hypothetical protein	
MSM1604	UDP-glucose pyrophosphorylase, GalU	
MSM1605	hypothetical protein	
MSM1606	arylsulfatase regulator, AslB	
MSM1607	conserved hypothetical protein	
MSM1608	predicted oxidoreductase, aldo/keto reductase family	
MSM1609	molybdate ABC transporter, substrate-binding component, ModA	
MSM1610	molybdate ABC transporter, permease component, ModC	
MSM1611	molybdate ABC transporter, ATPase component, ModB	
MSM1612	predicted UDP-glucose 6-dehydrogenase	
MSM1613	predicted UDP-glucose/GDP-mannose dehydrogenase	
MSM1614	predicted transcriptional regulator	
MSM1615	deoxyhypusine synthase, Dys	
MSM1616	conserved hypothetical membrane protein	
MSM1617	orotidine-5'-phosphate decarboxylase, PyrF	
MSM1618	cobalamin biosynthesis protein M, CbiM	
MSM1619	cobalt ABC transporter, substrate-binding component, CbiN	
MSM1620	cobalt ABC transporter, permease component, CbiQ	
MSM1621	cobalt ABC transporter, ATPase component, CbiO	
MSM1622	archaeal riboflavin synthase, RibC	
MSM1623	glycosyltransferase/dolichyl-phosphate mannose synthase, GT2 family	
MSM1624	conserved hypothetical protein	
MSM1625	thiol:fumarate reductase, subunit B, TfrB	
MSM1626	predicted fumarate reductase	
MSM1627	glycosyltransferase, GT2 family	
MSM1628	conserved hypothetical protein, aldolase family	
MSM1629	IMP dehydrogenase/GMP reductase, GuaB	
MSM1630	ribosomal protein L37Ae	
MSM1631	predicted DNA-directed RNA polymerase II, subunit RPC10	
MSM1632	predicted brix-domain ribosomal biogenesis protein	
MSM1633	conserved hypothetical protein	
MSM1634	prefoldin, beta subunit	
MSM1635	conserved hypothetical protein	
MSM1636	ProFAR isomerase-related protein	
MSM1637	conserved hypothetical membrane protein	
MSM1638	conserved hypothetical membrane protein	
MSM1639	heavy metal cation (Co/Zn/Cd) efflux system protein, CzcD family	
MSM1640	DNA intergrase/recombinase, phage integrase family	
MSM1641	hypothetical protein	
MSM1642	conserved hypothetical protein	
MSM1643	hypothetical protein	
MSM1644	hypothetical protein	
MSM1645	virulence protein	
MSM1646	putative ATPase (AAA+ superfamily)	
MSM1647	hypothetical protein	
MSM1648	hypothetical protein	
MSM1649	hypothetical protein	
MSM1650	hypothetical protein	
MSM1651	hypothetical protein	
MSM1652	hypothetical protein	
MSM1653	hypothetical protein	

MSM1654	putative Gp40-related protein, ERF family single-strand annealing protein
MSM1655	hypothetical protein
MSM1656	hypothetical protein
MSM1657	conserved hypothetical protein
MSM1658	hypothetical protein
MSM1659	hypothetical protein
MSM1660	hypothetical protein
MSM1661	hypothetical protein
MSM1662	hypothetical protein
MSM1663	hypothetical protein
MSM1664	hypothetical protein
MSM1665	hypothetical protein
MSM1666	hypothetical protein
MSM1667	hypothetical protein
MSM1668	hypothetical protein
MSM1669	hypothetical protein
MSM1670	hypothetical protein
MSM1671	large terminase subunit
MSM1672	bacteriophage capsid portal protein
MSM1673	conserved hypothetical protein
MSM1674	hypothetical protein
MSM1675	putative structural protein
MSM1676	hypothetical protein
MSM1677	putative major capsid protein gp5
MSM1678	hypothetical protein
MSM1679	hypothetical protein
MSM1680	hypothetical protein
MSM1681	hypothetical protein
MSM1682	hypothetical protein
MSM1683	hypothetical protein
MSM1684	phage-related minor tail protein
MSM1685	hypothetical protein
MSM1686	hypothetical protein
MSM1687	conserved hypothetical protein
MSM1688	hypothetical protein
MSM1689	putative collagen-like protein B
MSM1690	hypothetical protein
MSM1691	putative pseudomurein endoisopeptidase, PeiW
MSM1692	hypothetical protein
MSM1693	predicted ribokinase, PfkB family
MSM1694	predicted helicase
MSM1695	excinuclease ABC, subunit C, UvrC
MSM1696	conserved hypothetical protein
MSM1697	hypothetical protein
MSM1698	methyl coenzyme M reductase system, component A2-like
MSM1699	predicted universal stress protein, UspA
MSM1700	predicted ferredoxin
MSM1701	predicted FAD-dependent dehydrogenase, geranylgeranyl reductase family
MSM1702	UDP-glucose 4-epimerase
MSM1703	conserved hypothetical protein
MSM1704	glutamine phosphoribosylpyrophosphate amidotransferase, PurF
MSM1705	predicted collagenase, peptidase family U32
MSM1706	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase
MSM1707	nitrogenase NifH subunit, NifH
MSM1708	hypothetical protein
MSM1709	adhesin-like protein
MSM1710	seryl-tRNA synthetase, SerS
MSM1711	conserved hypothetical protein
MSM1712	predicted ferritin
MSM1713	predicted regulatory protein, amino acid-binding ACT domain family
MSM1714	coenzyme F390 synthetase
MSM1715	magnesium chelatase subunit
MSM1716	adhesin-like protein
MSM1717	predicted transporter
MSM1718	predicted biopolymer transport protein
MSM1719	conserved hypothetical protein
MSM1720	DNA-directed RNA polymerase, subunit M, RpoM
MSM1721	voltage gated chloride channel protein/cation transporter, TrkA family
MSM1722	nitroreductase
MSM1723	N5,N10-methenyl-tetrahydromethanopterin cyclohydrolase, Mch
MSM1724	conserved hypothetical membrane protein
MSM1725	conserved hypothetical membrane protein
MSM1726	conserved hypothetical membrane protein
MSM1727	multimeric flavodoxin



MSM1728	hypothetical protein	
MSM1729	conserved hypothetical protein	
MSM1730	conserved hypothetical membrane protein	
MSM1731	short chain dehydrogenase/reductase	
MSM1732	conserved hypothetical protein	
MSM1733	rubrerythrin	
MSM1734	predicted thymidylate synthase, ThyA	
MSM1735	adhesin-like protein	
MSM1736	permease, xanthine/uracil/vitamin C permease family	
MSM1737	putative transcription regulator	
MSM1738	putative adhesin-like protein	
MSM1739	conserved hypothetical membrane protein	
MSM1740	O-linked GlcNAc transferase	
MSM1741	conserved hypothetical protein	
MSM1742	predicted integrase, phage integrase-like family	
MSM1743	predicted type II restriction enzyme, methylase subunit	
MSM1744	predicted type II restriction enzyme, methylase subunit	
MSM1745	predicted type II restriction enzyme, methylase subunit	
MSM1746	predicted type II restriction enzyme, methylase subunit	
MSM1747	predicted type II restriction enzyme, methylase subunit	
MSM1748	predicted type II restriction enzyme, methylase subunit	
MSM1749	conserved hypothetical protein	
MSM1750	conserved hypothetical protein	
MSM1751	conserved hypothetical protein	
MSM1752	predicted restriction endonuclease	
MSM1753	conserved hypothetical protein	
MSM1754	predicted ATP-dependent protease La, Lon	
MSM1755	purine/pyrimidine phosphoribosyl transferase	
MSM1756	Smf protein	
MSM1757	hypothetical protein	
MSM1758	hypothetical protein	
MSM1759	hypothetical protein	
MSM1760	hypothetical protein	
MSM1761	predicted ATPase involved in DNA repair	
MSM1762	hypothetical protein	
MSM1763	predicted DNA-directed RNA polymerase, subunit M, RpoM	
MSM1764	conserved hypothetical protein	
MSM1765	conserved hypothetical protein	
MSM1766	O-linked GlcNAc transferase	
MSM1767	hypothetical protein	
MSM1768	hypothetical protein	
MSM1769	conserved hypothetical membrane protein	
MSM1770	conserved hypothetical membrane protein	
MSM1771	DNA helicase, UvrD/REP helicase family	
MSM1772	conserved hypothetical protein	
MSM1773	conserved hypothetical protein	
MSM1774	hypothetical protein	
MSM1775	putative topoisomerase IV, subunit A	
MSM1776	TPR repeat protein	
MSM1777	putative transcription regulator	
MSM1778	conserved hypothetical protein	
MSM1779	conserved hypothetical protein	
MSM1780	conserved hypothetical membrane protein	
MSM1781	conserved hypothetical protein	
MSM1782	hypothetical protein	
MSM1783	hypothetical protein	
MSM1784	hypothetical protein	
MSM1785	conserved hypothetical protein	
MSM1786	O-linked GlcNAc transferase	
MSM1787	O-linked GlcNAc transferase	
MSM1788	O-linked GlcNAc transferase	
MSM1789	predicted ATPase, AAA+ superfamily	
MSM1790	predicted ATPase, AAA+ superfamily	
MSM1791	conserved hypothetical protein	
MSM1792	nicotinate phosphoribosyltransferase	
MSM1793	conserved hypothetical protein	
MSM1794	predicted tubulin-like protein	
MSM1795	predicted ATPase, AAA+ superfamily	

[1]GeneChip-based genotyping of *M. smithii* strains done in duplicate; 'present' or 'absent' calls were determined using a perfect match/mismatch (PM/MM) model in dChip (see *Methods*). Note that the term 'absent' is based on different criteria than those used for the human microbiome dataset (see footnote 2).

[2]Metagenomic datasets from the microbiomes of two healthy lean adults (Gill et al., 2006) were tested for identity to *M. smithii* PS ORFs ; ORFs with reads that matched with >95% identity are called 'present', 80-95% identity are called 'divergent', and <80% identity are called 'absent'.

[3]Probeset for *M. smithii* gene not represented on GeneChip.

Table 3. Transcriptional regulators identified in the *M. smithii* proteome

ORF	COG	ANNOTATION
MSM0026	COG1396	predicted transcriptional regulator (possible epoxidase activity)
MSM0094		predicted transcription regulator (TetR family)
MSM0155	COG2061	predicted allosteric regulator of homoserine dehydrogenase
MSM0218	COG1321	iron dependent transcriptional regulator (Fe2+-binding)
MSM0233	COG0347	nitrogen regulatory protein P-II, GlnK
MSM0255		putative transcription regulator (winged helix DNA-binding domain)
MSM0269	COG2522	predicted transcriptional regulator (lambda repressor-like)
MSM0329	COG1396	DNA binding protein, xenobiotic response element family
MSM0354	COG1222	ATP-dependent 26S proteasome regulatory subunit, RPT1
MSM0364	COG0864	transcriptional regulator (nickel-responsive), NikR
MSM0383	COG1409	predicted phosphohydrolase, calcineurin-like superfamily
MSM0388	COG4747	amino acid regulator (ACT domain)
MSM0404	COG4742	predicted transcriptional regulator
MSM0413	COG1846	transcriptional regulator, MarR family
MSM0417	COG4068	predicted transmembrane protein with a zinc ribbon DNA-binding domain
MSM0452		predicted DNA-binding protein
MSM0453	COG1395	predicted transcriptional regulator
MSM0540	COG2865	predicted transcriptional regulator
MSM0564	COG0704	phosphate uptake regulator, PhoU
MSM0569	COG0704	phosphate transport system regulator related protein, PhoU
MSM0600	COG1846	transcriptional regulator, MarR family
MSM0635	COG2150	predicted regulator of amino acid metabolism
MSM0650	COG1309	transcriptional regulator, TetR/AcrR family
MSM0766	COG0340	biotin-[acetyl-CoA-carboxylase] ligase/biotin operon regulator bifunctional protein, BirA
MSM0775	COG2207	transcriptional regulator, AraC family
MSM0817	COG4742	predicted transcriptional regulator
MSM0818	COG4742	predicted transcriptional regulator
MSM0819	COG0640	putative transcription regulator, ArsR family (winged helix DNA-binding domain)
MSM0851	COG1548	predicted transcriptional regulator
MSM0862	COG1781	aspartate carbamoyltransferase regulatory chain, PyrI
MSM0864	COG1733	predicted transcriptional regulator
MSM0936	COG0603	transcription regulator-related ATPase, ExsB
MSM0966	COG1223	predicted 26S protease regulatory subunit (ATP-dependent), AAA+ family ATPase
MSM1030	COG0399	predicted pyridoxal phosphate-dependent enzyme
MSM1032	COG1522	transcriptional regulator, Lrp family
MSM1081	COG1112	transcriptional regulator, DNA2/NAM7 helicase family
MSM1090	COG1489	sugar fermentation stimulation protein, SfsA
MSM1106	COG0068	hydrogenase maturation factor, HypF
MSM1107	COG1777	predicted transcriptional regulator
MSM1126	COG0640	predicted transcriptional regulator, ArsR family (arsenic)
MSM1150	COG1476	predicted transcriptional regulator
MSM1207	COG2005	molybdate transport system regulatory protein
MSM1224	COG0440	acetolactate synthase, small subunit (regulatory), IlvH
MSM1230	COG1846	transcriptional regulator, MarR family
MSM1250	COG1695	predicted transcriptional regulator, PadR-like family
MSM1257	COG1339	predicted transcriptional regulator of riboflavin/FAD biosynthetic operon
MSM1292	COG2183	transcriptional accessory protein, S1 RNA binding family, Tex
MSM1315	COG2865	predicted transcriptional regulator
MSM1350	COG0640	predicted transcriptional regulator, ArsR family
MSM1390	COG0583	transcriptional regulator, LysR family
MSM1445	COG1378	predicted transcriptional regulator
MSM1499	COG1497	predicted transcriptional regulator
MSM1528	COG1396	predicted transcriptional regulator, HTH XRE-like family (xenobiotic)
MSM1536	COG0399	pleiotropic regulatory protein DegT (PLP-dependent)
MSM1568		putative transcription regulator
MSM1606	COG0641	arylsulfatase regulator, AslB
MSM1614	COG2524	predicted transcriptional regulator
MSM1713	COG4747	predicted regulatory protein, amino acid-binding ACT domain family
MSM1737		putative transcription regulator
MSM1777		putative transcription regulator

Table 4. Machinery for genome evolution in *M. smithii*

	ORF	ANNOTATION
Restriction-Modification System Subunits	MSM0157	predicted type I restriction-modification enzyme, subunit S
	MSM0158	type I restriction-modification system methylase, subunit S
	MSM1187	predicted type III restriction enzyme
	MSM1217	type II restriction endonuclease
	MSM1743	predicted type II restriction enzyme, methylase subunit
	MSM1744	predicted type II restriction enzyme, methylase subunit
	MSM1745	predicted type II restriction enzyme, methylase subunit
	MSM1746	predicted type II restriction enzyme, methylase subunit
	MSM1747	predicted type II restriction enzyme, methylase subunit
	MSM1748	predicted type II restriction enzyme, methylase subunit
	MSM1752	predicted restriction endonuclease
Recombination/ Repair	MSM0023	uncharacterized protein predicted to be involved in DNA repair
	MSM0097	Mg-dependent DNase, TatD
	MSM0120	purine NTPase involved in DNA repair, Rad50
	MSM0121	DNA repair exonuclease (SbcD/Mre11-family), Rad32
	MSM0163	conserved hypothetical proetin predicted to be involved in DNA repair
	MSM0164	conserved hypothetical protein predicted to be involved in DNA repair
	MSM0167	conserved hypothetical protein predicted to be involved in DNA repair
	MSM0168	conserved hypothetical protein predicted to be involved in DNA repair
	MSM0170	conserved hypothetical protein predicted to be involved in DNA repair
	MSM0405	predicted metal-dependent DNase, TatD-related family
	MSM0416	Mg-dependent DNase, TatD-related
	MSM0524	DNA mismatch repair ATPase, MutS
	MSM0543	DNA repair photolyase, SplB
	MSM0611	DNA repair protein, RadB
	MSM0693	ATPase involved in DNA repair, SbcC
	MSM0695	DNA repair helicase
	MSM0725	DNA repair flap structure-specific 5'-3' endonuclease
	MSM1193	single-stranded DNA-specific exonuclease, DHH family
	MSM1333	DNA repair protein RadA, RadA
	MSM1500	ssDNA exonuclease, RecJ
	MSM1640	DNA intergrase/recombinase, phage integrase family
	MSM1761	predicted ATPase involved in DNA repair
IS elements	MSM0527	IS element ISM1 (ICSNY family)
	MSM0528	IS element ISM1 (ICSNY family)
	MSM0532	IS element ISM1 (ICSNY family)
	MSM0533	IS element ISM1 (ICSNY family)
	MSM0534	IS element ISM1 (ICSNY family)
	MSM1518	IS element ISM1 (ICSNY family)
	MSM1519	IS element ISM1 (ICSNY family)
	MSM1520	IS element ISM1 (ICSNY family)
Transposases or remnants of transposases	MSM0008	putative transposase
	MSM0087	putative transposase
	MSM0110	predicted transposase
	MSM0230	putative transposase
	MSM0256	putative transposase
	MSM0342	putative transposase
	MSM0396	putative transposase
	MSM0458	transposase, homeodomain-like superfamily
	MSM0460	predicted transposase
	MSM0601	putative transposase
	MSM0629	putative transposase
	MSM0730	putative transposase
	MSM0871	putative transposase
	MSM1093	putative transposase
	MSM1115	putative transposase
	MSM1189	putative transposase
	MSM1419	putative transposase
	MSM1523	transposase
	MSM1566	putative transposase
	MSM1588	predicted transposase
	MSM1589	predicted transposase, RNaseH-like family
	MSM1596	putative transposase

Table 5. Publicly available finished genome sequences for members of Archaea

Group	Strain Designation	Abbreviation	Temperature	Habitat of Origin	GenBank Accession Number
Human Gut Methanogens	*Methanobrevibacter smithii* PS (ATCC 35021)	Msm	Mesophilic	Host-associated	CP000678
	Methanosphaera stadtmanae DSM 3091	Msp	Mesophilic	Host-associated	CP000102
Non-Gut Methanogens	*Methanothermobacter thermautotrophicus* Delta H	Mth	Thermophilic	Specialized	AE000666
	Methanocaldococcus jannaschii DSM 2661	Mja	Hyperthermophilic	Aquatic	L77117
	Methanococcoides burtonii DSM 6242	Mbu	Mesophilic	Aquatic	CP000300
	Methanococcus maripaludis S2	Mmr	Mesophilic	Aquatic	BX950229
	Methanopyrus kandleri AV19	Mka	Hyperthermophilic	Specialized	AE009439
	Methanosarcina acetivorans C2A	Mac	Mesophilic	Aquatic	AE010299
	Methanosarcina barkeri str. Fusaro	Mba	Mesophilic	Multiple	CP000099
	Methanosarcina mazei Go1	Mma	Mesophilic	Multiple	AE008384
	Methanospirillum hungatei JF-1	Mhu	Mesophilic	Multiple	CP000254
Other Archaea	*Aeropyrum pernix* K1	Apx	Hyperthermophilic	Specialized	BA000002
	Archaeoglobus fulgidus DSM 4304	Afu	Hyperthermophilic	Aquatic	AE000782
	Haloarcula marismortui ATCC 43049	Hma	Mesophilic	Aquatic	AY596297
	Halobacterium sp. NRC-1	Hal	Mesophilic	Specialized	AE004437
	Nanoarchaeum equitans Kin4-M	Neq	Hyperthermophilic	Host-associated	AE017199
	Natronomonas pharaonis DSM 2160	Nph	Mesophilic	Aquatic	CR936257
	Picrophilus torridus DSM 9790	Pto	Thermophilic	Specialized	AE017261
	Pyrobaculum aerophilum str. IM2	Pae	Hyperthermophilic	Aquatic	AE009441
	Pyrococcus abyssi GE5	Pab	Hyperthermophilic	Aquatic	AL096836
	Pyrococcus furiosus DSM 3638	Pfu	Hyperthermophilic	Aquatic	AE009950
	Pyrococcus horikoshii OT3	Pho	Hyperthermophilic	Aquatic	BA000001
	Sulfolobus acidocaldarius DSM 639	Sac	Thermophilic	Specialized	CP000077
	Sulfolobus solfataricus P2	Sso	Hyperthermophilic	Specialized	AE006641
	Sulfolobus tokodaii str. 7	Sto	Hyperthermophilic	Specialized	BA000023
	Thermococcus kodakarensis KOD1	Tko	Hyperthermophilic	Specialized	AP006878
	Thermoplasma acidophilum DSM 1728	Tac	Thermophilic	Specialized	AL139299
	Thermoplasma volcanium GSS1	Tvo	Thermophilic	Specialized	BA000011

Table 6. Representation of enriched gene ontology (GO) categories in the *M. smithii* and *M. stadtmanae* proteomes compared to the proteomes of all sequenced methanogenic archaea and all archaea

			M. smithii			*M. stadtmanae*		
Category	**GO**	**Description**	**No.**	**Arch**	**Meth**	**No.**	**Arch**	**Meth**
SURFACE VARIATION	GO:0007047	cell wall organization and biogenesis	**17**			**10**		
	GO:0042545	cell wall modification	**9**			**6**		
	GO:0030599	pectinesterase activity	**9**			**6**		
	GO:0042546	cell wall biosynthesis	**8**			**4**		
	GO:0009252	peptidoglycan biosynthesis	**5**			**4**		
CENTRAL METABOLISM	GO:0006082	organic acid metabolism	**109**			**105**		
	GO:0006730	one-carbon compound metabolism	**27**			**22**		
	GO:0045333	cellular respiration	**20**			**14**		
	GO:0015948	methanogenesis	**19**			**13**		
	GO:0019321	pentose metabolism	**5**			**1**		
COFACTOR/VITAMIN METABOLISM	GO:0046872	metal ion binding	**120**			**97**		
	GO:0005506	iron ion binding	**77**			**61**		
	GO:0050661	NADP binding	**11**			**4**		
	GO:0006767	water-soluble vitamin metabolism	**49**			**44**		
	GO:0042727	riboflavin and derivative biosynthesis	**11**			**4**		
	GO:0008703	5-amino-6-(5-phosphoribosylamino)uracil reductase activity	**8**			**1**		
AMINO ACID METABOLISM	GO:0006519	amino acid and derivative metabolism	**98**			**99**		
	GO:0008652	amino acid biosynthesis	**60**			**58**		
ENVIRONMENTAL SENSING	GO:0004672	protein kinase activity	**4**			**4**		
	GO:0007165	signal transduction	**3**			**0**		
	GO:0000160	two-component signal transduction system (phosphorelay)	**1**			**0**		
TRANSCRIPTIONAL CONTROL	GO:0006350	transcription	**45**			**35**		
	GO:0019222	regulation of metabolism	**36**			**29**		
	GO:0045449	regulation of transcription	**32**			**24**		
NUTRIENT TRANSPORT	GO:0046943	carboxylic acid transporter activity	**1**			**3**		
	GO:0015849	organic acid transport	**0**			**2**		
	GO:0005275	amine transporter activity	**0**			**0**		
PHOSPHATE METABOLISM	GO:0006796	phosphate metabolism	**24**			**23**		
RECOMBINATION	GO:0006310	DNA recombination	**8**			**7**		

P-value	>1E-2	<1E-2	<1E-3	<1E-4
Enrichment				
Depletion				

Abbreviations: 'non-gut-associated methanogens' (Meth) or 'all Archaea' (Arch) [see SI **Table 5**]; No., number of genes associated with gene ontology (GO) term.

Table 7. *M. smithii* genes in the significantly enriched GO categories listed in Table 6

■ - member of GO category
□ - not a member of GO category

ENRICHED COMPARED TO SEQUENCED ARCHAEA AND/OR NON-GUT METHANOGENS							
SURFACE VARIATION	***M. smithii* genes**	**Annotation**	**GO:0007047 - cell wall organization and biogenesis**	**GO:0042545 - cell wall modification**	**GO:0030599 - pectinesterase activity**	**GO:0042546 - cell wall biosynthesis**	**GO:0009252 - peptidoglycan biosynthesis**
	MSM0052	adhesin-like protein	■	■	■	□	□
	MSM0118	cell wall biosynthesis protein, MurD-like peptide ligase family	■	□	□	■	■
	MSM0266	adhesin-like protein	■	■	■	□	□
	MSM0359	cell wall biosynthesis protein, MurD-like peptide ligase family	■	□	□	■	■
	MSM0360	cell wall biosynthesis protein, phospho-N-acetylmuramoyl-pentapeptide-transferase family	■	□	□	■	■
	MSM1111	adhesin-like protein	■	■	■	□	□
	MSM1112	adhesin-like protein	■	■	■	□	□
	MSM1113	adhesin-like protein	■	■	■	□	□
	MSM1190	cell wall biosynthesis protein, UDP-N-acetylmuramate-alanine ligase family	■	□	□	■	■
	MSM1191	cell wall biosynthesis protein, MurD-like peptide ligase family	■	□	□	■	■
	MSM1327	predicted CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase	■	□	□	■	□
	MSM1585	adhesin-like protein	■	■	■	□	□
	MSM1586	adhesin-like protein	■	■	■	□	□
	MSM1587	adhesin-like protein	■	■	■	□	□
	MSM1590	adhesin-like protein	■	■	■	□	□
	MSM1594	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	■	□	□	■	□
	MSM1602	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	■	□	□	■	□
CENTRAL METABOLISM	***M. smithii* genes**	**Annotation**	**GO:0006082 - organic acid metabolism**	**GO:0006730 - one-carbon compound metabolism**	**GO:0045333 - cellular respiration**	**GO:0015948 - methanogenesis**	**GO:0019321 - pentose metabolism**
	MSM0027	glutamate synthase, domain 2 with rubredoxin	■	□	□	□	□
	MSM0071	methionyl-tRNA synthetase, MetG	■	□	□	□	□
	MSM0082	heterodisulfide reductase, subunit A, HdrA	□	■	■	■	□
	MSM0089	pyrroline-5-carboxylate reductase (NADP oxidoreductase, coenzyme F420-dependent), ProC	■	□	□	□	□
	MSM0102	cobalamin-independent methionine synthase, MetE	■	□	□	□	□
	MSM0154	homoserine dehydrogenase, ThrA	■	□	□	□	□

MSM0160	asparagine synthetase, AsnB	■				
MSM0174	O-acetylhomoserine sulfhydrylase (PLP-dependent), MET17	■				
MSM0175	homoserine O-acetyltransferase, MetX	■				
MSM0214	threonine synthase (pyridoxal-phosphate dependent), ThrC	■				
MSM0216	tryptophanyl-tRNA synthetase, TrpS	■				
MSM0231	3-dehydroquinate dehydratase	■				
MSM0265	O-acetylhomoserine sulfhydrylase	■				
MSM0268	cysteinyl-tRNA synthetase, CysS	■				
MSM0270	serine acetyltransferase, CysE	■				
MSM0271	cysteine synthase, CysK	■				
MSM0273	EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase)	■				
MSM0275	valyl-tRNA synthetase, ValS	■				
MSM0277	phenylalanyl-tRNA synthetase, beta subunit, PheT	■				
MSM0286	glycerol 1-phosphate dehydrogenase (Dehydroquinate synthase-like family)	■				
MSM0287	prolyl-tRNA synthetase, ProS	■				
MSM0307	sugar kinase (ribokinase/pfkB superfamily)					■
MSM0308	formylmethanofuran:tetrahydromethanopterin formyltransferase, FtrC		■			
MSM0334	L-asparaginase, GatD,	■				
MSM0343	GMP synthase (glutamine-hydrolysing), subunit A, GuaA	■				
MSM0350	2-isopropylmalate synthase, LeuA	■				
MSM0368	glutamate synthase (NADPH), subunit 2	■				
MSM0371	predicted glutamine amidotransferase involved in pyridoxine biosynthesis, Pdx2	■				
MSM0373	isocitrate/isopropylmalate dehydrogenase, LeuB	■				
MSM0375	acetylglutamate kinase, ArgB	■				
MSM0379	glutamate N-acetyltransferase, ArgJ	■				
MSM0388	amino acid regulator	■				
MSM0393	fumarate reductase, iron-sulfur protein			■		
MSM0403	glycyl-tRNA synthetase	■				
MSM0415	uridylate kinase, PyrH	■				
MSM0457	D-3-phosphoglycerate dehydrogenase, SerA	■				
MSM0488	carbamoylphosphate synthase, large subunit, CarB	■				
MSM0489	carbamoylphosphate synthase, small subunit, CarA	■				
MSM0513	tyrosyl-tRNA synthetase, TyrS	■				
MSM0516	corrinoid protein (methionine synthase-related), MtaC	■				
MSM0518	methylcobalamin:coenzyme M methyltransferase, MtaA		■			
MSM0556	dihydropteroate synthase	■				
MSM0572	H(2)-forming N5,N10-methylenetetrahydromethanopterin dehydrogenase (coenzyme F420-dependent), Mth		■	■	■	
MSM0573	biotin synthetase, BioB	■				
MSM0604	predicted archaeal aspartate/glutamate/uridylate kinase	■				
MSM0619	alanyl-tRNA synthetase, AlaS	■				
MSM0627	tetrahydromethanopterin S-methyltransferase, subunit H, MtrH		■			
MSM0641	prephenate dehydrogenase (NADP+)	■				

MSM0653	histidinol-phosphate aminotransferase, HisC	■				
MSM0719	phosphoserine phosphatase, HAD family, SerB	■				
MSM0722	2-isopropylmalate synthase, LeuA	■				
MSM0723	3-isopropylmalate dehydratase, LeuC	■				
MSM0727	S-adenosylhomocysteine hydrolase (adenosylhomocysteinase), AhcY		■			
MSM0829	aspartate-semialdehyde dehydrogenase, Asd	■				
MSM0830	dihydrodipicolinate reductas, DapB	■				
MSM0832	aspartokinase, alpha subunit	■				
MSM0834	chorismate mutase	■				
MSM0835	archaeal shikimate kinase	■				
MSM0847	archaeal 3-isopropylmalate dehydratase, small subunit, LeuD	■				
MSM0858	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide (ProFAR) isomerase, HisA	■				
MSM0860	aspartate-semialdehyde dehydrogenase, ArgC	■				
MSM0878	pyruvoyl-dependent arginine decarboxylase, PdaD	■				
MSM0888	glutamate dehydrogenase (NADP+), GdhA	■				
MSM0902	methyl-coenzyme M reductase, alpha subunit, McrA		■	■	■	
MSM0903	methyl-coenzyme M reductase, gamma subunit, McrG		■	■	■	
MSM0904	methyl-coenzyme M reductase, D subunit, McrD		■	■	■	
MSM0905	methyl-coenzyme M reductase, beta subunit, McrB		■	■	■	
MSM0939	pyruvate carboxylase, subunit B, PycB	■				
MSM0965	3-hydroxyacyl-CoA dehydrogenase, FadB	■				
MSM0967	glutamyl-tRNA reductase, HemA	■				
MSM0987	tyrosine decarboxylase, MfnA	■				
MSM0988	phosphoenolpyruvate synthase, PpsA	■				
MSM1001	methyl viologen-reducing hydrogenase, delta subunit, MvhD		■	■	■	
MSM1007	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H, MtrH		■			
MSM1008	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit G, MtrG		■	■	■	
MSM1011	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit B, MtrB		■	■	■	
MSM1012	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit C, MtrC		■	■	■	
MSM1015	methyl-coenzyme M reductase, alpha subunit, McrA		■	■	■	
MSM1016	methyl-coenzyme M reductase, gamma subunit, McrG		■	■	■	
MSM1017	methyl-coenzyme M reductase, C subunit, McrC		■	■	■	
MSM1018	methyl-coenzyme M reductase, D subunit, McrD		■	■	■	
MSM1019	methyl-coenzyme M reductase, beta subunit, McrB		■	■	■	
MSM1052	prephenate dehydratase, PheA	■				
MSM1084	argininosuccinate synthase, ArgG	■				

MSM1092	formylmethanofuran:tetrahydromethanopterin formyltransferase, Ftr		■			
MSM1103	phosphoribosyl-ATP pyrophosphohydrolase, HisE	■				
MSM1141	tryptophan synthase, alpha subunit, TrpA	■				
MSM1142	tryptophan synthase, beta subunit, TrpB	■				
MSM1143	indole-3-glycerol phosphate synthase, TrpC	■				
MSM1144	anthranilate phosphoribosyltransferase, TrpD	■				
MSM1145	anthranilate/para-aminobenzoate synthase component II, TrpG	■				
MSM1159	glutamine amidotransferase, HisH	■				
MSM1172	leucyl-tRNA synthetase, LeuS	■				
MSM1179	shikimate 5-dehydrogenase, AroE	■				
MSM1181	histidyl-tRNA synthetase, HisS	■				
MSM1182	phosphoribosyl-AMP cyclohydrolase, HisI	■				
MSM1202	branched-chain-amino-acid aminotransferase, IlvE	■				
MSM1204	coenzyme F420-dependent methylenetetrahydromethanopterin dehydrogenase, Mtd		■	■	■	
MSM1206	imidazoleglycerol-phosphate dehydrogenase, HisB	■				
MSM1214	threonyl-tRNA synthetase, ThrS	■				
MSM1222	ketol-acid reductoisomerase, IlvC	■				
MSM1224	acetolactate synthase, small subunit (regulatory), IlvH	■				
MSM1226	ornithine carbamoyltransferase, ArgF	■				
MSM1231	arginyl-tRNA synthetase, ArgS	■				
MSM1236	aspartyl-tRNA synthetase, AspS	■				
MSM1237	dihydroxy-acid dehydratase, IlvD	■				
MSM1238	histidinol dehydrogenase, HisD	■				
MSM1242	tryptophan synthase, beta subunit, TrpB	■				
MSM1246	isopropylmalate synthase, LeuA	■				
MSM1261	ATP phosphoribosyltransferase, HisG	■				
MSM1263	aspartate carbamoyltransferase, PyrB	■				
MSM1298	3-isopropylmalate dehydrogenase, LeuB	■				
MSM1299	3-isopropylmalate dehydratase, small subunit, LeuD	■				
MSM1300	3-isopropylmalate dehydratase, large subunit, LeuC	■				
MSM1327	predicted CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase	■				■
MSM1337	glycine hydroxymethyltransferase, GlyA	■				
MSM1340	archaeal S-adenosylmethionine synthetase, MetK		■			
MSM1341	isoleucyl-tRNA synthetase, IleS	■				
MSM1364	imidazoleglycerol-phosphate synthase, HisF	■				
MSM1368	N-acetylornithine aminotransferase, ArgD	■				
MSM1371	diaminopimelate decarboxylase, LysA	■				
MSM1372	diaminopimelate epimerase, DapF	■				
MSM1387	lysyl-tRNA synthetase (class I), LysS	■				
MSM1389	sugar kinase, ribokinase/pfkB superfamily					■
MSM1396	tungsten formylmethanofuran dehydrogenase, subunit E, FwdE		■	■	■	

	MSM1404	formate dehydrogenase, alpha subunit, FdhA	■	■				
	MSM1414	tungsten formylmethanofuran dehydrogenase, subunit C, FwdC		■	■	■		
	MSM1418	glutamine synthetase, GlnA	■					
	MSM1440	predicted archaeal kinase	■					
	MSM1452	glutamyl-tRNA synthetase, GltX	■					
	MSM1461	methyl viologen-reducing hydrogenase, delta subunit, MvhD		■	■	■		
	MSM1474	chorismate synthase, AroC	■					
	MSM1478	phenylalanyl-tRNA synthetase, PheS	■					
	MSM1594	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	■				■	
	MSM1602	glycosyltransferase/CDP-glycerol:poly(glycerophosphate) glycerophosphotransferase, GT2 family	■				■	
	MSM1636	ProFAR isomerase-related protein	■					
	MSM1710	seryl-tRNA synthetase, SerS	■					
	MSM1713	predicted regulatory protein, amino acid-binding ACT domain family	■					
COFACTOR/ VITAMIN METABOLISM	***M. smithii* genes**	**Annotation**	**GO:0046872 - metal ion binding**	**GO:0005506 - iron ion binding**	**GO:0050661 - NADP binding**	**GO:0006767 - water-soluble vitamin metabolism**	**GO:0042727 - riboflavin and derivative biosynthesis**	**GO:0008703 - 5-amino-6-(5-phosphoribosylamino)uracil reductase**
	MSM0027	glutamate synthase, domain 2 with rubredoxin	■					
	MSM0029	putative calcium-binding protein	■					
	MSM0052	adhesin-like protein			■	■	■	■
	MSM0065	riboflavin-specific deaminase			■	■	■	■
	MSM0079	CofH protein (7,8-didemethyl-8-hydroxy-5-deazariboflavin (FO)/F420 biosynthesis)	■	■				
	MSM0082	heterodisulfide reductase, subunit A, HdrA	■	■				
	MSM0084	heterodisulfide reductase, subunit C, HdrC	■	■				
	MSM0085	putative ferredoxin	■	■				
	MSM0101	precorrin-3 methylase, CbiF				■		
	MSM0107	hydrogenase expression/formation protein, HypB	■					
	MSM0108	hydrogenase nickel incorporation protein, HypA	■					
	MSM0129	nicotinamide mononucleotide adenylyltransferase, NadR				■		
	MSM0131	molybdenum-binding protein, MopI	■					
	MSM0135	coenzyme F420-reducing hydrogenase, beta subunit	■	■				
	MSM0136	putative ferredoxin	■	■				
	MSM0148	predicted oxidoreductase, aldo/keto reductase family	■	■				
	MSM0153	predicted phosphoglycerate mutase	■					
	MSM0187	rubredoxin	■					
	MSM0188	rubredoxin	■					
	MSM0198	inorganic pyrophosphatase	■					
	MSM0203	nucleoside diphosphate kinase, Ndk	■					
	MSM0209	ferredoxin	■	■				
	MSM0218	iron dependent transcriptional regulator (Fe2+-binding)	■	■				
	MSM0238	precorrin-6B methylase, CbiT				■		

MSM0262	desulfoferrodoxin (dfx)							
MSM0262	desulfoferrodoxin (dfx)							
MSM0266	adhesin-like protein							
MSM0272	endonuclease III							
MSM0284	ribose 5-phosphate isomerase, RpiA							
MSM0289	phosphomethylpyrimidine kinase (HMPP-kinase), ThiD							
MSM0306	polyferredoxin, iron-sulfur binding							
MSM0310	polyferredoxin, iron-sulfur binding							
MSM0311	polyferredoxin, iron-sulfur binding							
MSM0312	[NiFe]-hydrogenase-3-type complex, large subunit/NADH:quinone oxidoreductase (complex I), subunit 49K/NdhH/NuoD							
MSM0331	2-oxoisovalerate ferredoxin oxidoreductase, delta subunit							
MSM0357	conserved hypothetical membrane protein (possible Zinc-binding)							
MSM0368	glutamate synthase (NADPH), subunit 2							
MSM0376	alcohol dehydrogenase (zinc-binding), GroES-like							
MSM0385	anaerobic magnesium-protoporphyrin IX monomethyl ester cyclase, Elongator protein 3/MiaB/NifB family							
MSM0392	indolepyruvate ferredoxin oxidoreductase, alpha subunit							
MSM0393	fumarate reductase, iron-sulfur protein							
MSM0395	ferredoxin, iron-sulfur binding							
MSM0409	C4-type Zinc-finger protein							
MSM0424	transcription initiation factor TFIIB (zinc-binding)							
MSM0491	nicotinate-nucleotide pyrophosphorylase (carboxylating), NadC							
MSM0494	quinolinate synthetase, subunit A, NadA							
MSM0516	corrinoid protein (methionine synthase-related), MtaC							
MSM0517	methyltransferase activation protein, MapA							
MSM0517	methyltransferase activation protein, MapA							
MSM0538	pyruvate formate-lyase activating enzyme, PflA							
MSM0543	DNA repair photolyase, SplB							
MSM0544	predicted Fe-S oxidoreductase							
MSM0558	pyruvate:ferredoxin oxidoreductase, delta subunit, PorD							
MSM0561	formate dehydrogenase, iron-sulfur subunit							
MSM0562	formate dehydrogenase, iron-sulfur subunit							
MSM0573	biotin synthetase, BioB							
MSM0579	polyferredoxin, iron-sulfur binding							
MSM0583	cobalt ABC transporter, permease component, CbiM							
MSM0585	cobalt ABC transporter, permease component, CbiQ							
MSM0607	predicted ATPase, RNase L inhibitor family							
MSM0607	predicted ATPase, RNase L inhibitor family							
MSM0609	ferredoxin, iron-sulfur binding							
MSM0616	adhesin-like protein							

MSM0617	thiamine biosynthesis ATP pyrophosphatase, ThiI				■		
MSM0644	thiamine biosynthesis protein, ThiC				■		
MSM0652	pyruvate formate-lyase activating enzyme, PflA	■	■				
MSM0657	phosphoglycerate mutase, AP superfamily	■					
MSM0696	Fe-S oxidoreductase	■					
MSM0723	3-isopropylmalate dehydratase, LeuC	■	■				
MSM0723	3-isopropylmalate dehydratase, LeuC						
MSM0728	predicted oxidoreductase, aldo/keto reductase family	■	■				
MSM0771	cobalt ABC transporter, permease component, CbiQ				■		
MSM0783	tungsten formylmethanofuran dehydrogenase, subunit F, FwdF	■	■				
MSM0784	ferredoxin	■	■				
MSM0787	Fe-S oxidoreductase	■	■				
MSM0796	heterodisulfide reductase, subunit C, HdrC	■	■				
MSM0808	nitrogenase iron-molybdenum cofactor biosynthesis protein, NifB	■	■				
MSM0829	aspartate-semialdehyde dehydrogenase, Asd			■			
MSM0837	cobalamin biosynthesis protein D, CbiD				■		
MSM0842	histone acetyltransferase, radical SAM superfamily	■	■				
MSM0845	2-methylthioadenine synthetase, MiaB	■	■				
MSM0849	molybdenum cofactor biosynthesis-related protein, MoaA	■	■				
MSM0865	putative glucose-methanol-choline oxidoreductase (FAD-dependent)	■	■				
MSM0892	putative zinc-binding protein	■					
MSM0895	cation transport ATPase, HAD family	■					
MSM0896	precorrin-6X reductase, CbiJ				■		
MSM0916	hydroxyethylthiazole kinase, ThiM				■		
MSM0917	thiamine monophosphate synthase, ThiE	■			■		
MSM0922	Fe-S oxidoreductase	■	■				
MSM0928	2-oxoglutarate ferredoxin oxidoreductase, delta subunit, KorD	■	■				
MSM0933	cobalamin-5-phosphate synthase, CobS				■		
MSM0960	cation transport ATPase, HAD family	■					
MSM0961	heavy-metal cation transporting ATPase	■					
MSM0962	glyceraldehyde 3-phosphate dehydrogenase, GapA			■			
MSM0998	polyferredoxin, MvhB	■	■				
MSM0999	methyl viologen-reducing hydrogenase, alpha subunit, MvhA	■					
MSM1020	Fe-S oxidoreductase, Radical SAM family	■	■				
MSM1035	FO synthase subunit 1 (SAM-dependent), CofG (F420 biosynthesis)	■	■				
MSM1043	dihydroorotate dehydrogenase electron transfer subunit, PyrK	■	■				
MSM1055	coenzyme PQQ synthesis protein, SAM family	■	■				
MSM1063	energy-converting hydrogenase B, subunit N, EhbN	■					
MSM1065	energy-converting hydrogenase B, subunit L, EhbL	■	■				
MSM1066	energy-converting hydrogenase B, subunit K, EhbK	■	■				

MSM1099	biotin synthase related protein						
MSM1106	hydrogenase maturation factor, HypF						
MSM1111	adhesin-like protein						
MSM1112	adhesin-like protein						
MSM1122	coenzyme F420-reducing hydrogenase, gamma subunit, FrhG						
MSM1123	coenzyme F420-reducing hydrogenase, delta subunit, FrhD						
MSM1124	coenzyme F420-reducing hydrogenase, alpha subunit, FrhA						
MSM1127	cation transport ATPase, HAD family						
MSM1138	predicted glutamine amidotransferase, CobB/CobQ-like family						
MSM1153	cation transport ATPase, HAD family						
MSM1167	cobalt precorrin-6Y C5,15-methyltransferase, CbiE						
MSM1171	ammonia-dependent NAD+ synthetase, NadE						
MSM1174	heat shock protein HtpX (Zn-dependent)						
MSM1179	shikimate 5-dehydrogenase, AroE						
MSM1198	O-sialoglycoprotein endopeptidase						
MSM1200	phosphoribosyltransferase, CobT						
MSM1215	cobyrinic acid a,c-diamide synthase, CbiA						
MSM1223	carbonic anhydrase						
MSM1230	transcriptional regulator, MarR family						
MSM1234	cobalt-precorrin-8X methylmutase, CbiC						
MSM1238	histidinol dehydrogenase, HisD						
MSM1239	predicted DNA-binding protein						
MSM1241	chromosome partitioning ATPase						
MSM1254	cobyric acid synthase						
MSM1256	3,4-dihydroxy-2-butanone 4-phosphate synthase, RibB						
MSM1266	cobalamin biosynthesis protein D, CobD						
MSM1267	cobalamin biosynthesis protein G, CbiG						
MSM1273	cobalt precorrin-3B C17-methyltransferase, CbiH						
MSM1283	thiamine monphosphate kinase, ThiL						
MSM1284	pyruvate formate-lyase activating enzyme, PflA						
MSM1296	riboflavin synthase, beta subunit, RibH						
MSM1300	3-isopropylmalate dehydratase, large subunit, LeuC						
MSM1301	predicted Fe-S oxidoreductase						
MSM1336	heterodisulfide reductase, subunit A, HdrA						
MSM1338	archaeal flavoprotein						
MSM1348	rubrerythrin						
MSM1351	precorrin-2 C20-methyltransferase, CbiL						
MSM1354	DNA-directed RNA polymerase subunit M, RpoM						
MSM1380	NADP-dependent alcohol dehydrogenase						
MSM1386	cytosine deaminase						
MSM1388	thiamine biosynthesis protein, ThiC						
MSM1404	formate dehydrogenase, alpha subunit, FdhA						
MSM1405	formate dehydrogenase, beta subunit, FdhB						

	MSM1406	molybdopterin cofactor biosynthesis protein A, MoaA	■	■				
	MSM1408	tungsten formylmethanofuran dehydrogenase, subunit E, FwdE	■	■				
	MSM1409	tungsten formylmethanofuran dehydrogenase, subunit F, FwdF	■	■				
	MSM1410	tungsten formylmethanofuran dehydrogenase, subunit G, FwdG	■	■				
	MSM1411	tungsten formylmethanofuran dehydrogenase, subunit D, FwdD	■					
	MSM1436	ferredoxin	■	■				
	MSM1446	predicted hydroxylamine reductase, Hcp	■	■				
	MSM1450	predicted oxidoreductase, aldo/keto reductase family	■	■				
	MSM1460	energy-converting hydrogenase B, subunit K, EhbK	■	■				
	MSM1462	formate dehydrogenase, beta subunit, FdhB	■	■				
	MSM1488	cobalt ABC transporter, permease component, CbiM				■		
	MSM1497	predicted coenzyme PQQ synthesis protein	■	■				
	MSM1565	cobyric acid synthase, CobQ				■		
	MSM1567	adhesin-like protein			■	■	■	■
	MSM1590	adhesin-like protein			■	■	■	■
	MSM1606	arylsulfatase regulator, AslB	■	■				
	MSM1608	predicted oxidoreductase, aldo/keto reductase family	■	■				
	MSM1618	cobalamin biosynthesis protein M, CbiM				■		
	MSM1619	cobalt ABC transporter, substrate-binding component, CbiN				■		
	MSM1620	cobalt ABC transporter, permease component, CbiQ				■		
	MSM1622	archaeal riboflavin synthase, RibC				■	■	
	MSM1626	predicted fumarate reductase	■	■				
	MSM1655	hypothetical protein (phage)	■					
	MSM1700	predicted ferredoxin	■	■				
	MSM1712	predicted ferritin	■	■				
	MSM1720	DNA-directed RNA polymerase, subunit M, RpoM	■					
	MSM1733	rubrerythrin	■					
	MSM1792	nicotinate phosphoribosyltransferase				■		

AMINO ACID METABOLISM	***M. smithii* genes**	**Annotation**	**GO:0006519 - amino acid and derivative metabolism**	**GO:0008652 - amino acid biosynthesis**
	MSM0027	glutamate synthase, domain 2 with rubredoxin	■	■
	MSM0071	methionyl-tRNA synthetase, MetG	■	
	MSM0089	pyrroline-5-carboxylate reductase (NADP oxidoreductase, coenzyme F420-dependent), ProC	■	■
	MSM0102	cobalamin-independent methionine synthase, MetE	■	■
	MSM0154	homoserine dehydrogenase, ThrA	■	■
	MSM0160	asparagine synthetase, AsnB	■	■
	MSM0174	O-acetylhomoserine sulfhydrylase (PLP-dependent), MET17	■	
	MSM0175	homoserine O-acetyltransferase, MetX	■	■

MSM0214	threonine synthase (pyridoxal-phosphate dependent), ThrC		
MSM0216	tryptophanyl-tRNA synthetase, TrpS		
MSM0231	3-dehydroquinate dehydratase		
MSM0265	O-acetylhomoserine sulfhydrylase		
MSM0268	cysteinyl-tRNA synthetase, CysS		
MSM0270	serine acetyltransferase, CysE		
MSM0271	cysteine synthase, CysK		
MSM0273	EPSP synthase (3-phosphoshikimate 1-carboxyvinyltransferase)		
MSM0275	valyl-tRNA synthetase, ValS		
MSM0277	phenylalanyl-tRNA synthetase, beta subunit, PheT		
MSM0286	glycerol 1-phosphate dehydrogenase (Dehydroquinate synthase-like family)		
MSM0287	prolyl-tRNA synthetase, ProS		
MSM0334	L-asparaginase, GatD		
MSM0343	GMP synthase (glutamine-hydrolysing), subunit A, GuaA		
MSM0368	glutamate synthase (NADPH), subunit 2		
MSM0371	predicted glutamine amidotransferase involved in pyridoxine biosynthesis, Pdx2		
MSM0373	isocitrate/isopropylmalate dehydrogenase, LeuB		
MSM0375	acetylglutamate kinase, ArgB		
MSM0379	glutamate N-acetyltransferase, ArgJ		
MSM0388	amino acid regulator		
MSM0403	glycyl-tRNA synthetase		
MSM0415	uridylate kinase, PyrH		
MSM0457	D-3-phosphoglycerate dehydrogenase, SerA		
MSM0488	carbamoylphosphate synthase, large subunit, CarB		
MSM0489	carbamoylphosphate synthase, small subunit, CarA		
MSM0513	tyrosyl-tRNA synthetase, TyrS		
MSM0516	corrinoid protein (methionine synthase-related), MtaC		
MSM0604	predicted archaeal aspartate/glutamate/uridylate kinase		
MSM0619	alanyl-tRNA synthetase, AlaS		
MSM0641	prephenate dehydrogenase (NADP+)		
MSM0653	histidinol-phosphate aminotransferase, HisC		
MSM0719	phosphoserine phosphatase, HAD family, SerB		
MSM0723	3-isopropylmalate dehydratase, LeuC		
MSM0829	aspartate-semialdehyde dehydrogenase, Asd		
MSM0830	dihydrodipicolinate reductas, DapB		
MSM0832	aspartokinase, alpha subunit		
MSM0834	chorismate mutase		
MSM0835	archaeal shikimate kinase		
MSM0847	archaeal 3-isopropylmalate dehydratase, small subunit, LeuD		
MSM0858	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide (ProFAR) isomerase, HisA		
MSM0860	aspartate-semialdehyde dehydrogenase, ArgC		
MSM0876	arginase/agmatinase/formimionoglutamate hydrolase, SpeB		
MSM0878	pyruvoyl-dependent arginine decarboxylase, PdaD		

	MSM0888	glutamate dehydrogenase (NADP+), GdhA		
	MSM0967	glutamyl-tRNA reductase, HemA		
	MSM1052	prephenate dehydratase, PheA		
	MSM1084	argininosuccinate synthase, ArgG		
	MSM1103	phosphoribosyl-ATP pyrophosphohydrolase, HisE		
	MSM1141	tryptophan synthase, alpha subunit, TrpA		
	MSM1142	tryptophan synthase, beta subunit, TrpB		
	MSM1143	indole-3-glycerol phosphate synthase, TrpC		
	MSM1144	anthranilate phosphoribosyltransferase, TrpD		
	MSM1145	anthranilate/para-aminobenzoate synthase component II, TrpG		
	MSM1159	glutamine amidotransferase, HisH		
	MSM1172	leucyl-tRNA synthetase, LeuS		
	MSM1179	shikimate 5-dehydrogenase, AroE		
	MSM1181	histidyl-tRNA synthetase, HisS		
	MSM1182	phosphoribosyl-AMP cyclohydrolase, HisI		
	MSM1202	branched-chain-amino-acid aminotransferase, IlvE		
	MSM1206	imidazoleglycerol-phosphate dehydrogenase, HisB		
	MSM1214	threonyl-tRNA synthetase, ThrS		
	MSM1222	ketol-acid reductoisomerase, IlvC		
	MSM1224	acetolactate synthase, small subunit (regulatory), IlvH		
	MSM1226	ornithine carbamoyltransferase, ArgF		
	MSM1231	arginyl-tRNA synthetase, ArgS		
	MSM1236	aspartyl-tRNA synthetase, AspS		
	MSM1237	dihydroxy-acid dehydratase, IlvD		
	MSM1238	histidinol dehydrogenase, HisD		
	MSM1242	tryptophan synthase, beta subunit, TrpB		
	MSM1261	ATP phosphoribosyltransferase, HisG		
	MSM1263	aspartate carbamoyltransferase, PyrB		
	MSM1298	3-isopropylmalate dehydrogenase, LeuB		
	MSM1299	3-isopropylmalate dehydratase, small subunit, LeuD		
	MSM1300	3-isopropylmalate dehydratase, large subunit, LeuC		
	MSM1337	glycine hydroxymethyltransferase, GlyA		
	MSM1341	isoleucyl-tRNA synthetase, IleS		
	MSM1364	imidazoleglycerol-phosphate synthase, HisF		
	MSM1368	N-acetylornithine aminotransferase, ArgD		
	MSM1371	diaminopimelate decarboxylase, LysA		
	MSM1372	diaminopimelate epimerase, DapF		
	MSM1387	lysyl-tRNA synthetase (class I), LysS		
	MSM1418	glutamine synthetase, GlnA		
	MSM1440	predicted archaeal kinase		
	MSM1452	glutamyl-tRNA synthetase, GltX		
	MSM1474	chorismate synthase, AroC		
	MSM1478	phenylalanyl-tRNA synthetase, PheS		
	MSM1615	deoxyhypusine synthase, Dys		
	MSM1636	ProFAR isomerase-related protein		
	MSM1710	seryl-tRNA synthetase, SerS		
	MSM1713	predicted regulatory protein, amino acid-binding ACT domain family		

DEPLETED COMPARED TO SEQUENCED ARCHAEA AND/OR NON-GUT METHANOGENS

ENVIRONMENTAL SENSING	*M. smithii* genes	Annotation	GO:0004672 - protein kinase activity	GO:0007165 - signal transduction	GO:0000160 - two-component signal transduction
	MSM0485	universal stress protein, UspA	■	■	
	MSM0525	predicted unusual protein kinase, ubiquinone biosynthesis protein-related, AarF	■		
	MSM0869	GTPase, GTP1/OBG family		■	
	MSM0896	precorrin-6X reductase, CbiJ			■
	MSM0952	serine/threonine protein kinase, RIO1 family	■		
	MSM1198	O-sialoglycoprotein endopeptidase	■		

TRANSCRIPTIONAL CONTROL	*M. smithii* genes	Annotation	GO:0006350 - transcription	GO:0019222 - regulation of metabolism	GO:0045449 - regulation of transcription
	MSM0067	conserved hypothetical protein	■	■	■
	MSM0069	conserved hypothetical protein	■	■	■
	MSM0094	predicted transcription regulator (TetR family)	■	■	■
	MSM0196	archaeal DNA-dependent RNA polymerase, subunit E, RpoE	■		
	MSM0197	archaeal DNA-dependent RNA polymerase, subunit E, RpoE	■		
	MSM0218	iron dependent transcriptional regulator (Fe2+-binding)	■	■	■
	MSM0233	nitrogen regulatory protein P-II, GlnK		■	
	MSM0288	conserved hypothetical protein (DUF121 daomain)	■	■	■
	MSM0334	L-asparaginase, GatD,		■	
	MSM0364	transcriptional regulator (nickel-responsive), NikR	■	■	■
	MSM0413	transcriptional regulator, MarR family	■	■	■
	MSM0424	transcription initiation factor TFIIB (zinc-binding)	■	■	■
	MSM0600	transcriptional regulator, MarR family	■	■	■
	MSM0624	transcription antiterminator, NusG	■	■	■
	MSM0631	transcription initiation factor IIE, alpha subunit	■		
	MSM0650	transcriptional regulator, TetR/AcrR family	■	■	■
	MSM0720	transcription initiation factor TFIID TATA binding protein	■	■	■
	MSM0767	selenocysteine synthase, SelA		■	
	MSM0775	transcriptional regulator, AraC family	■	■	■
	MSM0817	predicted transcriptional regulator	■	■	■
	MSM0896	precorrin-6X reductase, CbiJ	■	■	■
	MSM0908	DNA-dependent RNA polymerase, subunit A, RpoA	■		
	MSM0909	DNA-dependent RNA polymerase, subunit A', RpoA	■		
	MSM0910	DNA-dependent RNA polymerase, subunit B', RpoB	■		
	MSM0911	DNA-dependent RNA polymerase, subunit B, RpoB	■		
	MSM0912	DNA-dependent RNA polymerase, subunit H, RpoH	■		
	MSM1032	transcriptional regulator, Lrp family	■	■	■

	MSM1087	NAD-dependent protein deacetylase, SIR2 family	
	MSM1107	predicted transcriptional regulator	
	MSM1126	predicted transcriptional regulator, ArsR family	
	MSM1137	DNA polymerase sliding clamp subunit, PCNA family, Pcn	
	MSM1207	molybdate transport system regulatory protein	
	MSM1230	transcriptional regulator, MarR family	
	MSM1315	predicted transcriptional regulator	
	MSM1350	predicted transcriptional regulator, ArsR family	
	MSM1354	DNA-directed RNA polymerase subunit M, RpoM	
	MSM1356	DNA-directed RNA polymerase, subunit L, RpoL	
	MSM1376	DNA-directed RNA polymerase subunit F	
	MSM1390	transcriptional regulator, LysR family	
	MSM1408	tungsten formylmethanofuran dehydrogenase, subunit E, FwdE	
	MSM1428	DNA-directed RNA polymerase, subunit D, RpoD	
	MSM1432	DNA-directed RNA polymerase, subunit N, RpoN	
	MSM1433	DNA-directed RNA polymerase, subunit K, RpoK	
	MSM1499	predicted transcriptional regulator	
	MSM1631	predicted DNA-directed RNA polymerase II, subunit RPC10	
	MSM1720	DNA-directed RNA polymerase, subunit M, RpoM	
	MSM1763	predicted DNA-directed RNA polymerase, subunit M, RpoM	
	MSM1764	conserved hypothetical protein	
	MSM1765	conserved hypothetical protein	
PHOSPHATE METABOLISM	***M. smithii* genes**	**Annotation**	**GO:0006796 - phosphate metabolism**
	MSM0060	predicted archaeal kinase (GHMP kinase family)	
	MSM0198	inorganic pyrophosphatase	
	MSM0313	[NiFe]-hydrogenase-3-type complex, small subunit/NADH:quinone oxidoreductase (complex I), subunit PSST/NdhK/NuoB	
	MSM0433	archaeal/vacuolar-type H+-transporting ATP synthase, subunit D	
	MSM0434	archaeal/vacuolar-type H+-transporting ATP synthase, subunit B	
	MSM0435	archaeal/vacuolar-type H+-transporting ATP synthase, subunit A	
	MSM0436	archaeal/vacuolar-type H+-transporting ATP synthase, subunit F	
	MSM0437	archaeal/vacuolar-type H+-transporting ATP synthase, subunit C	
	MSM0438	archaeal/vacuolar-type H+-transporting ATP synthase, subunit E	
	MSM0439	archaeal/vacuolar-type H+-transporting ATP synthase, subunit K	
	MSM0525	predicted unusual protein kinase, ubiquinone biosynthesis protein-related, AarF	

	MSM0835	archaeal shikimate kinase	
	MSM0842	histone acetyltransferase, radical SAM superfamily	
	MSM0848	predicted archaeal sugar kinase, GHMP kinase family	
	MSM0952	serine/threonine protein kinase, RIO1 family	
	MSM0988	phosphoenolpyruvate synthase, PpsA	
	MSM1000	methyl viologen-reducing hydrogenase, gamma subunit, MvhG	
	MSM1064	energy-converting hydrogenase B, subunit M, EhbM	
	MSM1071	energy-converting hydrogenase B, subunit F, EhbF	
	MSM1072	energy-converting hydrogenase B, subunit E, EhbE	
	MSM1122	coenzyme F420-reducing hydrogenase, gamma subunit, FrhG	
	MSM1198	O-sialoglycoprotein endopeptidase	
	MSM1424	polyphosphate kinase, ppk	
	MSM1439	mevalonate kinase	
RECOMBINATION	***M. smithii* genes**	**Annotation**	**GO:0006310 - DNA recombination**
	MSM0002	integrase-recombinase protein	
	MSM0428	integrase-recombinase protein, phage integrase family	
	MSM0611	DNA repair protein, RadB	
	MSM0645	ATP-dependent DNA ligase, Cdc9	
	MSM1333	DNA repair protein RadA, RadA	
	MSM1523	transposase	
	MSM1640	DNA intergrase/recombinase, phage integrase family	
	MSM1742	predicted integrase, phage integrase-like family	

Table 8. *M. smithii* proteins with homologs in other sequenced Methanobacteriales

M. smithii	*Methanosphaera stadtmanae*			*Methanothermobacter thermoautotrophicus*		
ORF	**ORF**	**ANNOTATION**	**E-value**	**ORF**	**ANNOTATION**	**E-value**
MSM0001	Msp_0220	predicted glycosyltransferase	4.2E-08	NONE		
MSM0002	Msp_1355	predicted site-specific recombinase/integrase	2.0E-08	MTH_893	integrase-recombinase protein	8.1E-16
MSM0003	Msp_0548	hypothetical membrane-spanning protein	6.8E-09	NONE		
MSM0004	Msp_0803	conserved hypothetical protein	2.3E-24	NONE		
MSM0005	Msp_0783	hypothetical membrane-spanning protein	3.7E-05	MTH_1439	unknown	6.2E-04
MSM0006	Msp_0725	hypothetical protein	1.3E-05	MTH_1277	unknown	3.3E-05
MSM0007	NONE			MTH_675	unknown	1.1E-34
MSM0008	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0009	NONE			MTH_675	unknown	8.1E-34
MSM0010	Msp_0813	conserved hypothetical protein	1.5E-36	MTH_676	unknown	1.7E-40
MSM0011	NONE			NONE		
MSM0012	Msp_0317	hypothetical protein	3.3E-04	NONE		
MSM0013	NONE			NONE		
MSM0014	NONE			MTH_1289	heat shock protein GrpE	2.6E-04
MSM0015	NONE			NONE		
MSM0016	NONE			NONE		
MSM0017	NONE			NONE		
MSM0018	NONE			NONE		
MSM0019	NONE			NONE		
MSM0020	Msp_1323	conserved hypothetical protein	1.4E-05	MTH_83	O-linked GlcNAc transferase	3.3E-07
MSM0021	Msp_0047	predicted short chain dehydrogenase	3.7E-40	NONE		
MSM0022	NONE			NONE		
MSM0023	Msp_0424	conserved hypothetical protein	1.6E-25	MTH_1084	conserved protein	4.4E-18
MSM0024	NONE			NONE		
MSM0025	Msp_0447	predicted acyl-CoA synthetase	3.7E-49	MTH_657	long-chain-fatty-acid-CoA ligase	8.7E-227
MSM0026	Msp_0265	conserved hypothetical protein	2.0E-16	MTH_659	epoxidase	4.1E-62
MSM0027	Msp_0667	putative glutamate synthase, subunit 2 with ferredoxin domain	7.9E-70	NONE	glutamate synthase (NADPH), alpha subunit	4.6E-79
MSM0028	Msp_0602	conserved hypothetical protein	1.9E-13	MTH_1876	conserved protein	1.7E-04
MSM0029	NONE			NONE		
MSM0030	Msp_0741	conserved hypothetical membrane-spanning protein	1.8E-72	MTH_1812	conserved protein	1.6E-44
MSM0031	Msp_1465	member of asn/thr-rich large protein family	2.9E-23	MTH_716	cell surface glycoprotein (s-layer protein)	3.7E-04
MSM0032	NONE			NONE		
MSM0033	Msp_0966	putative 2-dehydropantoate 2-reductase	6.8E-112	NONE		
MSM0034	Msp_0725	hypothetical protein	7.9E-06	NONE		
MSM0035	NONE			NONE		
MSM0036	NONE			NONE		
MSM0037	NONE			NONE		
MSM0038	NONE			NONE		
MSM0039	NONE			NONE		
MSM0040	Msp_1274	conserved hypothetical protein	5.5E-05	NONE		
MSM0041	NONE			NONE		
MSM0042	NONE			NONE		
MSM0043	Msp_0737	putative peptide methionine sulfoxide reductase MsrA/MsrB	1.6E-32	MTH_535	peptide methionine sulfoxide reductase	5.3E-16
MSM0044	Msp_0510	putative aspartate aminotransferase	2.0E-15	MTH_1894	aspartate aminotransferase homolog	3.9E-13

MSM0045	Msp_0283	predicted ATPase	3.9E-93	MTH_1176	nucleotide-binding protein (putative ATPase)	1.4E-70
MSM0046	Msp_1460	predicted NAD(FAD)-dependent dehydrogenase	8.4E-114	MTH_1354	NADH oxidase	2.0E-149
MSM0047	NONE			NONE		
MSM0048	Msp_0701	hypothetical protein	4.0E-20	NONE		
MSM0049	Msp_0665	F420H2:NADP oxidoreductase	3.1E-75	MTH_248	conserved protein	9.4E-56
MSM0050	Msp_1172	conserved hypothetical protein	1.7E-21	NONE		
MSM0051	Msp_1399	member of asn/thr-rich large protein family	4.0E-33	MTH_716	cell surface glycoprotein (s-layer protein)	3.9E-11
MSM0052	Msp_0145	member of asn/thr-rich large protein family	1.4E-53	MTH_716	cell surface glycoprotein (s-layer protein)	1.8E-11
MSM0053	Msp_0086	putative tRNA nucleotidyltransferase	5.0E-100	MTH_584	tRNA nucleotidyltransferase	2.5E-110
MSM0054	Msp_0089	predicted 2'-5' RNA ligase	7.2E-37	MTH_583	conserved protein	9.1E-42
MSM0055	Msp_0090	predicted 3-dehydroquinate synthase	3.5E-108	MTH_580	conserved protein	3.3E-124
MSM0056	Msp_0091	predicted fructose-bisphosphate aldolase	1.5E-100	MTH_579	conserved protein	2.9E-100
MSM0057	Msp_0762	member of asn/thr-rich large protein family	1.7E-13	MTH_716	cell surface glycoprotein (s-layer protein)	8.2E-07
MSM0058	Msp_0128	predicted helicase	8.6E-23	MTH_472	DNA helicase II	1.2E-90
MSM0059	Msp_0092	conserved hypothetical protein	9.4E-35	MTH_578	unknown	2.1E-49
MSM0060	Msp_1187	predicted archaeal kinase	8.2E-52	MTH_577	conserved protein	2.1E-49
MSM0061	Msp_0757	predicted ATPase	7.5E-97	NONE		
MSM0062	Msp_0554	hypothetical protein	2.2E-08	MTH_847	unknown	6.9E-08
MSM0063	Msp_1186	predicted hydrolase	1.3E-67	MTH_576	conserved protein	7.0E-51
MSM0064	Msp_0099	conserved hypothetical protein	4.6E-10	MTH_812	conserved protein	1.5E-09
MSM0065	Msp_1185	putative 5-amino-6-(5-phosphoribosylamino)uracil reductase	2.6E-55	MTH_235	riboflavin-specific deaminase	1.5E-66
MSM0066	Msp_0080	predicted glycosyltransferase	8.2E-107	MTH_590	N-acetylglucosamine-1-phosphate transferase	7.9E-107
MSM0067	NONE			NONE		
MSM0068	Msp_0407	conserved hypothetical protein	6.0E-04	MTH_521	unknown	8.4E-04
MSM0069	Msp_0081	conserved hypothetical protein	2.8E-26	MTH_589	conserved protein	3.1E-25
MSM0070	Msp_0082	conserved hypothetical protein	2.8E-99	MTH_588	conserved protein	4.8E-100
MSM0071	Msp_0083	MetG	5.3E-199	MTH_587	methionyl-tRNA synthetase	2.9E-235
MSM0072	Msp_0216	hypothetical membrane-spanning protein	2.2E-04	NONE		
MSM0073	Msp_0084	DNA primase, large subunit	1.4E-102	MTH_586	unknown	1.7E-118
MSM0074	NONE			NONE		
MSM0075	Msp_0085	DNA primase, small subunit	1.2E-96	NONE	DNA primase, small subunit	8.1E-105
MSM0076	Msp_0710	hypothetical protein	9.9E-04	NONE		
MSM0077	Msp_0357	putative thymidylate kinase	6.9E-16	MTH_1100	conserved protein	4.6E-47
MSM0078	NONE			MTH_1099	conserved protein	3.9E-50
MSM0079	Msp_0392	CofH	7.6E-81	MTH_820	conserved protein	1.0E-106
MSM0080	Msp_0278	ComD	1.0E-53	MTH_1206	phosphonopyruvate decarboxylase related protein	1.7E-47
MSM0081	Msp_0277	ComE	9.4E-51	MTH_1207	phosphonopyruvate decarboxylase related protein	1.7E-40
MSM0082	Msp_0127	HdrA2	1.3E-241	NONE	heterodisulfide reductase, subunit A	2.5E-133
MSM0083	Msp_0126	HdrB2	2.6E-94	NONE	heterodisulfide reductase, subunit B	8.6E-46
MSM0084	Msp_0125	HdrC2	2.6E-48	NONE	heterodisulfide reductase, subunit C	3.5E-17
MSM0085	Msp_1261	conserved hypothetical protein	6.6E-114	MTH_1684	conserved protein (contains ferredoxin domain)	2.1E-115
MSM0086	Msp_1270	ComA	5.2E-73	MTH_1674	conserved protein	3.5E-81
MSM0087	Msp_0233	conserved hypothetical protein	2.3E-22	NONE		
MSM0088	Msp_1322	conserved hypothetical protein	7.3E-44	MTH_727	conserved protein	1.6E-51
MSM0089	Msp_1314	ProC	8.2E-07	NONE		
MSM0090	NONE			MTH_224	conserved protein	8.6E-30
MSM0091	Msp_0129	putative 2,3-diphosphoglycerate synthase	8.6E-144	MTH_223	unknown	2.0E-172
MSM0092	Msp_0154	member of asn/thr-rich large protein family	5.6E-08	NONE		
MSM0093	Msp_1068	partially conserved hypothetical membrane-spanning protein	1.1E-58	MTH_1858	phage infection protein homolog	5.7E-98

MSM0094	Msp_0971	hypothetical protein	4.4E-09	MTH_1787	conserved protein	9.3E-17
MSM0095	Msp_1181	predicted phosphotransacetylase	1.3E-44	MTH_231	conserved protein	8.8E-44
MSM0096	Msp_1182	UppS	2.6E-96	MTH_232	conserved protein	2.3E-100
MSM0097	Msp_1183	predicted DNase	3.2E-57	MTH_233	conserved protein	3.4E-67
MSM0098	NONE			NONE		
MSM0099	Msp_0079	hypothetical membrane-spanning protein	2.1E-23	MTH_596	unknown	8.2E-25
MSM0100	Msp_0078	hypothetical membrane-spanning protein	7.3E-12	MTH_429	unknown	1.1E-13
MSM0101	Msp_0988	CbiF	9.8E-88	MTH_602	precorrin-3 methylase	1.5E-80
MSM0102	Msp_1236	MetE	3.4E-69	MTH_775	cobalamin-independent methionine synthase	3.8E-75
MSM0103	NONE			MTH_776	conserved protein	7.3E-33
MSM0104	NONE			MTH_777	conserved protein	2.7E-42
MSM0105	Msp_1234	conserved hypothetical membrane-spanning protein	3.8E-86	MTH_778	unknown	5.9E-118
MSM0106	Msp_1232	conserved hypothetical protein	1.8E-109	MTH_781	conserved protein	2.3E-132
MSM0107	Msp_1231	HypB	1.4E-79	MTH_782	hydrogenase expression/formation protein HypB	1.1E-84
MSM0108	Msp_1230	HypA	5.8E-35	MTH_783	hydrogenase expression/formation protein HypA	4.8E-36
MSM0109	Msp_0987	hypothetical membrane-spanning protein	8.6E-09	NONE		
MSM0110	Msp_0017	conserved hypothetical protein	1.5E-22	NONE		
MSM0111	NONE			NONE		
MSM0112	Msp_0367	predicted helicase	1.2E-208	NONE	ATP-dependent RNA helicase, eIF-4A family	1.4E-235
MSM0113	Msp_0128	predicted helicase	9.9E-137	MTH_472	DNA helicase II	6.1E-26
MSM0114	NONE			NONE		
MSM0115	Msp_1290	conserved hypothetical protein	8.0E-29	MTH_526	conserved protein	2.1E-51
MSM0116	Msp_1289	conserved hypothetical protein	3.5E-51	MTH_528	unknown	9.1E-42
MSM0117	Msp_1288	conserved hypothetical membrane-spanning protein	4.7E-56	MTH_529	unknown	1.5E-66
MSM0118	Msp_1286	conserved hypothetical protein	1.1E-86	MTH_532	UDP-N-acetylmuramyl tripeptide synthetase related protein	2.9E-86
MSM0119	Msp_0156	predicted nuclease	3.2E-18	MTH_538	unknown	2.5E-14
MSM0120	Msp_1095	DNA double-strand break repair protein Rad50	1.3E-92	MTH_540	intracellular protein transport protein	2.1E-27
MSM0121	Msp_1094	DNA double-strand break repair protein Mre11	3.7E-72	MTH_541	Rad32 related protein	1.2E-16
MSM0122	Msp_1093	predicted ATPase	1.7E-122	MTH_307	conserved protein	4.2E-124
MSM0123	Msp_1092	conserved hypothetical protein	2.4E-29	MTH_306	conserved protein	1.2E-32
MSM0124	Msp_1291	PcrB	5.1E-75	MTH_552	conserved protein	2.9E-84
MSM0125	Msp_1292	50S ribosomal protein L40e	5.5E-23	MTH_553	ribosomal protein L40	7.6E-22
MSM0126	Msp_1293	conserved hypothetical protein	9.4E-51	MTH_554	conserved protein	2.9E-54
MSM0127	NONE			NONE		
MSM0128	Msp_0853	conserved hypothetical membrane-spanning protein	2.3E-10	MTH_570	unknown	2.8E-31
MSM0129	Msp_0435	nicotinamide-nucleotide adenylyltransferase	8.1E-61	MTH_150	conserved protein	6.7E-62
MSM0130	NONE			MTH_149	molybdenum cofactor biosynthesis protein MoaE	6.6E-39
MSM0131	NONE			MTH_920	anion permease	1.5E-04
MSM0132	NONE			MTH_1797	conserved protein	7.9E-20
MSM0133	Msp_1198	predicted thioesterase	2.2E-42	MTH_658	unknown	4.8E-36
MSM0134	Msp_0565	predicted M42 glutamyl aminopeptidase	2.2E-115	NONE	endo-1,4-beta-glucanase	3.7E-116
MSM0135	Msp_0668	conserved hypothetical protein	9.1E-85	NONE	coenzyme F420-reducing hydrogenase, beta subunit homolog	4.5E-88
MSM0136	Msp_0147	ferredoxin	2.2E-06	NONE	tungsten formylmethanofuran dehydrogenase, subunit G	2.2E-06
MSM0137	Msp_0220	predicted glycosyltransferase	3.7E-12	MTH_540	intracellular protein transport protein	4.7E-05
MSM0138	NONE			MTH_491	conserved protein	2.6E-51
MSM0139	Msp_0448	predicted polysaccharide biosynthesis protein	7.6E-04	NONE		
MSM0140	Msp_0560	conserved hypothetical protein	4.0E-59	MTH_435	conserved protein	2.9E-68
MSM0141	Msp_0561	predicted dephospho-CoA kinase	5.5E-23	MTH_434	UMP/CMP kinase related protein	5.6E-42
MSM0142	Msp_0563	predicted ATPase of PP-loop superfamily	3.2E-66	MTH_432	conserved protein	2.9E-68

MSM0143	Msp_0564	partially conserved hypothetical membrane-spanning protein	1.3E-30	MTH_431	unknown	2.4E-34
MSM0144	NONE			NONE		
MSM0145	Msp_0451	hypothetical membrane-spanning protein	1.9E-13	MTH_422	unknown	1.6E-14
MSM0146	Msp_0452	conserved hypothetical membrane-spanning protein	7.0E-18	MTH_421	unknown	2.0E-21
MSM0147	Msp_0453	PyrG	2.2E-202	MTH_419	CTP synthase	2.9E-212
MSM0148	Msp_0739	predicted oxidoreductase	3.9E-93	MTH_907	conserved protein	3.1E-32
MSM0149	NONE			NONE		
MSM0150	NONE			NONE		
MSM0151	NONE			NONE		
MSM0152	Msp_1417	predicted Na+-driven multidrug efflux pump	1.1E-28	MTH_314	conserved protein	4.7E-23
MSM0153	Msp_0485	ApgM1	1.3E-110	MTH_418	phosphonopyruvate decarboxylase related protein	2.1E-106
MSM0154	Msp_0487	putative homoserine dehydrogenase	1.3E-101	MTH_417	homoserine dehydrogenase homolog	6.1E-100
MSM0155	Msp_0488	predicted allosteric regulator of homoserine dehydrogenase	1.1E-29	MTH_416	conserved protein	7.8E-36
MSM0156	Msp_0489	conserved hypothetical protein	2.6E-23	MTH_415	conserved protein	3.3E-21
MSM0157	Msp_0484	predicted type I restriction-modification system subunit	1.9E-09	NONE	type I restriction modification system, subunit S	5.3E-09
MSM0158	Msp_0483	hypothetical protein	2.3E-17	NONE	type I restriction modification system, subunit S	2.2E-13
MSM0159	Msp_0777	member of asn/thr-rich large protein family	2.1E-13	NONE		
MSM0160	Msp_0490	putative asparagine synthetase	7.9E-102	MTH_414	asparagine synthetase	2.3E-91
MSM0161	NONE			NONE		
MSM0162	NONE			NONE		
MSM0163	Msp_0425	conserved hypothetical protein	7.0E-23	MTH_1083	conserved protein	5.6E-26
MSM0164	Msp_0946	conserved hypothetical protein	1.3E-106	MTH_1084	conserved protein	4.6E-118
MSM0165	Msp_0945	predicted RecB family exonuclease	7.9E-54	MTH_1085	conserved protein	1.8E-45
MSM0166	Msp_0422	predicted helicase	2.3E-27	MTH_1086	conserved protein	9.1E-32
MSM0167	NONE			MTH_1087	unknown	8.4E-04
MSM0168	NONE			NONE		
MSM0169	Msp_0220	predicted glycosyltransferase	2.1E-04	NONE		
MSM0170	Msp_0944	conserved hypothetical protein	1.4E-63	MTH_1091	conserved protein	3.4E-35
MSM0171	Msp_0835	hypothetical membrane-spanning protein	2.7E-43	MTH_769	unknown	1.7E-34
MSM0172	NONE			NONE		
MSM0173	Msp_0145	member of asn/thr-rich large protein family	3.2E-34	MTH_1074	putative membrane protein	5.5E-31
MSM0174	Msp_0677	predicted O-acetylhomoserine sulfhydrylase	1.9E-123	NONE		
MSM0175	Msp_0676	MetX	2.3E-166	MTH_1820	homoserine O-acetyltransferase	1.5E-21
MSM0176	NONE			NONE		
MSM0177	NONE			NONE		
MSM0178	Msp_1385	conserved hypothetical protein	1.5E-27	NONE		
MSM0179	NONE			NONE		
MSM0180	NONE			MTH_698	unknown	1.6E-04
MSM0181	Msp_1174	50S ribosomal protein L37e	9.6E-26	MTH_648	ribosomal protein L37	2.8E-24
MSM0182	Msp_1175	putative snRNP Sm-like protein	1.5E-27	MTH_649	conserved protein	2.1E-33
MSM0183	Msp_1176	predicted RNA-binding protein	9.0E-46	MTH_650	conserved protein	8.6E-46
MSM0184	Msp_1177	predicted creatinine amidohydrolase	1.3E-51	MTH_651	conserved protein	1.6E-51
MSM0185	Msp_0547	hypothetical membrane-spanning protein	7.8E-08	MTH_515	unknown	4.3E-05
MSM0186	Msp_0345	conserved hypothetical protein	1.3E-14	NONE		
MSM0187	Msp_0444	rubredoxin	2.5E-09	MTH_156	rubredoxin	2.3E-13
MSM0188	Msp_0444	rubredoxin	3.4E-14	MTH_156	rubredoxin	3.5E-17
MSM0189	Msp_1301	predicted nucleoside-diphosphate-sugar pyrophosphorylase	4.6E-08	MTH_272	acetyl / acyl transferase related protein	1.3E-58
MSM0190	Msp_0617	predicted ATPase	3.1E-84	MTH_271	conserved protein	1.8E-75

MSM0191	Msp_1533	RpoM1	1.5E-04	NONE		
MSM0192	Msp_0618	ArgH	2.7E-147	MTH_269	argininosuccinate lyase	8.2E-160
MSM0193	Msp_0620	30S ribosomal protein S27Ae	1.8E-17	MTH_268	ribosomal protein S27a	8.1E-18
MSM0194	Msp_0621	30S ribosomal protein S24e	1.1E-26	MTH_267	ribosomal protein S24	1.6E-28
MSM0195	Msp_0622	conserved hypothetical protein	4.8E-31	MTH_266	conserved protein	1.3E-33
MSM0196	Msp_0623	RpoE2	9.0E-14	NONE	DNA-dependent RNA polymerase, subunit E''	1.5E-18
MSM0197	Msp_0624	RpoE1	2.2E-65	NONE	DNA-dependent RNA polymerase, subunit E'	1.3E-67
MSM0198	Msp_0625	inorganic pyrophosphatase	3.1E-68	MTH_263	inorganic pyrophosphatase	7.2E-65
MSM0199	Msp_0626	conserved hypothetical protein	2.4E-22	MTH_262	conserved protein	3.7E-29
MSM0200	Msp_0627	putative translation initiation factor 2, subunit gamma (aIF-2-gamma)(eIF2G)	3.3E-158	NONE	translation initiation factor eIF-2, gamma subunit	1.6E-163
MSM0201	Msp_0628	30S ribosomal protein S6e	9.9E-40	MTH_260	ribosomal protein S6	1.5E-41
MSM0202	Msp_0629	InfB	9.3E-202	MTH_259	translation initiation factor IF2 homolog	2.6E-218
MSM0203	Msp_0630	nucleoside diphosphate kinase	1.8E-56	MTH_258	nucleoside diphosphate kinase	1.9E-57
MSM0204	Msp_0631	50S ribosomal protein L24e	3.0E-22	MTH_257	ribosomal protein L24	8.2E-25
MSM0205	Msp_0632	30S ribosomal protein S28e	4.3E-30	MTH_256	ribosomal protein S28	2.2E-31
MSM0206	Msp_0633	50S ribosomal protein L7Ae	9.3E-44	MTH_255	ribosomal protein L7a	1.3E-44
MSM0207	NONE			MTH_1178	conserved protein	1.9E-41
MSM0208	NONE			MTH_1178	conserved protein	3.9E-08
MSM0209	Msp_0861	ferredoxin	7.3E-12	MTH_1106	ferredoxin	7.6E-22
MSM0210	Msp_0253	conserved hypothetical membrane-spanning protein	1.1E-04	NONE		
MSM0211	NONE			NONE		
MSM0212	NONE			NONE		
MSM0213	Msp_0769	archaeal histone	8.2E-20	MTH_821	histone HMtA1	3.7E-22
MSM0214	Msp_0588	ThrC	2.0E-153	MTH_253	threonine synthase	8.8E-163
MSM0215	Msp_0232	hypothetical membrane-spanning protein	2.4E-22	MTH_252	conserved protein	4.5E-24
MSM0216	Msp_0653	TrpS	5.0E-132	MTH_251	tryptophanyl-tRNA synthetase	1.8E-116
MSM0217	Msp_0652	EndA	5.0E-45	MTH_250	tRNA intron endonuclease	2.7E-49
MSM0218	Msp_0446	predicted metal-dependent transcriptional regulator	5.3E-57	MTH_214	iron repressor	6.4E-57
MSM0219	Msp_1129	partially conserved hypothetical membrane-spanning protein	1.0E-46	MTH_357	conserved protein	4.0E-67
MSM0220	Msp_0114	ThsB	1.7E-170	MTH_218	chaperonin	4.0E-183
MSM0221	Msp_0590	member of asn/thr-rich large protein family	6.9E-13	MTH_719	cell surface glycoprotein (s-layer protein)	4.2E-05
MSM0222	Msp_0787	FprA	2.5E-128	MTH_220	flavoprotein A homolog (II)	3.2E-133
MSM0223	NONE			MTH_557	unknown	1.4E-22
MSM0224	NONE			MTH_558	unknown	2.1E-28
MSM0225	Msp_1294	conserved hypothetical membrane-spanning protein	1.4E-47	MTH_559	conserved protein	1.4E-54
MSM0226	NONE			NONE		
MSM0227	Msp_0584	HmgA	2.2E-138	MTH_562	3-hydroxy-3-methylglutaryl CoA reductase	1.7E-143
MSM0228	Msp_0583	SucD	1.7E-99	NONE	succinyl-CoA synthetase, alpha subunit	1.3E-111
MSM0229	Msp_0582	conserved hypothetical protein	1.6E-69	MTH_564	conserved protein	1.5E-87
MSM0230	Msp_0233	conserved hypothetical protein	2.9E-21	NONE		
MSM0231	Msp_0577	AroD	9.9E-40	MTH_566	3-dehydroquinate dehydratase	2.9E-52
MSM0232	Msp_0145	member of asn/thr-rich large protein family	3.8E-05	MTH_567	unknown	7.5E-31
MSM0233	Msp_0664	nitrogen regulatory protein P-II	7.9E-31	MTH_664	nitrogen regulatory protein P-II	1.4E-36
MSM0234	Msp_0663	ammonium transporter	4.8E-150	MTH_663	ammonium transporter	1.2E-142
MSM0235	Msp_0119	hypothetical membrane-spanning protein	6.0E-04	MTH_181	unknown	1.4E-04
MSM0236	Msp_0434	predicted phosphohydrolase	1.2E-100	MTH_148	conserved protein	7.8E-123
MSM0237	Msp_0088	predicted 3-polyprenyl-4-hydroxybenzoate decarboxylase	3.1E-59	MTH_147	phenylacrylic acid decarboxylase	2.6E-53
MSM0238	Msp_0087	CbiT	4.2E-48	MTH_146	precorrin-8W decarboxylase	3.1E-48
MSM0239	NONE			MTH_145	conserved protein	6.9E-44

MSM0240	Msp_1289	conserved hypothetical protein	8.3E-07	MTH_143	molybdopterin-guanine dinucleotide biosynthesis MobA related protein	1.6E-30
MSM0241	Msp_1252	putative exosome complex, exonuclease 2 subunit	1.1E-61	MTH_682	conserved protein	5.6E-90
MSM0242	Msp_1251	putative exosome complex, exonuclease 1 subunit	1.4E-79	MTH_683	ribonuclease PH	1.1E-93
MSM0243	Msp_1250	putative exosome complex, RNA-binding subunit	1.6E-48	MTH_684	conserved protein	2.1E-90
MSM0244	Msp_1249	conserved hypothetical protein	1.8E-70	MTH_685	conserved protein	8.3E-80
MSM0245	Msp_1248	PsmA	6.3E-77	NONE	proteasome, alpha subunit	2.5E-94
MSM0246	Msp_1246	putative ribonuclease P, component 2	1.3E-19	MTH_687	conserved protein	2.3E-22
MSM0247	Msp_1245	putative ribonuclease P, component 3	2.1E-28	MTH_688	conserved protein	3.1E-41
MSM0248	Msp_0950	hypothetical protein	7.2E-05	NONE		
MSM0249	Msp_1548	hypothetical protein	1.8E-04	MTH_301	unknown	4.1E-23
MSM0250	Msp_0501	hypothetical membrane-spanning protein	1.0E-05	MTH_521	unknown	3.6E-10
MSM0251	Msp_0725	hypothetical protein	1.5E-04	NONE		
MSM0252	Msp_0824	predicted Na+-driven multidrug efflux pump	1.6E-96	MTH_314	conserved protein	3.7E-93
MSM0253	NONE			MTH_1725	unknown	1.4E-15
MSM0254	NONE			NONE		
MSM0255	NONE			NONE		
MSM0256	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0257	Msp_0975	hypothetical membrane-spanning protein	4.3E-30	NONE		
MSM0258	Msp_0724	hypothetical membrane-spanning protein	1.6E-04	NONE		
MSM0259	Msp_1548	hypothetical protein	1.1E-05	MTH_521	unknown	6.8E-04
MSM0260	Msp_0507	predicted archaea-specific RecJ-like exonuclease	2.0E-199	MTH_763	conserved protein	3.4E-225
MSM0261	Msp_1384	conserved hypothetical membrane-spanning protein	1.1E-04	MTH_759	unknown	1.5E-16
MSM0262	Msp_0788	desulfoferrodoxin	1.4E-26	MTH_757	rubredoxin oxidoreductase	3.4E-26
MSM0263	Msp_1003	predicted NifU protein	1.1E-47	NONE		
MSM0264	Msp_1002	IscS	6.6E-121	MTH_1389	nifS protein	1.6E-30
MSM0265	Msp_0677	predicted O-acetylhomoserine sulfhydrylase	1.5E-148	MTH_1188	pleiotropic regulatory protein DegT	3.1E-04
MSM0266	Msp_0145	member of asn/thr-rich large protein family	2.7E-50	MTH_911	probable surface protein	6.2E-09
MSM0267	Msp_0844	predicted multimeric flavodoxin	4.4E-53	MTH_135	conserved protein	2.7E-17
MSM0268	Msp_0124	CysS	1.2E-139	MTH_587	methionyl-tRNA synthetase	9.6E-08
MSM0269	Msp_0527	conserved hypothetical protein	8.0E-38	NONE		
MSM0270	Msp_0450	predicted serine acetyltransferase	8.1E-61	MTH_1588	ferripyochelin binding protein	2.0E-06
MSM0271	Msp_0449	cysteine synthase	2.2E-97	NONE	tryptophan synthase, beta subunit	3.1E-08
MSM0272	Msp_0497	putative endonuclease III	2.2E-67	MTH_764	endonuclease III	1.1E-70
MSM0273	Msp_0498	AroA	1.1E-102	MTH_766	5-enolpyruvylshikimate 3-phosphate synthase	2.5E-62
MSM0274	NONE			NONE		
MSM0275	Msp_0499	ValS	2.4E-235	MTH_767	valyl-tRNA synthetase	0.0E+00
MSM0276	Msp_0526	hypothetical membrane-spanning protein	8.1E-29	MTH_768	unknown	2.9E-22
MSM0277	Msp_0525	PheT	3.3E-151	MTH_770	phenylalanyl-tRNA synthetase	4.2E-172
MSM0278	NONE			NONE		
MSM0279	Msp_0522	conserved hypothetical protein	4.0E-36	MTH_771	conserved protein	2.7E-35
MSM0280	Msp_0757	predicted ATPase	4.4E-13	NONE		
MSM0281	Msp_0145	member of asn/thr-rich large protein family	2.1E-09	MTH_911	probable surface protein	2.9E-10
MSM0282	Msp_0141	member of asn/thr-rich large protein family	1.3E-23	MTH_911	probable surface protein	1.1E-17
MSM0283	NONE			MTH_436	unknown	1.1E-04
MSM0284	Msp_0995	RpiA	5.8E-74	MTH_608	ribose 5-phosphate isomerase	1.3E-74
MSM0285	Msp_0996	conserved hypothetical protein	1.3E-28	MTH_609	conserved protein	1.3E-35
MSM0286	Msp_0997	EgsA	7.9E-102	MTH_610	glycerol 1-phosphate dehydrogenase	1.5E-112
MSM0287	Msp_1004	ProS	8.6E-160	MTH_611	prolyl-tRNA synthetase	1.4E-155
MSM0288	Msp_1006	conserved hypothetical protein	1.7E-53	MTH_613	conserved protein	4.2E-60
MSM0289	Msp_1007	ThiD	3.6E-58	MTH_614	transcriptional regulator	5.1E-64

MSM0290	Msp_1000	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, ATB-binding protein	2.6E-71	MTH_920	anion permease	1.4E-31
MSM0291	Msp_1001	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, permease protein	1.9E-84	MTH_1730	phosphate transporter permease PstC homolog	4.8E-07
MSM0292	NONE			NONE		
MSM0293	Msp_0826	predicted cation transport ATPase	1.8E-198	MTH_1535	heavy-metal transporting CPx-type ATPase	1.2E-69
MSM0294	Msp_0825	hypothetical protein	4.2E-09	NONE		
MSM0295	NONE			NONE	nitrate assimilation protein, narQ	7.1E-49
MSM0296	NONE			MTH_691	conserved protein	1.2E-30
MSM0297	Msp_1244	predicted exosome subunit	1.1E-24	MTH_689	conserved protein	2.7E-26
MSM0298	Msp_1243	50S ribosomal protein L15e	2.1E-76	MTH_690	ribosomal protein L15	1.3E-67
MSM0299	NONE			NONE		
MSM0300	Msp_0851	predicted ABC-type dipeptide/oligopeptide/nickel transport system, solute-binding protein	1.5E-139	NONE		
MSM0301	Msp_0811	ABC-type dipeptide transport system, permease protein	2.3E-120	NONE		
MSM0302	Msp_0810	ABC-type dipeptide transport system, permease protein	1.7E-99	MTH_1729	phosphate transporter permease PstC	2.3E-05
MSM0303	Msp_0848	predicted ABC-type dipeptide/oligopeptide/nickel transport system, ATP-binding protein	3.4E-101	MTH_696	ABC transporter (glutamine transport ATP-binding protein)	1.4E-20
MSM0304	Msp_0847	predicted ABC-type dipeptide/oligopeptide/nickel transport system, ATP-binding protein	4.8E-63	NONE	methyl coenzyme M reductase system, component A2	7.3E-21
MSM0305	Msp_0431	GuaB	6.1E-10	MTH_406	conserved protein	7.6E-70
MSM0306	Msp_1447	EhbK	3.0E-18	MTH_405	polyferredoxin	1.6E-37
MSM0307	Msp_0071	predicted ribokinase	3.4E-62	MTH_404	ribokinase	3.5E-65
MSM0308	Msp_0070	formylmethanofuran-tetrahydromethanopterin formyltransferase	6.7E-89	MTH_403	formylmethanofuran:tetrahydromethanopterin formyltransferase II	1.7E-95
MSM0309	Msp_0069	conserved hypothetical membrane-spanning protein	2.4E-68	MTH_402	unknown	3.9E-57
MSM0310	Msp_1447	EhbK	1.7E-23	MTH_401	polyferredoxin	7.7E-77
MSM0311	Msp_1447	EhbK	2.1E-13	MTH_399	polyferredoxin	7.4E-111
MSM0312	Msp_1444	EhbN	2.2E-51	NONE	formate hydrogenlyase, subunit 5	7.8E-139
MSM0313	Msp_1445	EhbM	5.4E-32	NONE	formate hydrogenlyase, subunit 7	6.3E-66
MSM0314	NONE			MTH_396	conserved protein	2.9E-29
MSM0315	NONE			MTH_395	conserved protein	1.9E-18
MSM0316	Msp_0616	partially conserved hypothetical membrane-spanning protein	9.5E-04	MTH_394	unknown	5.8E-08
MSM0317	Msp_1443	EhbO	1.1E-16	NONE	NADH dehydrogenase (ubiquinone), subunit 1 related protein	1.9E-105
MSM0318	NONE			MTH_392	unknown	1.4E-15
MSM0319	Msp_1452	EhbF	4.0E-06	NONE	NADH dehydrogenase I, subunit N related protein	5.5E-83
MSM0320	NONE			MTH_390	conserved protein	7.0E-67
MSM0321	NONE			MTH_389	conserved protein	6.6E-55
MSM0322	NONE			MTH_388	unknown	1.5E-25
MSM0323	NONE			MTH_387	conserved protein	3.9E-18
MSM0324	NONE			MTH_386	unknown	6.4E-18
MSM0325	NONE			MTH_385	conserved protein	4.1E-55
MSM0326	NONE			MTH_384	unknown	3.5E-17
MSM0327	Msp_0067	putative UDP-glucose 4-epimerase	1.2E-73	MTH_380	UDP-glucose 4-epimerase homolog	1.7E-86
MSM0328	NONE			MTH_698	unknown	2.7E-10
MSM0329	Msp_0265	conserved hypothetical protein	7.4E-51	MTH_700	conserved protein	5.1E-64
MSM0330	Msp_0266	predicted acyl-CoA synthetase	1.1E-184	MTH_701	acetyl-CoA synthetase related protein	1.0E-138
MSM0331	Msp_1390	KorD	7.0E-07	NONE	2-oxoisovalerate oxidoreductase, gamma subunit	7.9E-20
MSM0332	Msp_1389	KorA	1.6E-56	NONE	2-oxoisovalerate oxidoreductase, beta subunit	6.4E-144

MSM0333	Msp_1388	KorB	2.0E-28	NONE	2-oxoisovalerate oxidoreductase, alpha subunit	8.0E-169
MSM0334	Msp_1411	GatD	9.1E-140	MTH_706	L-asparaginase I	6.4E-144
MSM0335	Msp_1412	GatE	8.1E-187	MTH_707	PET112-like protein	7.1E-209
MSM0336	NONE			NONE		
MSM0337	Msp_0145	member of asn/thr-rich large protein family	1.1E-08	NONE		
MSM0338	NONE			NONE		
MSM0339	NONE			NONE		
MSM0340	Msp_1413	predicted thioredoxin reductase	1.4E-70	MTH_708	thioredoxin reductase	6.9E-92
MSM0341	NONE			NONE		
MSM0342	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0343	Msp_1311	GMP synthase [glutamine hydrolyzing], subunit A	4.2E-64	NONE	GMP synthetase, subunit A	1.1E-68
MSM0344	NONE			NONE		
MSM0345	Msp_1312	GMP synthase [glutamine hydrolyzing], subunit B	3.4E-117	NONE	GMP synthetase, subunit B	7.1E-122
MSM0346	Msp_1315	conserved hypothetical protein	8.0E-125	MTH_720	unknown	3.1E-128
MSM0347	Msp_1316	conserved hypothetical protein	6.5E-43	MTH_721	conserved protein	8.6E-62
MSM0348	Msp_1317	conserved hypothetical protein	7.1E-14	MTH_722	conserved protein	2.3E-22
MSM0349	Msp_1317	conserved hypothetical protein	1.5E-05	MTH_722	conserved protein	1.2E-04
MSM0350	Msp_1318	predicted isopropylmalate/homocitrate/citramalate synthase	3.9E-155	MTH_723	2-isopropylmalate synthase	6.2E-162
MSM0351	NONE			NONE		
MSM0352	Msp_1319	predicted DNA modification methylase	1.4E-72	MTH_724	methyltransferase related protein	4.3E-83
MSM0353	Msp_1321	hypothetical membrane-spanning protein	4.8E-11	NONE		
MSM0354	Msp_1206	proteasome-activating nucleotidase	4.1E-144	MTH_728	ATP-dependent 26S protease regulatory subunit 4	1.2E-172
MSM0355	Msp_1207	predicted transcriptional regulator	7.4E-35	MTH_729	conserved protein	2.7E-33
MSM0356	Msp_1208	conserved hypothetical protein	2.3E-24	MTH_730	conserved protein	6.2E-27
MSM0357	Msp_1209	conserved hypothetical membrane-spanning protein	1.6E-128	MTH_731	unknown	1.5E-110
MSM0358	Msp_1210	conserved hypothetical membrane-spanning protein	7.3E-44	MTH_733	unknown	3.7E-45
MSM0359	Msp_1213	predicted UDP-N-acetylmuramyl tripeptide synthase	1.7E-108	MTH_530	UDP-N-acetylmuramyl tripeptide synthetase related protein	5.2E-14
MSM0360	Msp_1214	predicted UDP-N-acetylmuramyl pentapeptide phosphotransferase	1.9E-91	MTH_735	phospho-N-acetylmuramoyl-pentapeptide- transferase	2.8E-102
MSM0361	Msp_1215	partially conserved hypothetical protein, predicted carbamoyl-phosphate synthase, large chain	6.8E-96	MTH_736	conserved protein	2.0E-76
MSM0362	Msp_1216	partially conserved hypothetical protein	5.4E-16	NONE	coenzyme F420-reducing hydrogenase, delta subunit homolog	5.3E-30
MSM0363	Msp_1217	predicted RNA methylase	3.2E-50	MTH_738	conserved protein	1.0E-56
MSM0364	Msp_1218	putative nickel responsive regulator	3.0E-54	MTH_739	conserved protein	9.1E-58
MSM0365	Msp_1090	hypothetical protein	2.1E-23	MTH_741	unknown	1.8E-22
MSM0366	NONE			NONE		
MSM0367	Msp_0099	conserved hypothetical protein	6.0E-17	MTH_812	conserved protein	5.6E-26
MSM0368	Msp_0667	putative glutamate synthase, subunit 2 with ferredoxin domain	1.3E-193	NONE	glutamate synthase (NADPH), alpha subunit	1.3E-216
MSM0369	Msp_0669	putative glutamate synthase, subunit 3	1.2E-68	NONE	tungsten formylmethanofuran dehydrogenase, subunit C homolog	1.1E-82
MSM0370	Msp_0670	putative glutamate synthase, subunit 1	5.7E-115	MTH_191	glutamine PRPP amidotransferase	2.2E-127
MSM0371	Msp_0671	predicted glutamine amidotransferase	6.2E-54	MTH_190	conserved protein	3.3E-60
MSM0372	Msp_0673	partially conserved hypothetical protein	1.3E-23	MTH_187	conserved protein	2.8E-24
MSM0373	Msp_1484	LeuB	3.3E-96	MTH_184	isocitrate dehydrogenase	4.5E-104
MSM0374	Msp_0447	predicted acyl-CoA synthetase	8.3E-178	MTH_657	long-chain-fatty-acid-CoA ligase	5.0E-58
MSM0375	Msp_0550	ArgB	2.3E-111	MTH_183	acetylglutamate kinase	2.5E-110
MSM0376	Msp_0967	putative NADP-dependent alcohol dehydrogenase	6.2E-06	NONE		

MSM0377	Msp_0310	predicted GTP:adenosylcobinamide-phosphate guanylyltransferase	4.9E-07	MTH_1152	conserved protein	6.5E-05
MSM0378	NONE			MTH_1876	conserved protein	1.3E-24
MSM0379	Msp_0549	ArgJ	6.5E-107	MTH_182	glutamate N-acetyltransferase	1.9E-103
MSM0380	Msp_0506	hypothetical membrane-spanning protein	2.1E-05	MTH_181	unknown	1.8E-04
MSM0381	Msp_0546	conserved hypothetical membrane-spanning protein	2.8E-99	MTH_180	unknown	1.4E-114
MSM0382	Msp_0545	conserved hypothetical protein	3.7E-95	MTH_179	unknown	1.9E-103
MSM0383	Msp_0544	predicted phosphohydrolase	1.0E-62	MTH_178	Icc related protein	2.6E-53
MSM0384	Msp_0543	conserved hypothetical protein	4.1E-34	MTH_177	conserved protein	1.9E-34
MSM0385	Msp_0511	predicted Fe-S oxidoreductase	3.2E-07	MTH_1784	Mg-protoporphyrin IX monomethyl ester oxidative cyclase	9.9E-84
MSM0386	Msp_0148	predicted sodium:solute symporter	1.9E-178	MTH_1856	sodium/proline symporter (proline permease)	1.5E-181
MSM0387	Msp_1040	coenzyme F390 synthetase II	2.2E-145	MTH_1855	coenzyme F390 synthetase II	1.4E-162
MSM0388	Msp_1041	predicted regulatory protein	4.1E-34	MTH_1854	unknown	2.6E-37
MSM0389	Msp_0136	hypothetical protein	1.5E-06	NONE		
MSM0390	NONE			NONE		
MSM0391	Msp_1042	lorB	5.6E-53	NONE	indolepyruvate oxidoreductase, beta subunit	2.4E-50
MSM0392	Msp_1043	lorA	6.7E-185	NONE	indolepyruvate oxidoreductase, alpha subunit	4.1E-192
MSM0393	Msp_1044	TfrB	3.3E-135	MTH_1850	fumarate reductase	1.4E-155
MSM0394	Msp_1047	predicted rRNA methylase	2.2E-65	MTH_1849	conserved protein	1.2E-69
MSM0395	Msp_1581	partially conserved hypothetical protein	2.7E-46	MTH_745	unknown (contains ferredoxin domain)	3.9E-57
MSM0396	Msp_0233	conserved hypothetical protein	2.3E-22	NONE		
MSM0397	NONE			NONE		
MSM0398	Msp_1229	ribose-phosphate pyrophosphokinase	6.6E-04	MTH_1114	uracil phosphoribosyltransferase	6.6E-23
MSM0399	NONE			NONE		
MSM0400	NONE			NONE		
MSM0401	NONE			MTH_75	surface protease related protein	2.7E-27
MSM0402	Msp_1048	deoxycytidine triphosphate deaminase	3.5E-76	MTH_1847	deoxycytidine triphosphate deaminase	1.1E-75
MSM0403	Msp_1049	GlyS	2.1E-188	MTH_1846	glycyl-tRNA synthetase	7.6E-196
MSM0404	Msp_0799	predicted transcriptional regulator	1.6E-25	MTH_1843	unknown	9.1E-26
MSM0405	Msp_1050	predicted metal-dependent hydrolase	1.7E-58	MTH_1842	conserved protein	2.5E-46
MSM0406	Msp_1052	hypothetical protein	1.7E-10	MTH_1838	unknown	6.6E-23
MSM0407	Msp_1053	conserved hypothetical membrane-spanning protein	1.7E-115	MTH_1837	unknown	1.2E-124
MSM0408	Msp_0406	2-phosphoglycerate kinase-like/predicted small molecule-binding domain fusion	4.2E-80	MTH_1835	2-phosphoglycerate kinase homolog	2.3E-91
MSM0409	Msp_0407	conserved hypothetical protein	2.2E-42	MTH_1834	conserved protein	9.5E-47
MSM0410	Msp_0409	conserved hypothetical protein	3.9E-52	MTH_1833	unknown	4.6E-47
MSM0411	Msp_0145	member of asn/thr-rich large protein family	1.3E-25	MTH_1074	putative membrane protein	1.3E-115
MSM0412	Msp_0046	member of asn/thr-rich large protein family	1.3E-06	MTH_117	unknown	2.4E-41
MSM0413	Msp_0512	predicted transcriptional regulator	2.7E-21	MTH_313	transcriptional regulator	1.9E-16
MSM0414	Msp_0824	predicted Na+-driven multidrug efflux pump	2.8E-138	MTH_314	conserved protein	6.7E-110
MSM0415	Msp_1362	PyrH	3.5E-76	MTH_879	uridine monophosphate kinase	2.8E-79
MSM0416	Msp_0974	predicted Mg-dependent DNase	1.5E-93	MTH_233	conserved protein	8.0E-27
MSM0417	Msp_1361	hypothetical membrane-spanning protein	3.8E-15	MTH_880	unknown	3.2E-14
MSM0418	Msp_1045	conserved hypothetical protein	2.5E-34	MTH_507	conserved protein	2.5E-32
MSM0419	Msp_0253	conserved hypothetical membrane-spanning protein	1.4E-24	MTH_506	unknown	4.2E-21
MSM0420	Msp_0355	conserved hypothetical membrane-spanning protein	3.0E-22	MTH_882	conserved protein	1.1E-27
MSM0421	NONE			NONE		
MSM0422	Msp_0644	conserved hypothetical membrane-spanning protein	1.1E-36	MTH_883	unknown	6.3E-48
MSM0423	Msp_0645	predicted glycosyltransferase	6.9E-157	MTH_884	teichoic acid biosynthesis related protein	4.5E-184
MSM0424	Msp_1360	transcription initiation factor IIB (TFIIB)	8.1E-148	MTH_885	transcription initiation factor TFIIB	9.2E-152
MSM0425	Msp_1359	hypothetical protein	2.3E-15	MTH_886	conserved protein	3.4E-19

MSM0426	Msp_1358	predicted demethylmenaquinone methyltransferase	3.7E-33	MTH_888	conserved protein	3.2E-46
MSM0427	Msp_1356	predicted DNA primase	7.2E-108	MTH_891	conserved protein	2.9E-141
MSM0428	Msp_1355	predicted site-specific recombinase/integrase	2.5E-66	MTH_893	integrase-recombinase protein	7.7E-77
MSM0429	Msp_1354	conserved hypothetical protein	4.3E-46	MTH_905	conserved protein	1.8E-38
MSM0430	NONE			MTH_906	unknown	2.7E-17
MSM0431	Msp_1132	predicted ATP-dependent carboligase	1.7E-44	MTH_947	conserved protein	2.8E-40
MSM0432	Msp_1131	hypothetical membrane-spanning protein	5.5E-07	NONE		
MSM0433	Msp_1133	AhaD	1.6E-69	NONE	ATP synthase, subunit D	1.5E-73
MSM0434	Msp_1134	AhaB	1.4E-212	NONE	ATP synthase, subunit B	4.5E-214
MSM0435	Msp_1135	AhaA	1.4E-246	NONE	ATP synthase, subunit A	2.8E-260
MSM0436	Msp_1136	AhaF	8.6E-25	NONE	ATP synthase, subunit F	3.1E-25
MSM0437	Msp_1137	AhaC	1.5E-105	NONE	ATP synthase, subunit C	7.7E-116
MSM0438	Msp_1138	AhaE	3.2E-50	NONE	ATP synthase, subunit E	5.9E-54
MSM0439	Msp_1139	AhaK	7.0E-62	NONE	ATP synthase, subunit K	9.7E-70
MSM0440	Msp_1140	AhaI	1.9E-148	NONE	ATP synthase, subunit I	3.5E-191
MSM0441	Msp_1141	AhaH	7.6E-17	MTH_961	unknown	3.1E-18
MSM0442	NONE			NONE		
MSM0443	NONE			NONE		
MSM0444	NONE			NONE		
MSM0445	Msp_0408	putative nitroreductase protein	2.0E-55	MTH_120	NADPH-oxidoreductase	1.4E-13
MSM0446	NONE			MTH_962	citrate synthase I	6.2E-75
MSM0447	Msp_0338	fumarate hydratase	2.6E-15	NONE	fumarate hydratase, class I related protein	3.8E-75
MSM0448	NONE			MTH_964	unknown	4.6E-102
MSM0449	NONE			MTH_965	conserved protein	1.1E-86
MSM0450	Msp_0680	conserved hypothetical membrane-spanning protein	2.4E-38	NONE		
MSM0451	Msp_0679	conserved hypothetical membrane-spanning protein	7.8E-79	NONE		
MSM0452	Msp_1142	predicted DNA-binding protein	3.9E-132	MTH_966	conserved protein	1.8E-130
MSM0453	Msp_1143	putative transcriptional regulator	7.5E-58	MTH_967	conserved protein	1.3E-88
MSM0454	NONE			NONE		
MSM0455	Msp_1144	conserved hypothetical protein	2.2E-35	MTH_969	unknown	1.0E-43
MSM0456	Msp_1005	conserved hypothetical protein	2.3E-17	MTH_544	conserved protein	2.7E-35
MSM0457	Msp_1145	SerA	8.8E-158	MTH_970	phosphoglycerate dehydrogenase	1.3E-177
MSM0458	NONE			NONE		
MSM0459	NONE			NONE		
MSM0460	NONE			NONE		
MSM0461	Msp_0983	member of asn/thr-rich large protein family	3.0E-39	MTH_911	probable surface protein	2.9E-18
MSM0462	Msp_1146	partially conserved hypothetical protein	1.8E-38	MTH_971	unknown	1.0E-33
MSM0463	Msp_1147	conserved hypothetical protein	2.0E-57	MTH_972	conserved protein	3.7E-61
MSM0464	Msp_1148	predicted dinucleotide-utilizing protein	4.0E-59	MTH_973	conserved protein	1.1E-77
MSM0465	Msp_1149	conserved hypothetical protein	1.1E-17	MTH_974	unknown	4.1E-23
MSM0466	Msp_1150	predicted tRNA-binding protein	2.4E-68	MTH_975	conserved protein	1.4E-70
MSM0467	NONE			MTH_978	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	8.1E-137
MSM0468	NONE			MTH_1490	unknown	2.2E-10
MSM0469	NONE			MTH_1490	unknown	1.8E-11
MSM0470	Msp_1151	hypothetical membrane-spanning protein	1.4E-10	MTH_979	unknown	7.2E-10
MSM0471	Msp_1152	conserved hypothetical membrane-spanning protein	7.1E-53	MTH_980	conserved protein	5.9E-70
MSM0472	Msp_1153	PepQ	2.7E-69	MTH_981	aminopeptidase P	1.0E-65
MSM0473	Msp_0417	hypothetical membrane-spanning protein	2.5E-04	NONE		
MSM0474	NONE			NONE		
MSM0475	Msp_0417	hypothetical membrane-spanning protein	1.8E-04	NONE		

MSM0476	NONE			MTH_93	unknown	8.5E-04
MSM0477	NONE			NONE		
MSM0478	NONE			NONE		
MSM0479	Msp_1154	conserved hypothetical membrane-spanning protein	2.4E-45	MTH_986	conserved protein	2.1E-42
MSM0480	Msp_1155	conserved hypothetical protein	2.3E-95	MTH_987	conserved protein	6.0E-109
MSM0481	Msp_1274	conserved hypothetical protein	4.4E-53	MTH_989	conserved protein	2.2E-24
MSM0482	Msp_1275	predicted ATP-utilizing enzyme	4.6E-58	MTH_990	conserved protein	2.6E-51
MSM0483	NONE			MTH_991	unknown	8.6E-14
MSM0484	Msp_1276	conserved hypothetical protein	9.2E-76	MTH_992	inosine-5'-monophosphate dehydrogenase related protein IX	2.8E-86
MSM0485	Msp_1410	predicted universal stress protein	9.6E-26	MTH_993	conserved protein	1.0E-33
MSM0486	Msp_1199	predicted metal-dependent hydrolase	3.1E-84	MTH_994	N-ethylammeline chlorohydrolase related protein	4.2E-85
MSM0487	NONE			NONE		
MSM0488	Msp_1200	CarB	0.0E+00	NONE	carbamoyl-phosphate synthase, large subunit	0.0E+00
MSM0489	Msp_1201	CarA	1.5E-121	NONE	carbamoyl-phosphate synthase, small subunit	6.0E-125
MSM0490	Msp_0602	conserved hypothetical protein	1.0E-28	MTH_738	conserved protein	3.0E-06
MSM0491	Msp_0410	NadC	2.0E-64	MTH_1832	quinolinate phosphoribosyltransferase	7.7E-61
MSM0492	Msp_0411	putative ribonuclease Z	1.7E-76	MTH_1831	conserved protein	2.6E-92
MSM0493	Msp_0982	predicted mechanosensitive ion channel	6.7E-25	MTH_1830	conserved protein	1.7E-40
MSM0494	Msp_0643	NadA	3.6E-90	MTH_1827	quinolinate synthetase	6.8E-101
MSM0495	NONE			MTH_1821	unknown	2.7E-19
MSM0496	Msp_1526	putative homoserine O-acetyltransferase	1.2E-84	MTH_1820	homoserine O-acetyltransferase	1.3E-67
MSM0497	Msp_0157	hypothetical protein	6.9E-55	MTH_1816	conserved protein	2.6E-76
MSM0498	NONE			NONE		
MSM0499	Msp_1548	hypothetical protein	1.0E-05	MTH_1277	unknown	1.8E-06
MSM0500	Msp_0155	predicted amidohydrolase	3.1E-75	MTH_1811	N-carbamoyl-D-amino acid amidohydrolase	3.7E-77
MSM0501	Msp_0153	conserved hypothetical protein	1.8E-31	MTH_1806	phycocyanin alpha phycocyanobilin lyase CpcE	8.1E-34
MSM0502	Msp_0150	predicted helicase	2.9e-310	MTH_1802	ATP-dependent helicase	0.0E+00
MSM0503	Msp_0553	hypothetical protein	9.4E-19	MTH_1799	unknown	3.9E-18
MSM0504	Msp_0927	hypothetical protein	2.1E-05	MTH_1641	unknown	1.4E-06
MSM0505	NONE			NONE		
MSM0506	Msp_0240	predicted ATP-utilizing enzyme	3.0E-148	MTH_1201	conserved protein	3.4E-145
MSM0507	Msp_0365	predicted phosphoesterase	6.0E-49	MTH_1774	conserved protein	2.9E-52
MSM0508	Msp_0364	putative 23S rRNA methylase	1.9E-61	MTH_1773	cell division protein J	5.9E-70
MSM0509	Msp_0363	hypothetical membrane-spanning protein	1.4E-24	MTH_1772	unknown	9.1E-26
MSM0510	Msp_0362	predicted minichromosome maintenance protein	1.4E-255	MTH_1770	DNA replication initiator (Cdc21/Cdc54)	1.4E-260
MSM0511	Msp_0361	translation initiation factor aIF-2, beta subunit (eIF2B)	2.3E-54	NONE	translation initiation factor eIF-2, beta subunit	6.9E-60
MSM0512	Msp_0360	predicted NMD3-related protein	5.2E-73	MTH_1768	conserved protein	2.1E-90
MSM0513	Msp_0359	TyrS	2.4E-100	MTH_1767	tyrosyl-tRNA synthetase	1.1E-109
MSM0514	Msp_0358	hypothetical protein	3.5E-05	MTH_1766	unknown	1.1E-08
MSM0515	Msp_0186	MtaB2	1.3E-156	NONE		
MSM0516	Msp_0185	MtaC3	5.2E-89	NONE		
MSM0517	Msp_0190	MapA	8.7E-167	MTH_278	ferredoxin	7.0E-04
MSM0518	Msp_0112	MtaA2	2.1E-94	MTH_775	cobalamin-independent methionine synthase	3.4E-05
MSM0519	Msp_0183	hypothetical protein	1.2E-32	NONE		
MSM0520	Msp_0357	putative thymidylate kinase	2.1E-46	MTH_1765	thymidylate kinase	7.5E-47
MSM0521	NONE			NONE		
MSM0522	Msp_0984	predicted peptidase	2.7E-234	MTH_1763	collagenase	3.4E-99
MSM0523	Msp_0984	predicted peptidase	1.6E-96	MTH_1763	collagenase	6.8E-108
MSM0524	Msp_0354	MutS	4.3E-133	MTH_1762	DNA mismatch recognition protein MutS	1.9E-176
MSM0525	Msp_1282	predicted protein kinase	1.8E-104	MTH_1645	ABC transporter	3.1E-112

MSM0526	NONE			NONE		
MSM0527	Msp_0017	conserved hypothetical protein	3.5E-28	NONE		
MSM0528	Msp_0233	conserved hypothetical protein	1.4E-10	NONE		
MSM0529	Msp_0725	hypothetical protein	1.0E-04	NONE		
MSM0530	Msp_1323	conserved hypothetical protein	3.3E-04	MTH_72	O-linked GlcNAc transferase	5.5E-06
MSM0531	NONE			NONE		
MSM0532	Msp_0233	conserved hypothetical protein	3.4E-08	NONE		
MSM0533	Msp_0017	conserved hypothetical protein	3.1E-16	NONE		
MSM0534	NONE			NONE		
MSM0535	Msp_0466	hypothetical protein	7.1E-05	NONE		
MSM0536	NONE			NONE		
MSM0537	NONE			NONE		
MSM0538	Msp_1324	predicted glycyl radical activating enzyme	5.1E-07	MTH_1586	pyruvate formate-lyase activating enzyme	1.3E-05
MSM0539	Msp_0219	conserved hypothetical protein	3.1E-04	NONE		
MSM0540	NONE			NONE		
MSM0541	NONE			NONE		
MSM0542	Msp_1128	F420-dependent N5,N10-methylenetetrahydromethanopterin reductase	3.4E-94	NONE	coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase	1.4E-132
MSM0543	Msp_0646	predicted DNA repair photolyase	9.3E-28	NONE		
MSM0544	Msp_1127	predicted Fe-S oxidoreductase	4.4E-92	MTH_1751	conserved protein	1.3E-90
MSM0545	NONE			NONE		
MSM0546	Msp_1046	hypothetical membrane-spanning protein	2.6E-23	MTH_813	unknown	2.4E-27
MSM0547	Msp_0324	predicted nucleotidyltransferase	1.6E-08	MTH_1749	unknown	7.2E-81
MSM0548	Msp_1148	predicted dinucleotide-utilizing protein	4.4E-04	MTH_1747	conserved protein	5.4E-37
MSM0549	Msp_0830	Trk-type potassium transport system, membrane protein	3.9E-04	MTH_1746	cytochrome C-type biogenesis protein	2.1E-28
MSM0550	Msp_0656	hypothetical membrane-spanning protein	2.0E-04	MTH_1745	protein disulphide isomerase	7.9E-20
MSM0551	Msp_1124	conserved hypothetical protein	1.9E-68	MTH_1744	conserved protein	2.4E-73
MSM0552	Msp_0330	hypothetical protein	4.6E-10	MTH_1743	unknown	8.9E-12
MSM0553	Msp_0331	predicted ATPase	3.5E-92	MTH_1742	conserved protein	1.2E-80
MSM0554	Msp_0161	conserved hypothetical protein	2.8E-74	MTH_1815	conserved protein	2.6E-83
MSM0555	Msp_0192	predicted MoxR-like ATPase	3.9E-93	MTH_1814	conserved protein	1.9E-87
MSM0556	Msp_0333	predicted pterin-binding enzyme	4.1E-121	MTH_1741	conserved protein	1.1E-153
MSM0557	Msp_0334	PorC	2.1E-53	NONE	pyruvate oxidoreductase, gamma subunit	2.1E-65
MSM0558	Msp_0335	PorD	4.3E-30	NONE	pyruvate oxidoreductase, gamma subunit	1.2E-32
MSM0559	Msp_0336	PorA	2.1E-140	NONE	pyruvate oxidoreductase, alpha subunit	2.3E-148
MSM0560	Msp_0337	PorB	1.8E-118	NONE	pyruvate oxidoreductase, beta subunit	2.2E-127
MSM0561	Msp_1447	EhbK	8.6E-08	NONE	formate hydrogenlyase, iron-sulfur subunit I	4.5E-40
MSM0562	Msp_1447	EhbK	4.0E-09	NONE	formate hydrogenlyase, iron-sulfur subunit 2	5.3E-14
MSM0563	Msp_0338	fumarate hydratase	3.3E-96	NONE	fumarate hydratase, class I	8.3E-96
MSM0564	Msp_0339	predicted phosphate uptake regulator	4.8E-31	MTH_1734	phosphate transport system regulator	2.8E-47
MSM0565	Msp_0340	PstB	4.0E-107	MTH_1731	phosphate transport system ATP-binding	1.5E-105
MSM0566	Msp_0341	PstA	1.3E-94	MTH_1730	phosphate transporter permease PstC homolog	4.5E-111
MSM0567	Msp_0342	PstC	7.0E-94	MTH_1729	phosphate transporter permease PstC	4.8E-100
MSM0568	Msp_0343	PstS	1.6E-64	MTH_1727	phosphate-binding protein PstS	2.7E-81
MSM0569	Msp_0344	predicted phosphate uptake regulator	5.5E-62	MTH_1724	phosphate transport system regulator related protein	2.4E-82
MSM0570	Msp_0346	conserved hypothetical membrane-spanning protein	5.2E-17	MTH_1723	unknown	9.1E-26
MSM0571	NONE			MTH_1137	conserved protein (FlpA)	5.2E-165
MSM0572	NONE			NONE	H(2)-dependent N5,N10-methylenetetrahydromethanopterin dehydrogenase	2.4E-128
MSM0573	Msp_0296	CofG	1.4E-15	MTH_1143	biotin synthetase (BioB)	5.1E-112
MSM0574	NONE			MTH_1144	conserved protein	2.9E-38

MSM0575	Msp_1393	conserved hypothetical membrane-spanning protein	8.5E-05	MTH_1145	conserved protein	2.9E-38
MSM0576	NONE			MTH_1146	conserved protein	2.9E-38
MSM0577	NONE			MTH_1147	conserved protein	6.1E-52
MSM0578	NONE			MTH_1148	conserved protein	8.1E-34
MSM0579	Msp_1581	partially conserved hypothetical protein	7.5E-10	MTH_1106	ferredoxin	1.3E-10
MSM0580	Msp_0911	member of asn/thr-rich large protein family	2.5E-05	MTH_654	unknown	5.2E-39
MSM0581	Msp_0166	conserved hypothetical membrane-spanning protein	3.9E-29	MTH_655	conserved protein	6.7E-94
MSM0582	Msp_0737	putative peptide methionine sulfoxide reductase MsrA/MsrB	4.5E-122	MTH_535	peptide methionine sulfoxide reductase	2.4E-34
MSM0583	Msp_0655	CbiM2	2.7E-69	MTH_1707	cobalamin biosynthesis protein M	1.5E-64
MSM0584	Msp_0656	hypothetical membrane-spanning protein	2.2E-12	MTH_1706	unknown	3.4E-12
MSM0585	Msp_0657	CbiQ2	5.4E-55	MTH_1705	cobalt transport membrane protein	4.2E-60
MSM0586	Msp_0401	CbiO1	7.6E-81	MTH_1704	cobalt transport ATP-binding protein O	1.2E-85
MSM0587	Msp_1438	hypothetical protein	5.9E-10	NONE		
MSM0588	Msp_1441	FeoA	1.7E-12	MTH_1362	unknown	2.4E-11
MSM0589	Msp_1440	FeoB	3.6E-200	MTH_1361	ferrous iron transport protein B	5.7E-152
MSM0590	NONE			NONE		
MSM0591	NONE			NONE		
MSM0592	Msp_0202	conserved hypothetical membrane-spanning protein	2.3E-40	MTH_230	unknown	1.2E-48
MSM0593	Msp_0610	predicted ABC-type multidrug transport system, ATP-binding protein	3.9E-77	MTH_1487	ABC transporter (ATP-binding	2.0E-37
MSM0594	Msp_0609	conserved hypothetical membrane-spanning protein	2.7E-44	NONE		
MSM0595	Msp_0609	conserved hypothetical membrane-spanning protein	1.8E-40	NONE		
MSM0596	Msp_1163	predicted type II secretion protein F	3.0E-47	MTH_1703	unknown	4.9E-59
MSM0597	Msp_1162	predicted type II/IV secretion protein	4.1E-121	MTH_1702	secretory protein kinase	2.9E-157
MSM0598	Msp_1161	conserved hypothetical protein	3.5E-44	MTH_1701	unknown	5.6E-42
MSM0599	Msp_1160	conserved hypothetical membrane-spanning protein	1.3E-94	MTH_1700	conserved protein	8.9E-99
MSM0600	Msp_0512	predicted transcriptional regulator	7.9E-15	MTH_313	transcriptional regulator	5.5E-12
MSM0601	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0602	Msp_1159	elongation factor 1-beta (aEF-1beta) (ef1B)	2.2E-26	MTH_1699	translation elongation factor EF-1b	1.3E-28
MSM0603	Msp_1158	predicted Zn-ribbon RNA-binding protein	4.7E-17	MTH_1178	conserved protein	8.3E-04
MSM0604	Msp_1157	predicted amino acid kinase	1.7E-42	MTH_1698	delta 1-pyrroline-5-carboxylate synthetase	6.2E-43
MSM0605	Msp_1156	putative peptidyl-tRNA hydrolase	1.5E-29	MTH_1697	conserved protein	1.1E-36
MSM0606	NONE			NONE		
MSM0607	Msp_0613	predicted ATPase	4.1E-224	MTH_1695	RNase L inhibitor	6.8E-227
MSM0608	NONE			NONE		
MSM0609	Msp_0147	ferredoxin	2.6E-04	MTH_221	unknown	6.4E-25
MSM0610	Msp_0370	putative aspartate aminotransferase	8.5E-121	MTH_1694	aspartate aminotransferase related protein	9.6E-134
MSM0611	Msp_0369	RadB	3.9E-61	MTH_1693	DNA repair protein Rad51 homolog	3.6E-63
MSM0612	Msp_0096	conserved hypothetical protein	1.9E-36	MTH_1692	conserved protein	3.8E-43
MSM0613	Msp_0095	predicted phosphatidylglycerophosphate synthase	1.0E-46	MTH_1691	conserved protein	4.3E-44
MSM0614	Msp_0094	conserved hypothetical protein	2.1E-14	MTH_1690	unknown	1.7E-17
MSM0615	Msp_0675	conserved hypothetical protein	4.7E-159	MTH_1686	conserved protein	7.7E-164
MSM0616	Msp_0440	member of asn/thr-rich large protein family	1.1E-93	MTH_716	cell surface glycoprotein (s-layer protein)	1.4E-14
MSM0617	Msp_0160	ThiI	1.4E-102	MTH_1685	conserved protein	1.1E-118
MSM0618	Msp_1489	predicted potassium transport system, membrane component	3.0E-09	MTH_760	Na+/H+-exchanging protein:Na+/H+ antiporter	2.3E-16
MSM0619	Msp_1262	AlaS	7.0E-300	MTH_1683	alanyl-tRNA synthetase	1.5e-316
MSM0620	Msp_1263	50S ribosomal protein L12P	1.9E-36	MTH_1682	ribosomal protein Lp1	9.4E-40
MSM0621	Msp_1264	50S ribosomal protein L10P	5.3E-96	MTH_1681	ribosomal protein Lp0 (E.coli)	2.7E-106
MSM0622	Msp_1265	50S ribosomal protein L1P	9.5E-74	MTH_1680	ribosomal protein L10a (E.coli)	1.3E-81

MSM0623	Msp_1266	50S ribosomal protein L11P	1.3E-62	MTH_1679	ribosomal protein L12 (E.coli)	2.2E-63
MSM0624	Msp_1267	putative transcription antiterminator	1.3E-46	MTH_1678	transcription termination factor NusG	1.1E-61
MSM0625	Msp_1268	partially conserved hypothetical membrane-spanning protein	1.3E-12	MTH_1677	protein translocation complex sec61 gamma subunit related protein	1.1E-13
MSM0626	Msp_1269	FtsZ	8.7E-135	MTH_1676	cell division protein FtsZ	1.7E-143
MSM0627	Msp_0307	MtrH	8.5E-105	MTH_1156	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H	3.7E-116
MSM0628	NONE			MTH_1675	conserved protein	7.2E-49
MSM0629	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0630	Msp_1271	conserved hypothetical protein	7.1E-69	MTH_1670	conserved protein	4.2E-76
MSM0631	Msp_1272	predicted transcription initiation factor IIE, alpha subunit	3.4E-37	MTH_1669	conserved protein	4.6E-47
MSM0632	Msp_1273	conserved hypothetical protein	6.2E-38	MTH_1668	conserved protein	1.7E-40
MSM0633	Msp_1063	predicted RNA-binding protein	9.2E-92	MTH_1665	conserved protein	6.9E-92
MSM0634	Msp_1064	conserved hypothetical protein	1.8E-24	MTH_1664	conserved protein	6.2E-27
MSM0635	Msp_1069	predicted regulator of aminoacid metabolism	1.6E-41	MTH_1654	unknown	1.8E-45
MSM0636	Msp_1067	hypothetical protein	1.6E-23	MTH_1649	hydrogenase expression/formation protein HypC	1.2E-25
MSM0637	Msp_1077	predicted dihydrolipoamide dehydrogenase-related protein	2.4E-93	MTH_1648	dihydrolipoamide dehydrogenase	1.2E-92
MSM0638	Msp_1343	hypothetical membrane-spanning multicopy protein A 3	2.6E-78	MTH_1646	unknown	5.9E-54
MSM0639	Msp_1080	conserved hypothetical membrane-spanning protein	4.5E-67	MTH_1644	unknown	1.8E-52
MSM0640	Msp_1081	predicted release factor aRF1	2.2E-106	MTH_1642	cell division protein	9.6E-118
MSM0641	Msp_1083	putative prephenate dehydrogenase	4.4E-92	MTH_1640	chorismate mutase	1.8E-100
MSM0642	Msp_1084	CdcH	9.3E-273	MTH_1639	cell division control protein Cdc48	4.7E-299
MSM0643	Msp_0227	conserved hypothetical protein	3.3E-71	MTH_1574	conserved protein	5.2E-78
MSM0644	Msp_0228	ThiC1	1.2E-144	MTH_1576	thiamine biosynthesis protein	3.2E-158
MSM0645	Msp_0258	ATP-dependent DNA ligase	1.1E-148	MTH_1580	DNA ligase	3.9E-176
MSM0646	Msp_0504	conserved hypothetical membrane-spanning protein	5.5E-30	NONE		
MSM0647	Msp_0259	hypothetical protein	3.8E-15	MTH_1581	conserved protein	4.8E-20
MSM0648	Msp_0263	predicted phosphomannomutase	1.2E-169	MTH_1584	phosphomannomutase	9.9E-171
MSM0649	Msp_0970	hypothetical membrane-spanning protein	3.5E-44	MTH_559	conserved protein	1.0E-06
MSM0650	Msp_0971	hypothetical protein	1.2E-36	MTH_1787	conserved protein	1.3E-07
MSM0651	Msp_1323	conserved hypothetical protein	1.5E-98	MTH_1585	O-linked GlcNAc transferase	1.9E-105
MSM0652	Msp_1324	predicted glycyl radical activating enzyme	6.3E-45	MTH_1586	pyruvate formate-lyase activating enzyme	1.5E-50
MSM0653	Msp_1326	HisC	2.5E-112	MTH_1587	histidinol-phosphate aminotransferase	1.2E-119
MSM0654	Msp_1325	predicted carbonic anhydrase/acetyltransferase	1.8E-47	MTH_1588	ferripyochelin binding protein	4.6E-47
MSM0655	Msp_1301	predicted nucleoside-diphosphate-sugar pyrophosphorylase	3.0E-134	MTH_1589	glucose-1-phosphate thymidylyltransferase homolog	8.1E-137
MSM0656	Msp_1300	predicted phosphomannomutase	9.7E-136	MTH_1590	phosphomannomutase	7.6E-141
MSM0657	Msp_1299	ApgM2	6.1E-150	MTH_1591	phosphonopyruvate decarboxylase	6.0E-148
MSM0658	NONE			NONE		
MSM0659	Msp_1298	conserved hypothetical membrane-spanning protein	4.8E-63	MTH_1592	conserved protein	1.1E-77
MSM0660	Msp_1568	conserved hypothetical membrane-spanning protein	3.9E-52	NONE		
MSM0661	Msp_1297	30S ribosomal protein S3Ae	3.2E-66	MTH_1593	ribosomal protein S3a	8.4E-71
MSM0662	Msp_0712	hypothetical membrane-spanning protein	8.9E-07	NONE		
MSM0663	Msp_1295	predicted iron-molybdenum cluster-binding protein	1.4E-08	MTH_1594	conserved protein	1.2E-16
MSM0664	Msp_0540	predicted multimeric flavodoxin	2.4E-22	MTH_1595	conserved protein	5.0E-57
MSM0665	Msp_0642	predicted purine nucleoside phosphorylase	7.4E-74	MTH_1596	methylthioadenosine phosphorylase	3.7E-77
MSM0666	Msp_0641	conserved hypothetical membrane-spanning protein	6.7E-176	MTH_1597	conserved protein	3.5E-184
MSM0667	Msp_0587	hypothetical membrane-spanning protein	1.8E-05	MTH_520	unknown	3.7E-13
MSM0668	Msp_0637	conserved hypothetical protein	4.9E-22	MTH_1598	conserved protein	5.8E-40
MSM0669	NONE			NONE		
MSM0670	NONE			NONE		

MSM0671	Msp_0635	cell division control protein 6-like 2	2.7E-108	MTH_1599	Cdc6 related protein	5.4E-131
MSM0672	Msp_0661	conserved hypothetical protein	1.4E-56	MTH_1600	conserved protein	7.0E-67
MSM0673	Msp_1557	conserved hypothetical membrane-spanning protein	5.1E-27	NONE		
MSM0674	NONE			NONE		
MSM0675	NONE			NONE		
MSM0676	Msp_1557	conserved hypothetical membrane-spanning protein	9.7E-33	NONE		
MSM0677	Msp_0662	putative aspartate aminotransferase	1.3E-131	MTH_1601	aspartate aminotransferase	7.3E-136
MSM0678	Msp_0505	conserved hypothetical membrane-spanning protein	8.1E-29	MTH_519	unknown	1.1E-20
MSM0679	Msp_0587	hypothetical membrane-spanning protein	8.1E-12	MTH_520	unknown	8.1E-34
MSM0680	Msp_0757	predicted ATPase	2.4E-109	NONE		
MSM0681	NONE			NONE		
MSM0682	NONE			NONE		
MSM0683	Msp_0380	hypothetical protein	3.1E-13	MTH_626	unknown	9.7E-22
MSM0684	Msp_0381	hypothetical membrane-spanning protein	1.2E-09	MTH_625	unknown	1.5E-04
MSM0685	NONE			NONE		
MSM0686	Msp_0605	predicted thiamine pyrophosphate-requiring enzyme	2.1E-94	NONE	acetolactate synthase, large subunit homolog	8.5E-94
MSM0687	Msp_0604	predicted deoxycytidine triphosphate deaminase	1.6E-57	MTH_1605	deoxycytidine-triphosphate deaminase related protein	8.2E-57
MSM0688	Msp_1409	predicted tautomerase	3.2E-11	MTH_1606	unknown	1.7E-08
MSM0689	NONE			NONE		
MSM0690	Msp_0767	predicted helicase	2.1E-243	NONE	ATP-dependent RNA helicase, eIF-4A family	9.5E-09
MSM0691	Msp_0006	predicted NUDIX-related protein	1.4E-40	MTH_1336	mutator MutT protein homolog	4.1E-14
MSM0692	NONE			NONE		
MSM0693	Msp_0113	conserved hypothetical protein	1.4E-13	MTH_540	intracellular protein transport protein	7.2E-10
MSM0694	NONE			NONE		
MSM0695	Msp_0767	predicted helicase	1.0E-13	NONE	ATP-dependent RNA helicase, eIF-4A family	3.7E-10
MSM0696	Msp_1095	DNA double-strand break repair protein Rad50	4.0E-04	NONE		
MSM0697	NONE			NONE		
MSM0698	NONE			NONE		
MSM0699	Msp_0738	predicted Na+-dependent transporter	4.1E-137	MTH_1909	unknown	5.8E-04
MSM0700	Msp_0921	putative poly-gamma-glutamate biosynthesis protein	1.0E-108	NONE		
MSM0701	Msp_0601	partially conserved hypothetical protein, predicted GTPase	2.4E-116	MTH_1608	signal recognition particle protein (docking protein)	3.6E-111
MSM0702	Msp_0600	conserved hypothetical protein	1.5E-20	MTH_1609	conserved protein	1.1E-36
MSM0703	Msp_0599	RplX	4.1E-18	MTH_1610	ribosomal protein L18a	1.0E-17
MSM0704	Msp_0598	translation initiation factor 6 (aIF-6)	3.7E-56	MTH_1611	conserved protein	3.8E-59
MSM0705	Msp_0597	50S ribosomal protein L31e	1.4E-22	MTH_1612	ribosomal protein L31	4.7E-29
MSM0706	NONE			MTH_1613	ribosomal protein L39	1.2E-16
MSM0707	Msp_0596	predicted subunit of tRNA methyltransferase	2.8E-58	MTH_1614	conserved protein	3.8E-59
MSM0708	Msp_0595	partially conserved hypothetical protein	1.4E-31	MTH_1615	conserved protein	3.1E-32
MSM0709	Msp_0594	30S ribosomal protein S19e	1.5E-52	MTH_1616	ribosomal protein S19	5.9E-54
MSM0710	Msp_0593	hypothetical protein	1.3E-28	MTH_1617	conserved protein	1.3E-19
MSM0711	Msp_0592	putative ribonuclease P, subunit 4	8.7E-32	MTH_1618	conserved protein	3.0E-34
MSM0712	NONE			NONE		
MSM0713	Msp_0589	predicted nucleotide kinase	3.1E-36	MTH_1619	conserved protein (adenylate kinase related)	2.4E-34
MSM0714	Msp_0660	predicted GTPase	2.1E-46	NONE	GTP-binding protein, GTP1/OBG family	3.9E-50
MSM0715	Msp_0660	predicted GTPase	2.4E-77	NONE	GTP-binding protein, GTP1/OBG family	1.2E-87
MSM0716	Msp_0368	conserved hypothetical membrane-spanning protein	1.1E-141	MTH_1623	oligosaccharyl transferase STT3 subunit related protein	7.3E-88
MSM0717	Msp_0366	TopA	8.0E-228	MTH_1624	DNA topoisomerase I	3.1E-247
MSM0718	NONE			MTH_1625	unknown	4.6E-15
MSM0719	Msp_1096	putative phosphoserine phosphatase	2.7E-124	MTH_1626	phosphoserine phosphatase	1.3E-83
MSM0720	Msp_1097	TATA-box binding protein	5.0E-68	MTH_1627	TATA-binding transcription initiation factor	1.2E-73

MSM0721	Msp_1098	predicted adenylate cyclase	2.6E-39	MTH_1629	conserved protein	1.3E-42
MSM0722	Msp_1099	LeuA2	1.9E-91	MTH_1630	2-isopropylmalate synthase	1.5E-151
MSM0723	Msp_1100	LeuC2	2.7E-140	NONE	3-isopropylmalate dehydratase, LeuC subunit	5.8E-150
MSM0724	Msp_0326	hypothetical protein	9.1E-04	MTH_1632	conserved protein	1.0E-40
MSM0725	Msp_1086	flap structure-specific endonuclease	9.2E-92	MTH_1633	DNA repair protein Rad2	7.8E-100
MSM0726	NONE			MTH_1635	conserved protein	7.1E-42
MSM0727	Msp_1085	AhcY	1.3E-163	MTH_1636	S-adenosylhomocysteine hydrolase	3.7E-164
MSM0728	Msp_0524	predicted oxidoreductase	4.4E-92	MTH_907	conserved protein	2.5E-62
MSM0729	Msp_0231	predicted E1-like enzyme	2.1E-46	MTH_1571	molybdopterin biosynthesis protein MoeB homolog	1.7E-65
MSM0730	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0731	Msp_0113	conserved hypothetical protein	1.6E-13	MTH_511	DNA helicase II	4.6E-07
MSM0732	Msp_0873	TruB	3.2E-105	MTH_32	centromere/microtubule-binding protein	3.2E-110
MSM0733	Msp_0880	50S ribosomal protein L14e	2.3E-24	MTH_31	ribosomal protein L14	4.1E-23
MSM0734	Msp_0881	putative cytidylate kinase	1.8E-56	MTH_30	cytidylate kinase	3.8E-52
MSM0735	Msp_0882	50S ribosomal protein L34e	2.4E-29	MTH_29	ribosomal protein L34 (E.coli)	3.3E-37
MSM0736	Msp_0883	hypothetical membrane-spanning protein	1.2E-34	MTH_28	conserved protein	1.1E-50
MSM0737	Msp_0884	AdkA	1.1E-61	MTH_27	adenylate kinase	1.1E-63
MSM0738	Msp_0885	SecY	6.6E-153	MTH_26	preprotein translocase SecY	1.0E-145
MSM0739	Msp_0886	50S ribosomal protein L15P	1.9E-43	MTH_25	ribosomal protein L27a (E.coli)	4.1E-46
MSM0740	Msp_0887	50S ribosomal protein L30P	9.7E-49	MTH_24	ribosomal protein L7 (E.coli)	1.2E-53
MSM0741	Msp_0888	30S ribosomal protein S5P	3.5E-92	MTH_23	ribosomal protein S2 (E.coli)	3.7E-93
MSM0742	Msp_0889	50S ribosomal protein L18P	6.7E-57	MTH_22	ribosomal protein L5	8.9E-67
MSM0743	Msp_0890	50S ribosomal protein L19e	4.6E-58	MTH_21	ribosomal protein L19	1.5E-64
MSM0744	Msp_0891	50S ribosomal protein L32e	6.6E-34	MTH_20	ribosomal protein L32	3.1E-41
MSM0745	Msp_0892	50S ribosomal protein L6P	5.7E-60	MTH_19	ribosomal protein L9 (E.coli)	4.3E-67
MSM0746	Msp_0893	30S ribosomal protein S8P	9.5E-58	MTH_18	ribosomal protein S15a (E.coli)	1.2E-55
MSM0747	Msp_0894	30S ribosomal protein S14P	2.1E-21	MTH_17	ribosomal protein S29 (E.coli)	7.6E-22
MSM0748	Msp_0895	50S ribosomal protein L5P	2.4E-61	MTH_16	ribosomal protein L11 (E.coli)	2.9E-61
MSM0749	Msp_0896	30S ribosomal protein S4e	3.0E-70	MTH_15	ribosomal protein S4	1.8E-77
MSM0750	Msp_0897	50S ribosomal protein L24P	2.4E-29	MTH_14	ribosomal protein L26 (E.coli)	1.3E-35
MSM0751	Msp_0898	50S ribosomal protein L14P	1.4E-56	MTH_13	ribosomal protein L23 (E.coli)	1.0E-56
MSM0752	Msp_0899	30S ribosomal protein S17P	1.4E-42	MTH_12	ribosomal protein S11 (E.coli)	1.4E-45
MSM0753	Msp_0900	putative ribonuclease P, component 1	4.8E-24	MTH_11	conserved protein	8.7E-21
MSM0754	Msp_0901	protein translation factor SUI1-like protein	2.4E-45	MTH_10	ribosomal protein SUI1	3.6E-47
MSM0755	Msp_0902	50S ribosomal protein L29P	3.3E-16	MTH_9	ribosomal protein L35 (E.coli)	7.9E-20
MSM0756	Msp_0903	30S ribosomal protein S3P	6.8E-96	MTH_8	ribosomal protein S3 (E.coli)	1.2E-96
MSM0757	Msp_0904	50S ribosomal protein L22P	1.3E-46	MTH_7	ribosomal protein L17 (E.coli)	3.5E-56
MSM0758	Msp_0905	30S ribosomal protein S19P	1.4E-58	MTH_6	ribosomal protein S15 (E.coli)	1.3E-58
MSM0759	Msp_0906	50S ribosomal protein L2P	3.1E-107	MTH_5	ribosomal protein L8 (E.coli)	1.9E-105
MSM0760	Msp_0907	50S ribosomal protein L23P	2.8E-26	MTH_4	ribosomal protein L23a (E.coli)	5.4E-28
MSM0761	Msp_0908	50S ribosomal protein L1e	4.5E-99	MTH_3	ribosomal protein L4 (E.coli)	2.6E-99
MSM0762	Msp_0909	50S ribosomal protein L3P	1.5E-121	MTH_2	ribosomal protein L3 (E.coli)	1.1E-132
MSM0763	Msp_0910	conserved hypothetical protein	1.1E-79	MTH_1	conserved protein	1.2E-73
MSM0764	Msp_1319	predicted DNA modification methylase	1.7E-04	MTH_1918	possible protein methyltransferase	3.7E-45
MSM0765	Msp_0914	PycA	1.7E-186	MTH_1917	biotin carboxylase	5.5E-202
MSM0766	Msp_0915	partially conserved hypothetical protein	4.0E-36	MTH_1916	biotin acetyl-CoA carboxylase ligase/biotin operon repressor	5.3E-62
MSM0767	Msp_0916	predicted selenocysteine synthase	2.8E-99	MTH_1914	conserved protein	2.3E-100
MSM0768	Msp_0917	hypothetical protein	7.5E-04	MTH_1912	unknown	1.1E-11
MSM0769	Msp_0791	fumarate hydratase	3.1E-59	NONE	fumarate hydratase, class I related protein	1.5E-50

MSM0770	Msp_1112	CbiO2	1.2E-43	NONE	methyl coenzyme M reductase system, component A2 homolog	8.3E-64
MSM0771	Msp_0657	CbiQ2	1.4E-05	MTH_453	conserved protein	2.6E-12
MSM0772	NONE			MTH_452	unknown	9.2E-07
MSM0773	Msp_0958	predicted ABC-type polar amino acid transport system, ATP-binding protein	1.4E-26	MTH_1704	cobalt transport ATP-binding protein O	5.9E-25
MSM0774	Msp_0340	PstB	1.6E-26	MTH_1731	phosphate transport system ATP-binding	5.2E-26
MSM0775	Msp_0149	predicted transcriptional regulator	2.0E-34	NONE		
MSM0776	Msp_0790	conserved hypothetical membrane-spanning protein	2.2E-138	MTH_1909	unknown	2.8E-159
MSM0777	Msp_0491	hypothetical membrane-spanning protein	3.6E-10	MTH_1908	unknown	3.2E-16
MSM0778	Msp_0517	predicted RNA-binding protein	3.6E-184	MTH_1907	conserved protein	2.0E-188
MSM0779	Msp_0516	predicted Zn-dependent hydrolase of the beta-lactamase superfamily	2.3E-70	MTH_1902	conserved protein	3.5E-72
MSM0780	NONE			MTH_1901	unknown	2.9E-16
MSM0781	Msp_1151	hypothetical membrane-spanning protein	1.2E-09	MTH_1533	unknown	1.3E-10
MSM0782	Msp_1151	hypothetical membrane-spanning protein	2.4E-04	MTH_979	unknown	1.2E-05
MSM0783	Msp_1447	EhbK	3.3E-20	NONE	tungsten formylmethanofuran dehydrogenase, subunit F homolog	3.5E-88
MSM0784	Msp_0236	ferredoxin	5.5E-14	MTH_927	ferredoxin	5.1E-16
MSM0785	Msp_0514	putative phosphopantetheine adenylyltransferase	1.0E-37	MTH_1896	conserved protein	1.3E-42
MSM0786	Msp_1129	partially conserved hypothetical membrane-spanning protein	1.1E-49	MTH_412	conserved protein	1.3E-69
MSM0787	Msp_0511	predicted Fe-S oxidoreductase	7.6E-120	MTH_1895	conserved protein	8.7E-124
MSM0788	Msp_0510	putative aspartate aminotransferase	5.5E-117	MTH_1894	aspartate aminotransferase homolog	3.3E-108
MSM0789	Msp_0519	predicted Co/Zn/Cd cation transporter	7.6E-33	MTH_1893	cation efflux system protein (zinc/cadmium)	1.8E-77
MSM0790	Msp_1428	conserved hypothetical protein	1.3E-15	MTH_1884	conserved protein	3.0E-36
MSM0791	Msp_0443	2-phosphoglycerate kinase	3.6E-81	MTH_1883	2-phosphoglycerate kinase	3.7E-84
MSM0792	Msp_1010	predicted phosphoesterase	1.8E-47	MTH_1882	conserved protein	2.3E-52
MSM0793	Msp_1011	conserved hypothetical protein	1.9E-29	MTH_1881	conserved protein	4.4E-42
MSM0794	Msp_1012	conserved hypothetical protein	1.9E-20	MTH_1880	conserved protein	2.1E-28
MSM0795	Msp_1013	HdrB1	1.9E-116	NONE	heterodisulfide reductase, subunit B	4.3E-115
MSM0796	Msp_1014	HdrC1	1.6E-69	NONE	heterodisulfide reductase, subunit C	4.7E-77
MSM0797	Msp_1015	conserved hypothetical protein	2.5E-50	MTH_1877	conserved protein	1.6E-53
MSM0798	NONE			NONE		
MSM0799	Msp_0113	conserved hypothetical protein	1.6E-12	MTH_1626	phosphoserine phosphatase	2.2E-06
MSM0800	NONE			NONE		
MSM0801	Msp_1017	DphB	1.7E-74	MTH_1874	diphthine synthase	2.9E-77
MSM0802	Msp_1022	predicted methyltransferase	3.6E-81	MTH_1873	met-10+ protein	1.3E-74
MSM0803	NONE			MTH_633	conserved protein	4.3E-04
MSM0804	Msp_1023	putative translation initiation factor aIF-2B, subunit 1	5.0E-100	NONE	translation initiation factor eIF-2B, alpha subunit	2.2E-125
MSM0805	Msp_0958	predicted ABC-type polar amino acid transport system, ATP-binding protein	5.0E-100	MTH_696	ABC transporter (glutamine transport ATP-binding protein)	2.7E-35
MSM0806	Msp_0959	predicted ABC-type polar amino acid transport system, permease protein	2.1E-92	NONE		
MSM0807	Msp_0960	predicted ABC-type polar amino acid transport system, periplasmic substrate-binding protein	3.5E-108	NONE		
MSM0808	Msp_1024	conserved hypothetical protein	2.9E-104	MTH_1871	nitrogenase iron-molybdenum cofactor biosynthesis protein NifB	1.6E-115
MSM0809	Msp_1025	conserved hypothetical protein	2.3E-40	MTH_1870	conserved protein	3.1E-41
MSM0810	Msp_1026	predicted activator of 2-hydroxyglutaryl-CoA dehydratase	5.5E-165	MTH_1869	activator of (R)-2-hydroxyglutaryl-CoA	1.7E-175
MSM0811	Msp_1027	conserved hypothetical protein	1.7E-53	MTH_1868	conserved protein	1.2E-57

MSM0812	Msp_1029	conserved hypothetical protein	1.3E-39	MTH_1866	conserved protein	1.0E-40
MSM0813	Msp_1030	predicted peptidyl-prolyl cis-trans isomerase	2.6E-135	MTH_1865	conserved protein	2.3E-146
MSM0814	Msp_1032	predicted selenophosphate synthetase-related protein	3.3E-87	MTH_1864	phosphoribosylformylglycinamidine synthase II related protein	6.2E-91
MSM0815	Msp_1033	conserved hypothetical protein	4.5E-99	MTH_1863	conserved protein	4.4E-97
MSM0816	Msp_1034	predicted nucleic acid-binding protein	3.7E-33	MTH_1862	conserved protein	3.5E-40
MSM0817	Msp_0799	predicted transcriptional regulator	6.6E-34	MTH_1843	unknown	1.0E-33
MSM0818	Msp_0798	predicted transcriptional regulator	5.0E-36	MTH_1843	unknown	2.1E-26
MSM0819	NONE			MTH_1438	unknown	4.6E-15
MSM0820	NONE			MTH_1861	molybdenum cofactor biosynthesis MoaB	2.5E-46
MSM0821	Msp_1036	PyrE	3.1E-59	MTH_1860	uridine 5'-monophosphate synthase	5.2E-55
MSM0822	Msp_1035	hypothetical protein	3.1E-13	MTH_1859	unknown	1.4E-15
MSM0823	NONE			NONE		
MSM0824	NONE			NONE	N-terminal acetyltransferase complex, subunit ARD1	3.1E-06
MSM0825	Msp_0437	conserved hypothetical protein	4.7E-56	NONE		
MSM0826	Msp_0114	ThsB	8.2E-226	MTH_794	chaperonin	2.4E-231
MSM0827	Msp_0747	member of asn/thr-rich large protein family	5.9E-04	MTH_796	conserved protein	4.5E-33
MSM0828	Msp_0220	predicted glycosyltransferase	2.0E-14	MTH_540	intracellular protein transport protein	8.1E-06
MSM0829	Msp_0110	aspartate-semialdehyde dehydrogenase	6.6E-121	MTH_799	aspartate-semialdehyde dehydrogenase	2.3E-132
MSM0830	Msp_0109	DapB	1.0E-85	MTH_800	dihydrodipicolinate reductase	3.2E-87
MSM0831	Msp_0108	DapA	4.9E-86	MTH_801	dihydrodipicolinate synthase	2.0E-85
MSM0832	Msp_0107	putative aspartokinase	2.2E-129	MTH_802	aspartokinase II alpha subunit	6.7E-149
MSM0833	Msp_0106	30S ribosomal protein S17e	1.3E-19	MTH_803	ribosomal protein S17	1.5E-23
MSM0834	Msp_0105	putative chorismate mutase	3.8E-15	NONE	chorismate mutase, subunit A	9.3E-17
MSM0835	Msp_0104	AroK	4.7E-56	MTH_805	conserved protein (homoserine kinase related)	2.6E-76
MSM0836	Msp_0101	predicted glycosyltransferase	2.6E-64	MTH_450	LPS biosynthesis RfbU related protein	9.6E-31
MSM0837	Msp_0102	CbiD	6.5E-91	MTH_808	cobalamin biosynthesis protein D	4.0E-87
MSM0838	Msp_0103	putative thioredoxin	2.5E-18	MTH_807	thioredoxin	7.1E-19
MSM0839	Msp_0100	predicted helicase	2.1E-227	MTH_810	DNA helicase related protein	9.1E-248
MSM0840	Msp_0097	conserved hypothetical protein	3.0E-15	MTH_814	conserved protein	1.6E-14
MSM0841	Msp_0371	hypothetical protein	6.6E-11	MTH_815	unknown	2.2E-15
MSM0842	Msp_0372	predicted histone acetyltransferase	1.5E-187	MTH_817	conserved protein	6.2E-189
MSM0843	NONE			MTH_818	deoxyribose-phosphate aldolase	2.1E-26
MSM0844	Msp_0122	archaeal histone	3.5E-21	MTH_821	histone HMtA1	2.5E-23
MSM0845	Msp_0376	predicted 2-methylthioadenine synthetase	8.9E-126	MTH_826	conserved protein	3.8E-130
MSM0846	Msp_0375	conserved hypothetical protein	1.6E-39	MTH_828	conserved protein	1.6E-46
MSM0847	Msp_0374	LeuD2	4.1E-57	NONE	3-isopropylmalate dehydratase, LeuD subunit	7.4E-56
MSM0848	Msp_0373	predicted archaeal sugar kinase	1.5E-73	MTH_830	conserved protein	3.0E-82
MSM0849	Msp_0384	predicted Fe-S oxidoreductase	6.6E-169	MTH_831	molybdenum cofactor biosynthesis MoaA homolog	2.7E-177
MSM0850	Msp_0385	conserved hypothetical membrane-spanning protein	2.4E-45	MTH_832	conserved protein	1.4E-43
MSM0851	Msp_0386	predicted transcriptional regulator	1.1E-70	MTH_834	conserved protein	3.0E-98
MSM0852	Msp_0387	predicted ATP-utilizing enzyme	2.3E-40	MTH_835	conserved protein	1.0E-53
MSM0853	Msp_0217	predicted UDP-N-acetylglucosamine 2-epimerase	1.4E-120	MTH_837	UDP-N-acetylglucosamine 2-epimerase	1.3E-136
MSM0854	NONE			NONE		
MSM0855	Msp_0388	TruA	5.2E-50	MTH_840	pseudouridylate synthase I	1.6E-51
MSM0856	NONE			MTH_695	conserved protein	1.7E-08
MSM0857	Msp_1000	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, ATB-binding protein	1.5E-29	MTH_696	ABC transporter (glutamine transport ATP-binding protein)	3.3E-44
MSM0858	Msp_0389	HisA	6.3E-77	MTH_843	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	7.4E-79
MSM0859	Msp_0390	putative cytidylyltransferase	5.1E-43	MTH_844	autotrophic growth protein	1.5E-48

MSM0860	Msp_0552	ArgC	4.9E-109	MTH_846	N-acetyl-gamma-glutamyl-phosphate reductase	2.0E-108
MSM0861	Msp_0554	hypothetical protein	4.8E-31	MTH_847	unknown	3.3E-44
MSM0862	Msp_0521	PyrI	2.1E-44	MTH_850	aspartate carbamoyltransferase regulatory subunit	7.5E-47
MSM0863	Msp_1419	hypothetical protein	3.1E-20	NONE		
MSM0864	NONE			MTH_1285	conserved protein	2.7E-10
MSM0865	Msp_0159	conserved hypothetical protein	1.1E-79	MTH_853	conserved protein	2.4E-96
MSM0866	Msp_0402	predicted zinc metalloprotease	4.7E-143	MTH_856	zinc metalloproteinase	8.2E-144
MSM0867	Msp_0403	conserved hypothetical protein	1.1E-47	MTH_857	conserved protein	4.0E-48
MSM0868	NONE			NONE		
MSM0869	Msp_0404	predicted GTPase	3.0E-93	NONE	GTP-binding protein, GTP1/OBG family	8.2E-112
MSM0870	Msp_0405	putative small heat shock protein	1.2E-16	NONE	heat shock protein, class I	3.8E-20
MSM0871	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM0872	Msp_1054	predicted phosphosugar isomerase	1.2E-103	MTH_860	glucosamine--fructose-6-phosphate aminotransferase	5.6E-113
MSM0873	Msp_1309	conserved hypothetical protein	7.6E-17	MTH_863	conserved protein	5.4E-28
MSM0874	Msp_1308	adenine deaminase	1.5E-139	MTH_866	adenine deaminase	1.3E-132
MSM0875	Msp_1347	conserved hypothetical protein	6.0E-136	MTH_867	conserved protein	6.4E-144
MSM0876	Msp_0415	predicted arginase/agmatinase/formimionoglutamate hydrolase	1.3E-71	MTH_868	agmatine ureohydrolase	1.2E-73
MSM0877	Msp_1352	translation initiation factor 5A (aIF-5A)	4.4E-53	NONE	translation initiation factor, eIF-5A	1.7E-49
MSM0878	Msp_1327	PdaD	2.1E-37	MTH_870	conserved protein	3.4E-42
MSM0879	Msp_1330	PpnK	7.2E-60	MTH_872	conserved protein	9.0E-77
MSM0880	Msp_1331	predicted UDP-N-acetylmuramyl pentapeptide synthase	1.1E-47	MTH_873	UDP-N-acetylmuramyl tripeptide synthetase related protein	5.4E-81
MSM0881	Msp_1332	HemC	7.3E-83	MTH_874	porphobilinogen deaminase	2.0E-85
MSM0882	Msp_1333	predicted dehydrogenase	2.7E-101	NONE	3-chlorobenzoate-3,4-dioxygenase dyhydrogenase related protein	3.0E-130
MSM0883	Msp_1334	predicted orotate phosphoribosyltransferase	5.6E-53	MTH_876	orotate phosphoribosyltransferase	9.7E-70
MSM0884	Msp_0747	member of asn/thr-rich large protein family	1.5E-18	MTH_716	cell surface glycoprotein (s-layer protein)	4.1E-07
MSM0885	Msp_1465	member of asn/thr-rich large protein family	2.4E-39	MTH_716	cell surface glycoprotein (s-layer protein)	1.7E-08
MSM0886	NONE			NONE		
MSM0887	Msp_1410	predicted universal stress protein	2.5E-18	MTH_898	conserved protein	1.5E-18
MSM0888	Msp_1416	GdhA	2.6E-181	NONE		
MSM0889	NONE			NONE		
MSM0890	NONE			NONE		
MSM0891	Msp_1363	peptide chain release factor, subunit 1 (aRF-1)	3.4E-149	NONE	peptide chain release factor eRF, subunit 1	8.7E-156
MSM0892	Msp_1056	hypothetical membrane-spanning protein	5.4E-06	MTH_1905	unknown	3.2E-06
MSM0893	Msp_1202	predicted acetyltransferase	2.4E-29	NONE	N-terminal acetyltransferase complex, subunit ARD1	3.7E-38
MSM0894	Msp_1203	conserved hypothetical protein	5.7E-28	MTH_1000	conserved protein	1.2E-25
MSM0895	Msp_1204	predicted cation transport ATPase	3.9E-235	MTH_1001	cation-transporting P-ATPase PacL	9.8E-251
MSM0896	Msp_1205	CbiJ	6.5E-43	MTH_1002	cobalamin biosynthesis protein J	8.5E-39
MSM0897	Msp_1365	30S ribosomal protein S10P	1.6E-48	MTH_1059	ribosomal protein S20 (E.coli)	1.3E-49
MSM0898	Msp_1366	translation elongation factor 1-alpha (EF-Tu)	1.9E-185	NONE	translation elongation factor, EF-1 alpha	3.9E-192
MSM0899	Msp_1367	FusA	1.7e-319	NONE	translation elongation factor, EF-2	1.9e-318
MSM0900	Msp_1368	30S ribosomal protein S7P	3.3E-80	MTH_1056	ribosomal protein S5 (E.coli)	9.2E-81
MSM0901	Msp_1369	30S ribosomal protein S12P	4.4E-69	MTH_1055	ribosomal protein S23 (E.coli)	7.8E-68
MSM0902	Msp_0321	MrtA	5.7E-250	NONE	methyl coenzyme M reductase II, alpha subunit	2.0E-250
MSM0903	Msp_0320	MrtG	1.6E-103	NONE	methyl coenzyme M reductase II, gamma subunit	1.8E-116
MSM0904	Msp_0319	MrtD	1.9E-45	NONE	methyl coenzyme M reductase II, D protein	2.2E-40
MSM0905	Msp_0318	MrtB	9.8E-159	NONE	methyl coenzyme M reductase II, beta subunit	4.1E-181
MSM0906	Msp_1370	NusA	1.7E-44	MTH_1054	transcription termination factor NusA	2.5E-55
MSM0907	Msp_1371	50S ribosomal protein L30e	6.0E-33	MTH_1053	ribosomal protein L30	3.0E-36
MSM0908	Msp_1372	RpoA2	2.1E-126	NONE	DNA-dependent RNA polymerase, subunit A"	4.7E-141

MSM0909	Msp_1373	RpoA1	0.0E+00	NONE	DNA-dependent RNA polymerase, subunit A'	0.0E+00
MSM0910	Msp_1374	RpoB1	6.1E-253	NONE	DNA-dependent RNA polymerase, subunit B'	4.6E-276
MSM0911	Msp_1375	RpoB2	3.3E-103	NONE	DNA-dependent RNA polymerase, subunit B''	8.6E-220
MSM0912	Msp_1376	RpoH	7.6E-17	NONE	DNA-dependent RNA polymerase, subunit H	4.6E-15
MSM0913	NONE			NONE		
MSM0914	NONE			MTH_72	O-linked GlcNAc transferase	3.0E-04
MSM0915	NONE			NONE		
MSM0916	Msp_0682	ThiM1	1.2E-73	NONE		
MSM0917	Msp_0683	hypothetical protein	7.7E-56	NONE		
MSM0918	Msp_1381	phosphoglycerate kinase	1.1E-120	MTH_1042	3-phosphoglycerate kinase	4.3E-131
MSM0919	Msp_1382	TpiA	4.9E-77	MTH_1041	triosephosphate isomerase	3.2E-71
MSM0920	Msp_1103	member of asn/thr-rich large protein family	4.2E-04	NONE		
MSM0921	Msp_0548	hypothetical membrane-spanning protein	1.1E-05	NONE		
MSM0922	Msp_1383	predicted Fe-S oxidoreductase	1.7E-97	MTH_1039	conserved protein	4.9E-98
MSM0923	Msp_0540	predicted multimeric flavodoxin	1.2E-16	MTH_135	conserved protein	1.3E-17
MSM0924	Msp_1386	SucC	3.4E-101	NONE	succinyl-CoA synthetase, beta subunit	3.7E-116
MSM0925	Msp_1387	KorC	9.5E-58	NONE	2-oxoglutarate oxidoreductase, gamma subunit	8.8E-60
MSM0926	Msp_1388	KorB	1.3E-99	NONE	2-oxoglutarate oxidoreductase, beta subunit	2.2E-102
MSM0927	Msp_1389	KorA	4.5E-138	NONE	2-oxoglutarate oxidoreductase, alpha subunit	6.2E-130
MSM0928	Msp_1390	KorD	3.0E-15	NONE	ferredoxin (putative 2-oxoglutarate oxidoreductase, delat subunit)	8.6E-14
MSM0929	Msp_0791	fumarate hydratase	3.7E-17	NONE	fumarate hydratase, class I	3.5E-40
MSM0930	Msp_0325	predicted peptidyl-prolyl cis-trans isomerase 2	3.5E-67	MTH_1125	fkbp-type peptidyl-prolyl cis-trans isomerase	1.8E-77
MSM0931	Msp_0801	conserved hypothetical protein	7.0E-94	MTH_448	unknown	4.8E-68
MSM0932	Msp_1167	conserved hypothetical protein	4.7E-49	MTH_1113	conserved protein	1.6E-58
MSM0933	Msp_1168	CobS	1.2E-50	MTH_1112	cobalamin (5'-phosphate) synthase	1.9E-41
MSM0934	Msp_1169	hypothetical protein	1.1E-06	MTH_1111	conserved protein	1.5E-41
MSM0935	Msp_1170	conserved hypothetical protein	4.5E-106	MTH_1109	conserved protein	4.2E-92
MSM0936	Msp_1171	predicted ATPase	6.3E-77	MTH_1108	conserved protein	1.0E-65
MSM0937	NONE			NONE		
MSM0938	NONE			NONE		
MSM0939	Msp_1173	PycB	1.4E-212	NONE	oxaloacetate decarboxylase, alpha subunit	2.8E-221
MSM0940	Msp_1166	predicted myo-inositol-1-phosphate synthase	5.3E-151	MTH_1105	conserved protein	9.4E-159
MSM0941	Msp_0634	predicted prenyltransferase	2.3E-70	MTH_1098	bacteriochlorophyll synthase related protein	4.2E-69
MSM0942	Msp_0616	partially conserved hypothetical membrane-spanning protein	5.0E-52	MTH_371	unknown	5.1E-35
MSM0943	NONE			MTH_466	unknown	5.6E-09
MSM0944	NONE			NONE		
MSM0945	Msp_1285	hydrogenase expression/formation protein	9.3E-147	MTH_1072	hydrogenase expression/formation protein HypD	2.2E-141
MSM0946	Msp_0215	predicted glycosyltransferase	6.1E-04	MTH_1071	conserved protein	3.9E-50
MSM0947	Msp_1284	predicted modulator of DNA gyrase	3.7E-95	MTH_1070	conserved protein	1.5E-96
MSM0948	Msp_0220	predicted glycosyltransferase	4.0E-04	NONE		
MSM0949	Msp_1351	predicted transcriptional activator	6.7E-18	MTH_628	unknown	1.6E-19
MSM0950	NONE			MTH_1003	molybdenum cofactor biosynthesis protein MoeA	6.8E-101
MSM0951	Msp_1335	translation initiation factor 1A (aIF-1A) (eIF1A)	1.6E-41	NONE	translation initiation factor, eIF-1A	1.3E-44
MSM0952	Msp_1337	predicted serine/threonine protein kinase	5.1E-59	MTH_1005	conserved protein	1.1E-75
MSM0953	NONE			MTH_630	unknown	1.5E-04
MSM0954	Msp_1338	predicted RNA-binding protein	1.4E-56	MTH_1006	conserved protein	2.0E-60
MSM0955	Msp_1339	type II DNA topoisomerase VI, subunit B	2.4E-203	MTH_1007	conserved protein	1.5E-213
MSM0956	Msp_1340	type II DNA topoisomerase VI, subunit A	4.3E-149	MTH_1008	conserved protein	1.8E-155
MSM0957	Msp_0119	hypothetical membrane-spanning protein	6.8E-20	MTH_524	unknown	4.9E-35

MSM0958	Msp_1110	CobN	5.3E-11	MTH_515	unknown	1.1E-08
MSM0959	Msp_0994	conserved hypothetical protein	3.0E-31	NONE		
MSM0960	Msp_0678	predicted cation transport ATPase	4.8E-134	MTH_411	cadmium efflux ATPase	1.9E-80
MSM0961	Msp_0224	predicted cation transport ATPase	9.6E-07	MTH_1535	heavy-metal transporting CPx-type ATPase	1.4E-08
MSM0962	Msp_1346	glyceraldehyde 3-phosphate dehydrogenase	4.7E-127	MTH_1009	glyceraldehyde 3-phosphate dehydrogenase	5.9E-134
MSM0963	Msp_0992	putative endonuclease IV	9.5E-06	MTH_1010	endonuclease IV	6.6E-71
MSM0964	Msp_1349	predicted phosphohydrolase	8.0E-19	MTH_1179	conserved protein	1.1E-38
MSM0965	Msp_0718	preducted 3-hydroxyacyl-CoA dehydrogenase	2.6E-126	NONE		
MSM0966	Msp_1415	putative 26S protease, regulatory subunit	6.5E-107	MTH_1011	ATP-dependent 26S protease regulatory subunit 8	7.4E-111
MSM0967	Msp_1408	HemA	4.6E-90	MTH_1012	glutamyl-tRNA reductase	3.2E-94
MSM0968	Msp_1407	predicted siroheme synthase	2.4E-45	MTH_1013	conserved protein	1.9E-41
MSM0969	Msp_1406	predicted metal-binding transcription factor	4.9E-54	MTH_1014	conserved protein	5.6E-58
MSM0970	Msp_0784	hypothetical protein	1.3E-21	NONE		
MSM0971	Msp_0393	methyl-coenzyme M reductase, component A2	7.6E-191	NONE	methyl coenzyme M reductase system, component A2	4.3E-209
MSM0972	Msp_1405	conserved hypothetical protein	1.3E-46	MTH_1016	conserved protein	5.5E-51
MSM0973	Msp_1404	putative GTP cyclohydrolase III	9.2E-76	MTH_1017	conserved protein	1.3E-88
MSM0974	Msp_1403	CofD	3.6E-90	MTH_1018	conserved protein	8.0E-98
MSM0975	Msp_1402	CofE	3.8E-63	MTH_1019	conserved protein	1.6E-76
MSM0976	Msp_1398	PurO	2.8E-51	MTH_1020	conserved protein	1.0E-51
MSM0977	Msp_1397	conserved hypothetical membrane-spanning protein	3.7E-24	MTH_1021	unknown	3.2E-30
MSM0978	Msp_1396	predicted biopolymer transport protein	1.5E-77	MTH_1022	biopolymer transport protein	4.1E-94
MSM0979	Msp_1395	RnhB	1.6E-48	MTH_1023	ribonuclease HII	9.8E-61
MSM0980	Msp_1517	DnaK	5.3E-16	MTH_1024	rod shape-determining protein	7.3E-136
MSM0981	NONE			MTH_1025	unknown	2.6E-51
MSM0982	Msp_1394	partially conserved hypothetical membrane-spanning protein	2.4E-38	MTH_1027	CDP-diacylglycerol-serine O-phosphatidyltransferase	8.2E-41
MSM0983	Msp_1393	conserved hypothetical membrane-spanning protein	8.7E-48	MTH_1028	unknown	1.7E-70
MSM0984	NONE			MTH_1030	unknown	1.4E-45
MSM0985	Msp_1392	conserved hypothetical protein	1.1E-29	MTH_1031	conserved protein	6.3E-34
MSM0986	Msp_0760	putative bile salt acid hydrolase	4.3E-110	NONE		
MSM0987	Msp_0329	MfnA	3.9E-100	MTH_1116	glutamate decarboxylase	6.1E-123
MSM0988	Msp_0328	PpsA	1.7E-273	MTH_1118	phosphoenolpyruvate synthase	2.0E-250
MSM0989	Msp_0327	50S ribosomal protein L10e	2.8E-58	MTH_1119	ribosomal protein L10	2.1E-65
MSM0990	Msp_1000	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, ATB-binding protein	4.7E-40	MTH_920	anion permease	4.2E-37
MSM0991	Msp_1001	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, permease protein	2.4E-11	MTH_478	sulfate transport system permease protein	4.1E-09
MSM0992	Msp_0326	hypothetical protein	1.0E-12	MTH_1121	unknown	8.9E-12
MSM0993	Msp_0601	partially conserved hypothetical protein, predicted GTPase	3.9E-04	MTH_1123	unknown	1.9E-15
MSM0994	Msp_0324	predicted nucleotidyltransferase	3.4E-101	MTH_1126	conserved protein	2.7E-90
MSM0995	Msp_0590	member of asn/thr-rich large protein family	8.7E-33	MTH_716	cell surface glycoprotein (s-layer protein)	1.3E-09
MSM0996	Msp_0983	member of asn/thr-rich large protein family	2.6E-26	MTH_716	cell surface glycoprotein (s-layer protein)	1.1E-09
MSM0997	Msp_0323	PyrC	1.1E-97	MTH_1127	dihydroorotase	7.8E-100
MSM0998	Msp_1447	EhbK	1.0E-30	MTH_1133	polyferredoxin (MvhB)	4.4E-145
MSM0999	Msp_0316	MvhA	3.4E-181	NONE	methyl viologen-reducing hydrogenase, alpha subunit	2.1E-207
MSM1000	Msp_0315	MvhG	3.2E-128	NONE	methyl viologen-reducing hydrogenase, gamma subunit	5.5E-138
MSM1001	Msp_0314	MvhD1	3.9E-61	NONE	methyl viologen-reducing hydrogenase, delta subunit	1.6E-67
MSM1002	Msp_0312	conserved hypothetical protein	1.2E-130	MTH_1150	ABC transporter subunit Ycf24	3.5E-152
MSM1003	Msp_0313	predicted ABC-type transport system	3.2E-82	MTH_1149	ABC transporter subunit Ycf16	8.0E-98
MSM1004	Msp_0311	conserved hypothetical protein	1.2E-27	MTH_1151	unknown	9.3E-33

MSM1005	Msp_0310	predicted GTP:adenosylcobinamide-phosphate guanylyltransferase	4.0E-36	MTH_1152	conserved protein	7.0E-35
MSM1006	Msp_0308	conserved hypothetical protein	2.2E-90	MTH_1153	conserved protein	5.2E-165
MSM1007	Msp_0307	MtrH	2.1E-108	MTH_1156	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H	2.9E-125
MSM1008	Msp_0306	MtrG	5.7E-12	MTH_1157	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit G	4.2E-21
MSM1009	Msp_0305	MtrF	5.5E-07	MTH_1158	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit F	9.3E-17
MSM1010	Msp_0304	MtrA	9.0E-62	MTH_1159	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit A	9.8E-93
MSM1011	Msp_0303	MtrB	1.0E-12	MTH_1160	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit B	1.7E-31
MSM1012	Msp_0302	MtrC	7.6E-49	MTH_1161	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit C	7.2E-81
MSM1013	Msp_0301	MtrD	2.0E-57	MTH_1162	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit D	1.0E-81
MSM1014	Msp_0300	MtrE	9.5E-74	MTH_1163	N5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit E	1.5E-121
MSM1015	Msp_0321	MrtA	7.6E-207	NONE	methyl coenzyme M reductase I, alpha subunit	1.7E-253
MSM1016	Msp_0320	MrtG	6.2E-86	NONE	methyl coenzyme M reductase I, gamma subunit	2.9E-109
MSM1017	Msp_0299	McrC	2.8E-67	NONE	methyl coenzyme M reductase I, C protein	2.6E-83
MSM1018	Msp_0319	MrtD	7.4E-19	NONE	methyl coenzyme M reductase I, D protein	1.1E-34
MSM1019	Msp_0318	MrtB	1.6E-133	NONE	methyl coenzyme M reductase I, beta subunit	3.4E-177
MSM1020	Msp_0298	predicted Fe-S oxidoreductase	2.0E-119	MTH_1170	conserved protein	1.7E-136
MSM1021	Msp_0284	conserved hypothetical protein	1.7E-99	MTH_1180	conserved protein	6.7E-117
MSM1022	Msp_0285	conserved hypothetical protein	8.5E-34	MTH_1181	unknown	2.0E-23
MSM1023	Msp_0973	ComB2	1.3E-44	MTH_1182	conserved protein	2.7E-42
MSM1024	Msp_0287	conserved hypothetical membrane-spanning protein	1.9E-98	MTH_1183	pheromone shutdown protein TraB	4.4E-58
MSM1025	Msp_0288	hypothetical protein	1.5E-20	MTH_1184	unknown	3.0E-20
MSM1026	NONE			MTH_1224	inosine-5'-monophosphate dehydrogenase related protein III	5.6E-04
MSM1027	NONE			MTH_1155	Na+/Ca+ exchanging protein related	2.1E-42
MSM1028	Msp_0289	predicted ATPase	9.5E-74	MTH_1186	conserved protein	2.0E-85
MSM1029	Msp_0693	conserved hypothetical protein	1.3E-39	MTH_1187	conserved protein	3.2E-23
MSM1030	Msp_0290	predicted pyridoxal phosphate-dependent enzyme	1.3E-124	MTH_1188	pleiotropic regulatory protein DegT	6.1E-123
MSM1031	Msp_0291	N2,N2-dimethylguanosine tRNA methyltransferase	1.1E-109	NONE	N2,N2-dimethylguanosine tRNA methyltransferase	4.1E-110
MSM1032	Msp_0293	predicted transcriptional regulator	9.3E-44	MTH_1193	transcriptional regulator	2.9E-52
MSM1033	Msp_0294	conserved hypothetical protein	1.8E-109	MTH_1196	conserved protein	7.7E-116
MSM1034	Msp_0295	conserved hypothetical protein	6.0E-17	MTH_1197	conserved protein	1.1E-22
MSM1035	Msp_0296	CofG	4.2E-96	MTH_1198	biotin synthetase related protein	6.4E-105
MSM1036	Msp_0297	predicted methyltransferase	2.3E-70	MTH_1200	met-10+ related protein	5.7E-72
MSM1037	Msp_0282	PsmB	7.5E-58	NONE	proteasome, beta subunit	7.8E-68
MSM1038	Msp_0281	predicted exonuclease	5.4E-245	MTH_1203	cleavage and polyadenylation specificity factor	3.5E-278
MSM1039	Msp_0280	PurM	1.6E-103	MTH_1204	phosphoribosylformylglycinamidine cyclo-ligase	4.0E-112
MSM1040	Msp_0279	ComC	7.6E-104	MTH_1205	malate dehydrogenase	5.7E-104
MSM1041	Msp_1507	putative DNA polymerase	6.8E-167	MTH_1208	DNA-dependent DNA polymerase family B (PolB1)	5.1E-183
MSM1042	NONE			MTH_1211	conserved protein	4.0E-71
MSM1043	Msp_1420	PyrK	4.4E-69	NONE	cytochrome-c3 hydrogenase, gamma subunit	1.6E-74
MSM1044	Msp_1421	PyrD	7.4E-90	MTH_1213	dihydroorotate oxidase	1.3E-106
MSM1045	Msp_0220	predicted glycosyltransferase	1.9E-12	MTH_1626	phosphoserine phosphatase	2.4E-05

MSM1046	Msp_1422	predicted ribosomal biogenesis protein	1.2E-89	MTH_1214	pre-mRNA splicing protein PRP31	1.4E-88
MSM1047	Msp_1423	FlpA	5.3E-64	MTH_1215	fibrillarin-like pre-rRNA processing protein	2.5E-62
MSM1048	Msp_1424	predicted phosphopantothenoylcysteine synthetase/decarboxylase	1.9E-43	MTH_1216	pantothenate metabolism flavoprotein	2.3E-52
MSM1049	Msp_1424	predicted phosphopantothenoylcysteine synthetase/decarboxylase	2.0E-55	MTH_1216	pantothenate metabolism flavoprotein	2.2E-54
MSM1050	Msp_1425	conserved hypothetical membrane-spanning protein	4.7E-11	MTH_1218	unknown	3.3E-21
MSM1051	Msp_1426	hypothetical membrane-spanning protein	3.5E-05	MTH_1219	unknown	9.0E-19
MSM1052	Msp_1427	PheA	2.5E-59	MTH_1220	chorismate mutase	1.1E-70
MSM1053	Msp_1428	conserved hypothetical protein	4.4E-60	MTH_1222	inosine-5'-monophosphate dehydrogenase related protein I	4.5E-72
MSM1054	Msp_1429	conserved hypothetical protein	2.2E-74	MTH_1224	inosine-5'-monophosphate dehydrogenase related protein III	1.3E-83
MSM1055	Msp_1431	partially conserved hypothetical protein	1.9E-36	MTH_1227	coenzyme PQQ synthesis protein III	1.9E-57
MSM1056	Msp_1432	putative 6-pyruvoyl tetrahydrobiopterin synthase	1.4E-38	MTH_1228	conserved protein	4.6E-47
MSM1057	Msp_1433	conserved hypothetical protein	2.1E-53	MTH_1229	conserved protein	2.1E-49
MSM1058	Msp_1434	conserved hypothetical protein	5.6E-85	MTH_1231	conserved protein	1.1E-95
MSM1059	Msp_0945	predicted RecB family exonuclease	1.2E-06	MTH_1233	unknown	1.4E-36
MSM1060	Msp_1436	EhbQ	4.9E-61	MTH_1235	conserved protein	1.2E-69
MSM1061	Msp_1442	EhbP	6.3E-22	MTH_1236	conserved protein	1.6E-28
MSM1062	Msp_1443	EhbO	6.1E-79	NONE	NADH dehydrogenase (ubiquinone), subunit 1 related protein	5.8E-111
MSM1063	Msp_1444	EhbN	8.0E-141	NONE	formate hydrogenlyase, subunit 5	2.8E-143
MSM1064	Msp_1445	EhbM	1.0E-62	NONE	formate hydrogenlyase, subunit 7	1.6E-67
MSM1065	Msp_1446	EhbL	8.6E-41	MTH_1240	ferredoxin-like protein	3.4E-51
MSM1066	Msp_1447	EhbK	7.7E-72	MTH_1241	polyferredoxin	1.7E-97
MSM1067	Msp_1448	EhbJ	4.5E-12	MTH_1242	unknown	5.5E-19
MSM1068	Msp_1449	EhbI	4.2E-48	MTH_1243	conserved protein	1.0E-49
MSM1069	Msp_1450	EhbH	3.5E-21	MTH_1244	conserved protein	5.0E-25
MSM1070	Msp_1451	EhbG	4.8E-15	MTH_1245	unknown	6.6E-16
MSM1071	Msp_1452	EhbF	1.1E-134	NONE	NADH dehydrogenase I, subunit N	8.4E-142
MSM1072	Msp_1453	EhbE	2.0E-32	MTH_1247	conserved protein	4.5E-40
MSM1073	Msp_1454	EhbD	4.1E-18	MTH_1248	conserved protein	9.4E-24
MSM1074	Msp_1455	EhbC	1.4E-10	MTH_1249	conserved protein	1.5E-18
MSM1075	Msp_1456	EhbB	2.2E-10	MTH_1250	unknown	1.1E-13
MSM1076	Msp_1457	EhbA	1.2E-27	MTH_1251	conserved protein	6.8E-37
MSM1077	Msp_1336	predicted permease	2.3E-05	NONE		
MSM1078	Msp_1336	predicted permease	9.6E-97	MTH_900	conserved protein	3.1E-32
MSM1079	Msp_1458	conserved hypothetical membrane-spanning protein	2.1E-28	MTH_1252	conserved protein	1.6E-35
MSM1080	NONE			MTH_1253	unknown	2.5E-48
MSM1081	Msp_0795	partially conserved hypothetical protein	1.4E-56	MTH_1634	transcriptional control factor (enhancer-binding protein)	5.0E-176
MSM1082	NONE			NONE		
MSM1083	Msp_0202	conserved hypothetical membrane-spanning protein	4.5E-35	MTH_230	unknown	1.0E-33
MSM1084	Msp_1459	ArgG	7.4E-138	MTH_1254	argininosuccinate synthase	2.1E-136
MSM1085	Msp_1240	AqpM2	1.8E-54	MTH_103	water channel protein	1.5E-71
MSM1086	NONE			MTH_101	unknown	3.8E-194
MSM1087	NONE			NONE		
MSM1088	NONE			NONE		
MSM1089	Msp_0506	hypothetical membrane-spanning protein	3.3E-04	NONE		
MSM1090	Msp_1057	SfsA	6.0E-33	MTH_1521	sugar fermentation stimulation protein	3.6E-31
MSM1091	Msp_1501	predicted sugar kinase	3.6E-97	MTH_1256	conserved protein	1.4E-114

MSM1092	Msp_1502	formylmethanofuran-tetrahydromethanopterin formyltransferase	1.2E-91	MTH_1259	formylmethanofuran:tetrahydromethanopterin formyltransferase	1.3E-127
MSM1093	Msp_0233	conserved hypothetical protein	2.3E-22	NONE		
MSM1094	Msp_1503	conserved hypothetical membrane-spanning protein	2.8E-81	MTH_1261	conserved protein	7.2E-97
MSM1095	Msp_0830	Trk-type potassium transport system, membrane protein	2.6E-62	MTH_1264	TRK system potassium uptake protein TrkH	2.1E-122
MSM1096	Msp_0250	TrkA1	3.1E-52	MTH_1265	TRK system potassium uptake protein TrkA	3.6E-79
MSM1097	Msp_1505	putative Zn-dependent hydrolase	2.3E-40	MTH_1267	conserved protein	1.2E-53
MSM1098	Msp_1418	putative archaeal holliday junction resolvase	1.4E-38	MTH_1270	conserved protein	1.4E-43
MSM1099	Msp_0270	predicted biotin synthase related protein	7.4E-106	MTH_1279	conserved protein	2.3E-75
MSM1100	NONE			MTH_627	unknown	7.2E-10
MSM1101	Msp_0269	GatB	1.4E-175	MTH_1280	PET112-like protein	3.6E-182
MSM1102	Msp_0268	conserved hypothetical protein	3.4E-78	MTH_1282	inosine-5'-monophosphate dehydrogenase related protein VI	2.3E-93
MSM1103	Msp_0267	HisE	4.8E-31	MTH_1283	phosphoribosyl-AMP cyclohydrolase homolog	3.0E-34
MSM1104	Msp_1506	predicted acetyltransferase	2.6E-11	MTH_1284	conserved protein	3.2E-16
MSM1105	Msp_1492	conserved hypothetical protein	7.0E-62	MTH_1286	phosphoribosylaminoimidazole carboxylase related protein	1.7E-65
MSM1106	Msp_1497	HypF	8.5E-208	MTH_1287	transcriptional regulator HypF homolog	2.3E-219
MSM1107	Msp_1519	predicted transcriptional regulator	6.6E-34	MTH_1288	unknown	1.8E-52
MSM1108	Msp_1518	GrpE	2.1E-44	MTH_1289	heat shock protein GrpE	1.6E-44
MSM1109	Msp_1517	DnaK	8.6E-247	MTH_1290	DnaK protein (Hsp70)	7.7E-251
MSM1110	Msp_1516	DnaJ	3.0E-118	MTH_1291	DnaJ protein	1.0E-122
MSM1111	Msp_0145	member of asn/thr-rich large protein family	5.9E-49	MTH_716	cell surface glycoprotein (s-layer protein)	7.7E-12
MSM1112	Msp_0762	member of asn/thr-rich large protein family	1.6E-40	MTH_716	cell surface glycoprotein (s-layer protein)	3.3E-11
MSM1113	Msp_0762	member of asn/thr-rich large protein family	2.9E-70	MTH_716	cell surface glycoprotein (s-layer protein)	1.2E-05
MSM1114	Msp_0145	member of asn/thr-rich large protein family	1.3E-24	MTH_716	cell surface glycoprotein (s-layer protein)	3.3E-15
MSM1115	Msp_0017	conserved hypothetical protein	2.2E-21	NONE		
MSM1116	Msp_1108	member of asn/thr-rich large protein family	4.2E-137	MTH_911	probable surface protein	1.5E-12
MSM1117	Msp_1110	CobN	8.5E-304	MTH_514	cobalamin biosynthesis protein N	1.4E-239
MSM1118	Msp_1494	hypothetical membrane-spanning protein	1.5E-18	MTH_1294	unknown	2.5E-23
MSM1119	Msp_1495	hypothetical membrane-spanning protein	4.1E-25	MTH_1295	unknown	4.8E-36
MSM1120	Msp_1496	methionine aminopeptidase	3.4E-53	MTH_1296	methionine aminopeptidase	2.8E-86
MSM1121	Msp_1305	FrhB	3.9E-77	NONE	coenzyme F420-reducing hydrogenase, beta subunit	2.1E-97
MSM1122	Msp_1304	FrhG	4.6E-81	NONE	coenzyme F420-reducing hydrogenase, gamma subunit	2.2E-102
MSM1123	Msp_1514	putative coenzyme F420 hydrogenase, delta subunit-like protein	9.3E-44	NONE	coenzyme F420-reducing hydrogenase, delta subunit	4.7E-61
MSM1124	Msp_1302	FrhA	9.4E-138	NONE	coenzyme F420-reducing hydrogenase, alpha subunit	8.8E-163
MSM1125	Msp_1110	CobN	2.3E-10	MTH_1301	unknown	3.8E-11
MSM1126	Msp_0120	predicted transcriptional regulator	3.1E-20	MTH_1795	transcriptional regulator	1.1E-20
MSM1127	Msp_0121	predicted cation transport ATPase	1.2E-162	MTH_411	cadmium efflux ATPase	1.2E-119
MSM1128	NONE			NONE		
MSM1129	Msp_1523	conserved hypothetical protein	2.3E-118	MTH_1305	conserved protein	3.6E-134
MSM1130	Msp_1028	conserved hypothetical protein	4.5E-44	MTH_1868	conserved protein	1.4E-15
MSM1131	Msp_1524	conserved hypothetical protein	1.1E-56	MTH_1306	conserved protein	1.1E-59
MSM1132	Msp_1525	ribosome biogenesis protein Nop10	2.3E-15	MTH_1307	unknown	4.0E-16
MSM1133	Msp_1527	putative translation initiation factor 2, alpha subunit (aIF-2-alpha) (eIF2A)	3.4E-94	NONE	translation initiation factor eIF-2, alpha subunit	3.5E-104
MSM1134	Msp_1528	30S ribosomal protein S27e	2.3E-17	MTH_1309	ribosomal protein S27	8.1E-18
MSM1135	Msp_1529	50S ribosomal protein L44e	1.6E-41	MTH_1310	ribosomal protein L36a	2.7E-42
MSM1136	Msp_1530	partially conserved hypothetical protein	1.6E-30	MTH_1311	unknown	2.1E-49
MSM1137	Msp_1531	DNA polymerase sliding clamp (PCNA)	1.5E-73	MTH_1312	proliferating-cell nuclear antigen	6.0E-93
MSM1138	Msp_0580	predicted glutamine amidotransferase	5.2E-73	MTH_787	cobyric acid synthase	9.2E-10

MSM1139	Msp_0581	predicted UDP-N-acetylmuramyl tripeptide synthase	3.6E-90	MTH_530	UDP-N-acetylmuramyl tripeptide synthetase related protein	6.8E-16
MSM1140	Msp_0417	hypothetical membrane-spanning protein	2.7E-04	NONE		
MSM1141	Msp_1075	TrpA	7.3E-44	NONE	tryptophan synthase, subunit alpha	6.5E-48
MSM1142	Msp_1074	TrpB	6.4E-123	NONE	tryptophan synthase, beta subunit	1.3E-120
MSM1143	Msp_1072	TrpC	1.7E-42	MTH_1657	indole-3-glycerol phosphate synthase	1.4E-38
MSM1144	Msp_1076	TrpD	2.0E-71	MTH_1661	anthranilate phosphoribosyltransferase	2.3E-68
MSM1145	Msp_1071	TrpG	7.4E-51	MTH_1656	anthranilate synthase component II	1.1E-43
MSM1146	Msp_1070	TrpE	6.5E-78	MTH_1655	anthranilate synthase component I	9.9E-84
MSM1147	NONE			NONE		
MSM1148	NONE			MTH_1189	conserved protein	8.2E-08
MSM1149	Msp_0607	hypothetical membrane-spanning protein	6.0E-33	MTH_1192	conserved protein	2.8E-31
MSM1150	Msp_0608	predicted transcriptional regulator	9.4E-19	MTH_1328	conserved protein	1.3E-17
MSM1151	Msp_1247	PurB	6.0E-159	MTH_1537	adenylosuccinate lyase	8.4E-174
MSM1152	Msp_0879	hypothetical membrane-spanning protein	2.8E-04	MTH_1538	unknown	6.4E-25
MSM1153	Msp_0224	predicted cation transport ATPase	1.1E-205	MTH_1535	heavy-metal transporting CPx-type ATPase	5.1E-199
MSM1154	Msp_0200	predicted metal-dependent hydrolase	1.2E-07	MTH_1534	aryldialkylphosphatase related protein	5.0E-89
MSM1155	Msp_0225	conserved hypothetical protein	1.4E-40	MTH_1530	conserved protein	1.7E-42
MSM1156	Msp_0221	TruD	6.2E-125	MTH_1529	conserved protein	4.6E-134
MSM1157	Msp_1512	hypothetical membrane-spanning protein	3.5E-05	MTH_1526	conserved protein	8.9E-04
MSM1158	Msp_1511	HypE2	8.9E-126	MTH_1525	hydrogenase expression/formation protein HypE related protein	4.2E-156
MSM1159	Msp_1510	HisH	3.0E-38	MTH_1524	imidazoleglycerol-phosphate synthase	9.1E-58
MSM1160	Msp_1461	predicted nitrogenase molybdenum-iron protein	3.8E-118	MTH_1522	nitrogenase alpha chain (NifD) related protein	8.9E-131
MSM1161	Msp_0719	partially conserved hypothetical membrane-spanning protein	2.8E-05	NONE		
MSM1162	NONE			NONE		
MSM1163	NONE			NONE		
MSM1164	Msp_1463	predicted GTPase	1.4E-143	MTH_1515	GTP-binding protein	2.4E-153
MSM1165	Msp_1472	predicted phosphohydrolase	2.2E-67	MTH_1179	conserved protein	9.0E-10
MSM1166	Msp_1474	conserved hypothetical membrane-spanning protein	1.5E-146	NONE		
MSM1167	Msp_1464	CbiE	6.8E-48	MTH_1514	precorrin-6Y methylase	3.9E-50
MSM1168	Msp_0590	member of asn/thr-rich large protein family	1.7E-16	MTH_75	surface protease related protein	2.1E-11
MSM1169	NONE			NONE		
MSM1170	Msp_0169	putative arsenical prump-driving ATPase	5.3E-96	MTH_1511	arsenical pump-driving ATPase	6.9E-108
MSM1171	Msp_0170	NadE	1.1E-63	MTH_1510	NH(3)-dependent NAD+ synthetase	1.3E-60
MSM1172	Msp_0171	LeuS	0.0E+00	MTH_1508	leucyl-tRNA synthetase	0.0E+00
MSM1173	Msp_0004	predicted tRNA(1-methyladenosine) methyltransferase	1.0E-62	MTH_1414	protein-L-isoaspartate methyltransferase homolog	1.4E-77
MSM1174	Msp_0309	HtpX	1.8E-38	MTH_569	heat shock protein X	2.1E-67
MSM1175	Msp_0548	hypothetical membrane-spanning protein	6.6E-11	NONE		
MSM1176	Msp_0413	RfcS	2.2E-115	NONE	replication factor C, small subunit	3.7E-125
MSM1177	Msp_0414	RfcL	1.1E-113	NONE	replication factor C, large subunit	3.8E-123
MSM1178	Msp_0578	conserved hypothetical protein	4.1E-34	MTH_239	unknown	9.7E-38
MSM1179	Msp_0647	AroE	1.8E-72	MTH_242	shikimate 5-dehydrogenase	1.2E-71
MSM1180	NONE			MTH_1189	conserved protein	1.6E-08
MSM1181	Msp_0648	HisS	5.1E-114	MTH_244	histidyl-tRNA synthetase	3.8E-130
MSM1182	Msp_0649	HisI	1.6E-39	MTH_245	phosphoribosyl-AMP cyclohydrolase	1.0E-40
MSM1183	Msp_0650	predicted ATPase	1.5E-155	MTH_246	twitching mobility (PilT) related protein	8.0E-185
MSM1184	Msp_0651	predicted sugar phosphate isomerase/epimerase or endonuclease	8.7E-48	MTH_247	conserved protein	4.5E-49
MSM1185	Msp_1499	putative methylated-DNA--protein-cysteine methyltransferase	1.3E-12	MTH_618	O6-methylguanidine-DNA methyltransferase	2.8E-15

MSM1186	Msp_1489	predicted potassium transport system, membrane component	9.9E-111	NONE		
MSM1187	Msp_0007	predicted ERCC4-like helicase	5.4E-213	NONE	ATP-dependent RNA helicase, eIF-4A family	3.5E-241
MSM1188	Msp_0590	member of asn/thr-rich large protein family	1.4E-49	MTH_716	cell surface glycoprotein (s-layer protein)	6.9E-13
MSM1189	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM1190	Msp_1211	partially conserved hypothetical membrane-spanning protein	6.7E-128	MTH_530	UDP-N-acetylmuramyl tripeptide synthetase related protein	3.1E-57
MSM1191	Msp_1212	predicted UDP-N-acetylmuramoylalanine--D-glutamate ligase	7.9E-102	MTH_531	UDP-N-acetylmuramyl tripeptide synthetase related protein	1.3E-40
MSM1192	Msp_0008	conserved hypothetical protein	9.1E-124	MTH_1421	conserved protein	5.0E-137
MSM1193	Msp_0009	putative single-stranded-DNA-specific exonuclease	9.9E-111	MTH_1422	conserved protein	9.3E-136
MSM1194	Msp_0010	30S ribosomal protein S15P	5.3E-48	MTH_1423	ribosomal protein S13 (E.coli)	2.1E-49
MSM1195	Msp_0011	putative xanthosine triphosphate pyrophosphatase	1.9E-61	MTH_1424	conserved protein	1.2E-62
MSM1196	Msp_0635	cell division control protein 6-like 2	9.7E-06	NONE		
MSM1197	NONE			NONE		
MSM1198	Msp_0013	putative O-sialoglycoprotein endopeptidase	7.7E-159	MTH_1425	O-sialoglycoprotein endopeptidase	1.9E-174
MSM1199	Msp_0999	hypothetical protein	7.0E-06	NONE		
MSM1200	Msp_0012	predicted phosphoribosyltransferase	1.4E-88	MTH_1426	conserved protein	3.4E-99
MSM1201	Msp_0014	UppP	6.0E-72	MTH_1428	bacitracin resistance protein	1.1E-43
MSM1202	Msp_0015	IlvE	4.0E-114	MTH_1430	branched-chain amino-acid aminotransferase	5.2E-110
MSM1203	Msp_0724	hypothetical membrane-spanning protein	6.7E-09	MTH_470	conserved protein	7.9E-05
MSM1204	Msp_0163	F420-dependent methylenetetrahydromethanopterin dehydrogenase	4.0E-82	NONE	coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin dehydrogenase	2.2E-102
MSM1205	Msp_0417	hypothetical membrane-spanning protein	5.3E-04	MTH_1490	unknown	3.5E-17
MSM1206	Msp_0164	HisB	2.5E-57	MTH_1467	imidazoleglycerol-phosphate dehydratase	9.7E-54
MSM1207	NONE			MTH_1470	molybdenum transport protein ModA related protein	2.2E-17
MSM1208	Msp_0165	predicted polysaccharide biosynthesis protein	5.0E-116	MTH_1471	O-antigen transporter homolog	3.2E-87
MSM1209	Msp_0540	predicted multimeric flavodoxin	6.7E-25	MTH_1473	conserved protein	4.7E-54
MSM1210	Msp_0925	predicted arabinose efflux permease	7.5E-22	MTH_195	efflux pump antibiotic resistance protein	2.5E-24
MSM1211	Msp_0260	hypothetical protein	4.6E-16	MTH_1626	phosphoserine phosphatase	4.3E-06
MSM1212	NONE			NONE		
MSM1213	Msp_1498	formaldehyde activating enzyme fused to 3-hexulose-6-phosphate synthase	8.3E-162	MTH_1474	D-arabino 3-hexulose 6-phosphate formaldehyde lyase related protein	6.3E-169
MSM1214	Msp_1573	ThrS	7.3E-202	MTH_1455	threonyl-tRNA synthetase	1.3E-225
MSM1215	Msp_0162	CbiA	1.7E-147	NONE	cobyrinic acid a,c-diamide synthase	9.4E-143
MSM1216	Msp_0166	conserved hypothetical membrane-spanning protein	1.3E-74	MTH_1461	conserved protein	2.1E-67
MSM1217	Msp_0019	partially conserved hypothetical protein	5.0E-45	MTH_1434	unknown	1.3E-55
MSM1218	Msp_0020	SurE	1.2E-68	MTH_1435	survival protein SurE	1.5E-73
MSM1219	NONE			NONE		
MSM1220	NONE			MTH_1440	unknown	8.6E-14
MSM1221	Msp_0021	conserved hypothetical protein	5.2E-89	MTH_1441	conserved protein	3.4E-106
MSM1222	Msp_0022	IlvC	6.9E-126	MTH_1442	ketol-acid reductoisomerase	2.7E-122
MSM1223	Msp_0591	predicted carbonic anhydrase	8.1E-13	MTH_1582	carbonic anhydrase	3.7E-38
MSM1224	Msp_0025	IlvH1	1.1E-45	NONE	acetolactate synthase, small subunit	4.1E-55
MSM1225	Msp_0026	IlvB1	6.3E-180	NONE	acetolactate synthase, large subunit	3.5E-207
MSM1226	Msp_0031	ArgF	2.3E-102	MTH_1446	ornithine carbamoyltransferase	4.6E-102
MSM1227	Msp_0030	PurD	1.1E-150	MTH_1445	glycinamide ribonucleotide synthetase	4.2E-147
MSM1228	Msp_0513	predicted Na+-driven multidrug efflux pump	5.6E-108	MTH_314	conserved protein	2.8E-95
MSM1229	Msp_0513	predicted Na+-driven multidrug efflux pump	1.1E-125	MTH_314	conserved protein	3.1E-105
MSM1230	Msp_0512	predicted transcriptional regulator	5.3E-25	MTH_313	transcriptional regulator	2.2E-17
MSM1231	Msp_1574	ArgS	1.4E-157	MTH_1447	arginyl-tRNA synthetase	9.3E-175

MSM1232	Msp_1575	putative signal peptidase	3.6E-42	MTH_1448	signal peptidase	2.7E-42
MSM1233	Msp_1180	HemL	5.8E-138	MTH_228	glutamate-1-semialdehyde aminotransferase	2.1E-136
MSM1234	Msp_1179	CbiC	8.2E-68	MTH_227	precorrin isomerase	7.1E-58
MSM1235	Msp_0093	predicted flavoprotein	2.5E-59	NONE		
MSM1236	Msp_0135	AspS	1.9E-164	MTH_226	aspartyl-tRNA synthetase	1.2E-165
MSM1237	Msp_1576	IlvD	7.2E-195	MTH_1449	dihydroxy-acid dehydratase	3.4E-177
MSM1238	Msp_0134	HisD	2.7E-131	MTH_225	histidinol dehydrogenase	2.7E-138
MSM1239	Msp_1569	predicted DNA-binding protein	2.7E-92	MTH_1458	unknown	5.1E-96
MSM1240	Msp_1570	conserved hypothetical protein	8.9E-23	MTH_1457	unknown	3.0E-24
MSM1241	Msp_1571	predicted ATPase	5.2E-82	MTH_1456	chromosome partitioning protein Soj	1.9E-73
MSM1242	Msp_1074	TrpB	7.2E-37	NONE	tryptophan synthase, beta subunit homolog	1.0E-168
MSM1243	NONE			MTH_1477	unknown	3.1E-73
MSM1244	Msp_1491	predicted metal-dependent phosphoesterase	1.9E-45	MTH_1478	conserved protein	8.9E-28
MSM1245	Msp_0198	AlbA	2.2E-26	MTH_1483	conserved protein	3.8E-27
MSM1246	Msp_0199	LeuA1	8.3E-162	MTH_1481	isopropylmalate synthase	2.8E-175
MSM1247	Msp_0197	conserved hypothetical membrane-spanning protein	2.6E-78	MTH_1485	serine/threonine protein kinase related protein	1.2E-92
MSM1248	Msp_0196	ABC-type multidrug transport system, permease protein	4.6E-74	MTH_1486	conserved protein	1.5E-82
MSM1249	Msp_0195	ABC-type multidrug transport system, ATP-binding protein	1.6E-94	MTH_1487	ABC transporter (ATP-binding	5.1E-103
MSM1250	Msp_0194	predicted transcriptional regulator	3.6E-19	MTH_1488	unknown	1.6E-19
MSM1251	Msp_0651	predicted sugar phosphate isomerase/epimerase or endonuclease	7.5E-26	MTH_1489	conserved protein	8.8E-60
MSM1252	Msp_0191	MapB	8.0E-38	MTH_1493	cation transporting P-type ATPase related protein	1.8E-54
MSM1253	Msp_0181	GatA	2.1E-165	MTH_1496	amidase	1.1E-164
MSM1254	Msp_0174	predicted cobyric acid synthase	7.3E-115	NONE	cobyrinic acid a,c-diamide synthase related protein	8.9E-115
MSM1255	NONE			NONE		
MSM1256	Msp_0175	RibB	2.5E-59	MTH_1499	GTP cyclohydrolase II	2.8E-63
MSM1257	Msp_0177	predicted transcriptional regulator	1.7E-19	MTH_1500	conserved protein	9.4E-24
MSM1258	Msp_0180	TfrA	2.0E-174	NONE	succinate dehydrogenase, flavoprotein subunit	3.9E-185
MSM1259	Msp_0200	predicted metal-dependent hydrolase	1.0E-115	MTH_1505	N-ethylammeline chlorohydrolase homolog	9.3E-120
MSM1260	Msp_0383	archaeal histone	8.8E-16	MTH_1696	histone HMtA2	8.4E-16
MSM1261	Msp_0178	HisG	1.4E-88	MTH_1506	ATP phosphoribosyltransferase	1.3E-90
MSM1262	NONE			NONE		
MSM1263	Msp_0003	PyrB	8.4E-98	MTH_1413	aspartate carbamoyltransferase	5.1E-96
MSM1264	Msp_0001	cell division control protein 6-like 1	4.9E-141	MTH_1412	Cdc6 related protein	8.2E-160
MSM1265	NONE			MTH_1410	unknown	1.4E-31
MSM1266	Msp_1588	CobD	4.4E-76	MTH_1409	cobalamin biosynthesis protein B	7.6E-54
MSM1267	Msp_1587	CbiG	2.3E-70	MTH_1408	cobalamin biosynthesis protein G	3.0E-50
MSM1268	Msp_1586	conserved hypothetical protein	2.7E-21	MTH_1407	conserved protein	2.6E-28
MSM1269	NONE			NONE		
MSM1270	Msp_1585	predicted class II aldolase	4.7E-40	MTH_1406	fuculose-1-phosphate aldolase	4.9E-43
MSM1271	Msp_1584	PolB	4.5E-131	MTH_1405	DNA polymerase delta small subunit	3.6E-156
MSM1272	Msp_1583	hypothetical membrane-spanning protein	5.8E-19	MTH_1404	unknown	4.3E-28
MSM1273	Msp_1582	CbiH	2.5E-98	MTH_1403	precorrin-3 methylase	1.2E-101
MSM1274	NONE			MTH_1402	conserved protein	6.4E-73
MSM1275	Msp_0962	hypothetical membrane-spanning protein	2.4E-04	MTH_1401	unknown	5.4E-108
MSM1276	Msp_1558	hypothetical protein	1.7E-10	MTH_1400	unknown	1.3E-16
MSM1277	Msp_1559	conserved hypothetical membrane-spanning protein	8.0E-38	MTH_1399	unknown	2.0E-46
MSM1278	Msp_0757	predicted ATPase	4.3E-101	NONE		
MSM1279	Msp_1562	conserved hypothetical protein	1.5E-50	MTH_1398	conserved protein	2.3E-52
MSM1280	Msp_1561	conserved hypothetical protein	5.0E-52	MTH_1397	conserved protein	1.2E-25
MSM1281	Msp_1563	CbiX	7.5E-42	MTH_1397	conserved protein	8.6E-30

MSM1282	Msp_0590	member of asn/thr-rich large protein family	3.1E-13	MTH_716	cell surface glycoprotein (s-layer protein)	2.7E-05
MSM1283	Msp_1564	ThiL	6.8E-48	MTH_1396	thiamine monphosphate kinase	3.1E-57
MSM1284	Msp_1565	predicted pyruvate-formate lyase-activating enzyme	1.5E-66	MTH_1395	pyruvate formate-lyase activating enzyme related protein	3.5E-81
MSM1285	Msp_0615	partially conserved hypothetical membrane-spanning protein	6.8E-05	NONE		
MSM1286	Msp_1479	predicted 3-octaprenyl-4-hydroxybenzoate carboxy-lyase	5.7E-147	MTH_1394	conserved protein	3.5E-152
MSM1287	Msp_1480	PurE	6.4E-68	MTH_1393	phosphoribosylaminoimidazole carboxylase	1.9E-80
MSM1288	NONE			NONE		
MSM1289	Msp_1168	CobS	6.5E-04	NONE		
MSM1290	Msp_0054	predicted glycosyltransferase	1.4E-33	MTH_374	dolichyl-phosphate mannose synthase related protein	7.5E-31
MSM1291	NONE			NONE		
MSM1292	Msp_0920	predicted transcriptional accessory protein	9.5E-232	NONE	translation initiation factor eIF-2, alpha subunit	2.1E-04
MSM1293	Msp_0965	predicted nitroreductase	3.3E-16	MTH_120	NADPH-oxidoreductase	2.1E-33
MSM1294	Msp_1481	conserved hypothetical membrane-spanning protein	3.4E-124	MTH_1392	dolichyl-phosphate mannoosyltransferase related protein	5.8E-150
MSM1295	Msp_1482	conserved hypothetical membrane-spanning protein	7.0E-94	MTH_1391	conserved protein	3.8E-114
MSM1296	Msp_1483	RibH	2.0E-50	MTH_1390	riboflavin synthase beta subunit	1.4E-54
MSM1297	Msp_0219	conserved hypothetical protein	3.0E-70	NONE		
MSM1298	Msp_1484	LeuB	3.8E-109	MTH_1388	3-isopropylmalate dehydrogenase	3.2E-103
MSM1299	Msp_1485	LeuD1	3.1E-43	NONE	3-isopropylmalate dehydratase, LeuC subunit	3.3E-60
MSM1300	Msp_1486	LeuC1	1.3E-165	NONE	3-isopropylmalate dehydratase, LeuD subunit	1.7E-175
MSM1301	NONE			NONE		
MSM1302	NONE			NONE		
MSM1303	Msp_0214	predicted UDP-N-acetyl-D-mannosaminuronate dehydrogenase	2.3E-143	MTH_836	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	2.8E-79
MSM1304	Msp_1116	predicted dTDP-4-dehydrorhamnose reductase	9.6E-42	MTH_1792	dTDP-4-dehydrorhamnose reductase	1.9E-73
MSM1305	Msp_0762	member of asn/thr-rich large protein family	5.3E-36	MTH_716	cell surface glycoprotein (s-layer protein)	2.2E-12
MSM1306	Msp_0590	member of asn/thr-rich large protein family	3.5E-45	MTH_716	cell surface glycoprotein (s-layer protein)	1.8E-07
MSM1307	Msp_1102	predicted dTDP-glucose pyrophosphorylase	4.1E-41	MTH_1791	glucose-1-phosphate thymidylyltransferase	1.4E-123
MSM1308	Msp_0539	predicted dTDP-4-dehydrorhamnose 3,5-epimerase	1.9E-68	NONE	dTDP-4-dehydrorhamnose 3,5-epimerase	5.4E-60
MSM1309	Msp_1114	predicted dTDP-D-glucose 4,6-dehydratase	4.5E-106	NONE	dTDP-glucose 4,6-dehydratase	3.0E-137
MSM1310	Msp_0212	predicted glycosyltransferase	1.8E-54	MTH_884	teichoic acid biosynthesis related protein	7.1E-10
MSM1311	Msp_0496	predicted glycosyltransferase	2.8E-34	MTH_136	dolichyl-phosphate mannose synthase	2.2E-05
MSM1312	Msp_0500	predicted glycosyltransferase	4.8E-79	MTH_172	conserved protein	6.5E-19
MSM1313	Msp_0492	predicted glycosyltransferase	6.1E-57	MTH_338	LPS biosynthesis RfbU related protein	2.9E-07
MSM1314	NONE			NONE		
MSM1315	NONE			NONE		
MSM1316	Msp_0495	predicted glycosyltransferase	2.3E-33	MTH_884	teichoic acid biosynthesis related protein	8.9E-09
MSM1317	Msp_0500	predicted glycosyltransferase	2.9E-07	NONE		
MSM1318	Msp_0927	hypothetical protein	2.1E-30	NONE		
MSM1319	Msp_0928	hypothetical protein	3.0E-31	NONE		
MSM1320	Msp_0492	predicted glycosyltransferase	4.1E-58	NONE		
MSM1321	Msp_0500	predicted glycosyltransferase	4.4E-76	MTH_172	conserved protein	9.5E-17
MSM1322	Msp_0492	predicted glycosyltransferase	6.5E-62	MTH_338	LPS biosynthesis RfbU related protein	9.6E-12
MSM1323	Msp_0495	predicted glycosyltransferase	5.3E-34	MTH_884	teichoic acid biosynthesis related protein	2.0E-08
MSM1324	Msp_0215	predicted glycosyltransferase	1.0E-32	MTH_884	teichoic acid biosynthesis related protein	1.5E-08
MSM1325	Msp_0204	predicted ABC-type polysaccharide/polyol phosphate export system, permease protein	1.2E-64	MTH_1092	putative membrane protein	6.6E-06
MSM1326	Msp_0205	predicted ABC-type polysaccharide/polyol phosphate export system, ATP-binding protein	3.7E-79	MTH_1370	ABC transporter (ATP-binding protein)	2.0E-16
MSM1327	NONE			MTH_361	teichoic acid biosynthesis protein RodC related protein	2.4E-17
MSM1328	Msp_0212	predicted glycosyltransferase	2.9E-26	MTH_884	teichoic acid biosynthesis related protein	2.0E-12

MSM1329	Msp_0206	predicted glycosyltransferase	5.2E-82	MTH_172	conserved protein	2.5E-46
MSM1330	Msp_0207	predicted glycosyltransferase	9.1E-69	MTH_172	conserved protein	1.1E-20
MSM1331	Msp_0208	predicted bacterial sugar transferase	9.0E-117	NONE		
MSM1332	Msp_1487	predicted ssDNA-binding protein	6.2E-157	MTH_1385	replication factor A related protein	7.8E-152
MSM1333	Msp_1488	RadA	6.9E-142	MTH_1383	DNA repair protein RadA	6.4E-144
MSM1334	Msp_1477	predicted permease	1.4E-56	MTH_1382	conserved protein	1.2E-57
MSM1335	NONE			NONE		
MSM1336	Msp_1476	HdrA1	6.9E-277	NONE	heterodisulfide reductase, subunit A	2.0E-298
MSM1337	Msp_1475	GlyA	5.9E-145	MTH_1380	serine hydroxymethyltransferase	6.5E-151
MSM1338	Msp_1473	predicted flavoprotein	3.4E-53	MTH_1379	conserved protein (contains ferredoxin domain)	5.0E-73
MSM1339	Msp_1471	conserved hypothetical protein	2.5E-11	MTH_1377	conserved protein	9.7E-22
MSM1340	Msp_1470	S-adenosylmethionine synthetase	2.2E-138	MTH_1376	conserved protein	3.7E-148
MSM1341	Msp_1468	IleS	0.0E+00	MTH_1375	isoleucyl-tRNA synthetase	0.0E+00
MSM1342	Msp_1467	PurL	5.9E-239	MTH_1374	phosphoribosylformylglycinamidine synthase II	4.4E-255
MSM1343	NONE			MTH_1369	molybdenum cofactor biosynthesis MoeA	2.5E-110
MSM1344	Msp_1466	predicted membrane-associated Zn-dependent protease	1.4E-81	MTH_1368	conserved protein	3.4E-99
MSM1345	NONE			NONE		
MSM1346	Msp_0822	hypothetical protein	1.6E-06	NONE		
MSM1347	NONE			NONE		
MSM1348	Msp_0789	rubrerythrin	2.7E-04	MTH_1351	conserved protein	4.2E-37
MSM1349	Msp_0787	FprA	2.9E-136	MTH_1350	flavoprotein AI	2.7E-152
MSM1350	Msp_0061	conserved hypothetical protein	5.4E-32	MTH_1349	conserved protein	3.1E-48
MSM1351	Msp_0038	CbiL	1.1E-58	MTH_1348	precorrin-2 methyltransferase	9.8E-61
MSM1352	Msp_0036	putative ATP-dependent helicase	1.1E-175	MTH_1347	probable ATP-dependent helicase	3.4E-212
MSM1353	Msp_1532	hypothetical membrane-spanning protein	1.6E-08	MTH_1313	unknown	9.0E-13
MSM1354	Msp_1533	RpoM1	4.7E-33	MTH_1314	transcription elongation factor TFIIS	4.8E-36
MSM1355	Msp_1534	putative ADP-ribose pyrophosphatase	4.9E-38	MTH_1315	mutator MutT protein	1.1E-34
MSM1356	Msp_1535	RpoL	2.1E-14	NONE	DNA-dependent RNA polymerase, subunit L	5.5E-19
MSM1357	Msp_1536	predicted RNA-binding protein	2.6E-32	MTH_1318	conserved protein	1.6E-46
MSM1358	Msp_1537	predicted diphthamide synthase, subunit DPH2	6.1E-95	MTH_1319	conserved protein	1.1E-109
MSM1359	Msp_1538	putative adenine phosphoribosyltransferase	5.0E-52	MTH_1320	adenine phosphoribosyltransferase	2.2E-54
MSM1360	Msp_1539	signal recognition particle, 54 kDa protein	2.0E-151	MTH_1321	signal recognition particle protein SRP54	5.8E-159
MSM1361	Msp_1541	predicted pseudouridylate synthase	4.0E-82	MTH_1322	conserved protein	1.0E-104
MSM1362	NONE			MTH_809	molybdenum cofactor biosynthesis protein MoaC	2.2E-47
MSM1363	Msp_0229	SecG	2.2E-12	NONE		
MSM1364	Msp_0032	HisF	1.6E-112	MTH_1343	imidazoleglycerol-phosphate synthase (cyclase)	3.7E-109
MSM1365	Msp_0034	putative 3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase	2.1E-37	MTH_1342	8-oxoguanine DNA glycosylase	1.1E-68
MSM1366	NONE			MTH_758	S-D-lactoylglutathione methylglyoxal lyase	7.2E-26
MSM1367	Msp_0035	predicted peptidyl-prolyl cis-trans isomerase 1	2.3E-63	MTH_1338	peptidyl-prolyl cis-trans isomerase B	1.9E-57
MSM1368	Msp_0037	ArgD	6.6E-121	MTH_1337	N-acetylornithine aminotransferase	8.1E-121
MSM1369	Msp_0006	predicted NUDIX-related protein	4.5E-12	MTH_1336	mutator MutT protein homolog	1.0E-17
MSM1370	Msp_0715	conserved hypothetical membrane-spanning protein	9.6E-97	NONE		
MSM1371	Msp_1578	LysA	2.9E-152	MTH_1335	diaminopimelate decarboxylase	2.3E-155
MSM1372	Msp_1579	DapF	1.3E-74	MTH_1334	diaminopimelate epimerase	2.8E-86
MSM1373	Msp_1545	conserved hypothetical protein	3.2E-50	MTH_1329	methyltransferase related protein	4.1E-46
MSM1374	Msp_1544	KsgA	1.6E-62	MTH_1326	dimethyladenosine transferase	1.3E-56
MSM1375	NONE			MTH_1325	conserved protein	2.9E-61
MSM1376	Msp_1543	conserved hypothetical protein	5.1E-20	MTH_1324	conserved protein	2.1E-28
MSM1377	Msp_1542	50S ribosomal protein L21e	3.3E-32	MTH_1323	ribosomal protein L21	2.7E-35
MSM1378	Msp_0981	conserved hypothetical protein	7.4E-19	NONE		

MSM1379	Msp_0967	putative NADP-dependent alcohol dehydrogenase	1.4E-24	NONE		
MSM1380	Msp_0967	putative NADP-dependent alcohol dehydrogenase	4.6E-74	NONE		
MSM1381	Msp_0967	putative NADP-dependent alcohol dehydrogenase	2.2E-11	NONE		
MSM1382	Msp_0504	conserved hypothetical membrane-spanning protein	2.7E-53	NONE		
MSM1383	Msp_0254	anaerobic ribonucleotide-triphosphate reductase	1.6E-307	MTH_1539	anaerobic ribonucleoside-triphosphate reductase	9.9E-306
MSM1384	Msp_0255	PolC	3.9E-290	MTH_1536	conserved protein	0.0E+00
MSM1385	Msp_0113	conserved hypothetical protein	7.7E-16	MTH_1626	phosphoserine phosphatase	2.3E-09
MSM1386	NONE			NONE		
MSM1387	Msp_0249	LysS	4.8E-205	MTH_1542	conserved protein	2.6E-202
MSM1388	Msp_0251	ThiC2	1.0E-156	MTH_1543	thiamine biosynthesis protein	5.3E-172
MSM1389	Msp_0252	predicted ribokinase	1.3E-78	MTH_1544	ribokinase	3.8E-91
MSM1390	Msp_0248	conserved hypothetical protein	2.5E-50	MTH_1545	conserved protein	1.5E-55
MSM1391	Msp_0247	predicted sugar phosphate isomerase	1.2E-52	MTH_1546	conserved protein	1.3E-51
MSM1392	NONE			NONE	nitrate assimilation protein, narQ	4.4E-58
MSM1393	NONE			NONE		
MSM1394	Msp_0355	conserved hypothetical membrane-spanning protein	1.5E-04	NONE		
MSM1395	Msp_0340	PstB	3.1E-27	MTH_605	ABC transporter	3.2E-30
MSM1396	NONE			MTH_1345	conserved protein	4.7E-22
MSM1397	Msp_0432	member of asn/thr-rich large protein family	7.3E-30	MTH_911	probable surface protein	3.0E-12
MSM1398	Msp_0762	member of asn/thr-rich large protein family	4.2E-21	MTH_716	cell surface glycoprotein (s-layer protein)	2.4E-10
MSM1399	Msp_0911	member of asn/thr-rich large protein family	5.8E-13	MTH_716	cell surface glycoprotein (s-layer protein)	4.7E-13
MSM1400	Msp_0615	partially conserved hypothetical membrane-spanning protein	5.3E-05	MTH_672	unknown	1.6E-04
MSM1401	Msp_1106	conserved hypothetical membrane-spanning protein	5.9E-42	MTH_671	unknown	1.9E-48
MSM1402	Msp_1107	conserved hypothetical membrane-spanning protein	4.2E-16	MTH_670	unknown	2.4E-11
MSM1403	NONE			NONE		
MSM1404	Msp_0243	FwdB	5.2E-23	NONE	formate dehydrogenase, alpha subunit homolog	1.9E-153
MSM1405	Msp_0639	FdhB	5.0E-84	NONE	formate dehydrogenase, beta subunit related protein FlpB	7.8E-84
MSM1406	Msp_0384	predicted Fe-S oxidoreductase	2.7E-19	MTH_1550	molybdenum cofactor biosynthesis MoaA	2.6E-99
MSM1407	Msp_0488	predicted allosteric regulator of homoserine dehydrogenase	9.7E-04	MTH_1551	molybdopterin-guanine dinucleotide biosynthesis protein B related	2.3E-36
MSM1408	Msp_0147	ferredoxin	7.5E-10	NONE	tungsten formylmethanofuran dehydrogenase, subunit H	8.3E-48
MSM1409	Msp_1447	EhbK	6.0E-18	NONE	tungsten formylmethanofuran dehydrogenase, subunit F	3.1E-97
MSM1410	Msp_0241	FwdG	1.8E-22	NONE	tungsten formylmethanofuran dehydrogenase, subunit G	2.7E-19
MSM1411	Msp_0242	FwdD	5.4E-39	NONE	tungsten formylmethanofuran dehydrogenase, subunit D	6.9E-21
MSM1412	Msp_0243	FwdB	1.6E-156	NONE	tungsten formylmethanofuran dehydrogenase, subunit B	5.3E-117
MSM1413	Msp_0244	FwdA	6.4E-203	NONE	tungsten formylmethanofuran dehydrogenase, subunit A	1.7E-182
MSM1414	Msp_0245	FwdC	1.9E-66	NONE	tungsten formylmethanofuran dehydrogenase, subunit C	2.9E-52
MSM1415	Msp_0246	hypothetical protein	3.9E-13	MTH_1568	unknown	1.1E-08
MSM1416	Msp_0246	hypothetical protein	6.8E-09	MTH_1568	unknown	1.6E-05
MSM1417	Msp_0235	conserved hypothetical membrane-spanning protein	2.9E-150	MTH_1569	conserved protein	6.5E-151
MSM1418	Msp_0234	GlnA	3.8E-157	MTH_1570	glutamine synthetase	4.7E-164
MSM1419	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM1420	Msp_0128	predicted helicase	5.7E-11	MTH_511	DNA helicase II	1.5E-13
MSM1421	Msp_1566	conserved hypothetical membrane-spanning protein	4.4E-92	NONE		
MSM1422	Msp_1568	conserved hypothetical membrane-spanning protein	3.5E-67	NONE		
MSM1423	Msp_0721	partially conserved hypothetical protein	5.9E-42	NONE		
MSM1424	Msp_0720	polyphosphate kinase	2.4E-258	NONE		
MSM1425	Msp_0871	30S ribosomal protein S13P	7.7E-56	MTH_34	ribosomal protein S18 (E.coli)	2.9E-54
MSM1426	Msp_0870	30S ribosomal protein S4P	6.5E-59	MTH_35	ribosomal protein S9 (E.coli)	4.4E-65
MSM1427	Msp_0869	30S ribosomal protein S11P	2.5E-59	MTH_36	ribosomal protein S14 (E.coli)	2.9E-61

MSM1428	Msp_0868	RpoD	6.3E-61	NONE	DNA-dependent RNA polymerase, subunit D	9.1E-74
MSM1429	Msp_0867	50S ribosomal protein L18e	1.1E-33	MTH_38	ribosomal protein L18 (E.coli)	5.5E-35
MSM1430	Msp_0866	50S ribosomal protein L13P	1.3E-51	MTH_39	ribosomal protein S16 (E.coli)	7.1E-58
MSM1431	Msp_0865	30S ribosomal protein S9P	2.9E-56	MTH_39	ribosomal protein S16 (E.coli)	1.3E-56
MSM1432	Msp_0864	RpoN	9.4E-19	NONE	DNA-dependent RNA polymerase, subunit N	1.3E-24
MSM1433	Msp_0863	RpoK	6.9E-16	NONE	DNA-dependent RNA polymerase, subunit K	2.4E-18
MSM1434	NONE			NONE		
MSM1435	Msp_0862	enolase	2.2E-113	MTH_43	enolase	3.0E-121
MSM1436	Msp_0861	ferredoxin	3.0E-15	MTH_1106	ferredoxin	6.2E-20
MSM1437	Msp_0860	ribosomal protein S2P	3.9E-84	MTH_44	ribosomal protein Sa (E.coli)	5.5E-83
MSM1438	Msp_0859	conserved hypothetical protein	1.9E-59	MTH_45	conserved protein	5.1E-64
MSM1439	Msp_0858	putative mevalonate kinase	2.1E-60	MTH_46	mevalonate kinase	4.6E-63
MSM1440	Msp_0857	predicted archaeal kinase	9.2E-60	MTH_47	conserved protein	3.6E-70
MSM1441	Msp_0856	isopentenyl-diphosphate delta-isomerase	6.2E-118	MTH_48	conserved protein	4.1E-117
MSM1442	Msp_0855	predicted hydrolase	8.3E-178	MTH_49	conserved protein	8.6E-188
MSM1443	Msp_0854	IdsA	1.3E-90	MTH_50	bifunctional short chain isoprenyl diphosphate synthase	4.1E-94
MSM1444	NONE			NONE		
MSM1445	Msp_1125	predicted transcriptional regulator	1.4E-38	MTH_1454	conserved protein	2.9E-45
MSM1446	Msp_1126	putative hydroxylamine reductase	1.8E-152	MTH_1453	6Fe-6S prismane-containing protein	3.6E-173
MSM1447	Msp_0002	conserved hypothetical protein	1.1E-31	MTH_1452	unknown	2.3E-36
MSM1448	Msp_1545	conserved hypothetical protein	1.9E-08	MTH_146	precorrin-8W decarboxylase	1.7E-05
MSM1449	Msp_0219	conserved hypothetical protein	7.9E-04	MTH_83	O-linked GlcNAc transferase	9.2E-05
MSM1450	Msp_0524	predicted oxidoreductase	8.4E-25	MTH_907	conserved protein	6.8E-08
MSM1451	Msp_0039	predicted glycosyltransferase	2.2E-06	MTH_83	O-linked GlcNAc transferase	3.2E-10
MSM1452	Msp_0923	GltX	1.1E-184	MTH_51	glutamyl-tRNA synthetase	8.5E-181
MSM1453	NONE			NONE		
MSM1454	Msp_0226	hypothetical protein	9.5E-14	NONE	heterodisulfide reductase, subunit C	6.6E-06
MSM1455	Msp_0924	predicted aspartate/tyrosine/aromatic aminotransferase	3.8E-166	MTH_52	aspartate aminotransferase related protein	6.6E-158
MSM1456	NONE			NONE		
MSM1457	NONE			NONE		
MSM1458	NONE			NONE		
MSM1459	Msp_0925	predicted arabinose efflux permease	7.3E-115	MTH_195	efflux pump antibiotic resistance protein	7.7E-93
MSM1460	Msp_1447	EhbK	1.8E-33	MTH_1133	polyferredoxin (MvhB)	5.8E-143
MSM1461	Msp_0638	MvhD2	1.3E-53	NONE	methyl viologen-reducing hydrogenase, delta subunit homolog FlpD	2.7E-58
MSM1462	Msp_0639	FdhB	1.2E-119	NONE	formate dehydrogenase, beta subunit related protein FlpB	1.9E-135
MSM1463	Msp_0640	FdhA	4.1E-50	NONE	formate dehydrogenase, alpha subunit related protein FlpC	2.0E-39
MSM1464	NONE			MTH_1141	conserved protein (FlpE)	1.2E-18
MSM1465	Msp_0925	predicted arabinose efflux permease	1.3E-115	MTH_195	efflux pump antibiotic resistance protein	9.5E-95
MSM1466	NONE			NONE		
MSM1467	NONE			NONE		
MSM1468	Msp_0986	PurA	7.6E-136	MTH_615	adenylosuccinate synthetase	9.4E-143
MSM1469	Msp_1164	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, periplasmic solute-binding protein	2.4E-91	MTH_924	molybdate-binding periplasmic protein	5.9E-06
MSM1470	NONE			NONE		
MSM1471	Msp_0919	predicted acyl-CoA synthetase	2.3E-237	NONE	succinyl-CoA synthetase, alpha subunit	2.5E-07
MSM1472	NONE			MTH_752	conserved protein	3.7E-77
MSM1473	Msp_0575	predicted metal-dependent hydrolase	2.9E-79	MTH_751	conserved protein	9.4E-72
MSM1474	Msp_0579	AroC	7.2E-124	MTH_748	chorismate synthase	4.7E-125
MSM1475	Msp_0497	putative endonuclease III	1.0E-14	MTH_746	endonuclease III related protein	2.1E-51
MSM1476	Msp_0416	HemB	6.2E-102	MTH_744	porphobilinogen synthase	3.6E-102

MSM1477	Msp_0428	predicted ATP:dephospho-CoA triphosphoribosyl transferase	1.7E-58	MTH_743	conserved protein	5.9E-70
MSM1478	Msp_0429	PheS	2.6E-165	MTH_742	phenylalanyl-tRNA synthetase	5.5E-170
MSM1479	NONE			MTH_212	exodeoxyribonuclease	2.4E-73
MSM1480	Msp_1260	predicted hydrolase	1.5E-59	MTH_209	conserved protein	1.1E-77
MSM1481	Msp_1281	conserved hypothetical protein	6.5E-59	MTH_208	DNA-dependent DNA polymerase family B (PolB2)	2.0E-69
MSM1482	NONE			NONE		
MSM1483	Msp_0195	ABC-type multidrug transport system, ATP-binding protein	2.0E-41	MTH_1093	ABC transporter (ATP-binding	1.4E-54
MSM1484	Msp_0196	ABC-type multidrug transport system, permease protein	8.1E-29	MTH_1486	conserved protein	1.0E-19
MSM1485	Msp_0440	member of asn/thr-rich large protein family	3.3E-06	NONE		
MSM1486	Msp_1280	30S ribosomal protein S8e	6.6E-34	MTH_207	ribosomal protein S8	1.5E-41
MSM1487	NONE			MTH_199	unknown	9.6E-31
MSM1488	Msp_0977	conserved hypothetical protein	3.1E-27	MTH_200	cobalamin biosynthesis protein M related protein	3.0E-50
MSM1489	Msp_0474	hypothetical protein	1.2E-09	MTH_1346	unknown	1.3E-177
MSM1490	Msp_0474	hypothetical protein	7.1E-06	MTH_201	unknown	4.9E-11
MSM1491	Msp_0474	hypothetical protein	9.8E-08	MTH_1346	unknown	1.3E-159
MSM1492	Msp_1279	HypE1	1.0E-122	MTH_205	hydrogenase expression/formation protein HypE	3.2E-126
MSM1493	Msp_1278	conserved hypothetical membrane-spanning protein	1.3E-21	MTH_204	conserved protein	4.3E-19
MSM1494	NONE			NONE		
MSM1495	Msp_1089	predicted nuclease	1.8E-40	MTH_494	thermonuclease precursor	8.5E-39
MSM1496	Msp_0024	hypothetical protein	4.5E-67	NONE		
MSM1497	NONE			MTH_1785	coenzyme PQQ synthesis protein	6.4E-57
MSM1498	Msp_1228	predicted helicase	2.1E-131	NONE	ATP-dependent RNA helicase, eIF-4A family	3.8E-114
MSM1499	Msp_1188	predicted transcriptional regulator	8.1E-61	MTH_163	conserved protein	2.5E-62
MSM1500	Msp_1189	RecJ	1.5E-114	MTH_164	single-stranded DNA exonuclease RecJ related protein	1.1E-116
MSM1501	Msp_1190	signal recognition particle, 19 kDa protein	4.0E-20	MTH_165	signal recognition particle 19 kDa protein	9.3E-17
MSM1502	Msp_0223	predicted UDP-galactopyranose mutase	3.6E-65	MTH_344	UDP-galactopyranose mutase	2.4E-80
MSM1503	Msp_0215	predicted glycosyltransferase	4.0E-39	MTH_884	teichoic acid biosynthesis related protein	2.4E-06
MSM1504	Msp_1191	HemD	2.2E-49	MTH_166	uroporphyrinogen III synthase	1.1E-52
MSM1505	NONE			NONE		
MSM1506	NONE			NONE		
MSM1507	Msp_0215	predicted glycosyltransferase	5.6E-34	MTH_884	teichoic acid biosynthesis related protein	7.4E-10
MSM1508	NONE			NONE		
MSM1509	NONE			NONE		
MSM1510	NONE			NONE		
MSM1511	NONE			NONE		
MSM1512	Msp_0060	putative lipooligosaccharide cholinephosphotransferase	7.0E-62	NONE		
MSM1513	Msp_0662	putative aspartate aminotransferase	2.7E-37	MTH_1601	aspartate aminotransferase	1.9E-41
MSM1514	Msp_1333	predicted dehydrogenase	1.3E-06	NONE	3-chlorobenzoate-3,4-dioxygenase dyhydrogenase related protein	8.7E-09
MSM1515	Msp_0060	putative lipooligosaccharide cholinephosphotransferase	1.1E-24	NONE		
MSM1516	Msp_1326	HisC	1.7E-26	MTH_1587	histidinol-phosphate aminotransferase	5.5E-22
MSM1517	NONE			MTH_1495	ornithine cyclodeaminase	1.2E-15
MSM1518	Msp_0017	conserved hypothetical protein	1.2E-11	NONE		
MSM1519	NONE			NONE		
MSM1520	NONE			NONE		
MSM1521	NONE			NONE		
MSM1522	NONE			NONE		
MSM1523	NONE			NONE		
MSM1524	NONE			NONE		
MSM1525	NONE			NONE		

MSM1526	Msp_0772	hypothetical membrane-spanning protein	2.3E-15	MTH_252	conserved protein	7.1E-19
MSM1527	NONE			NONE		
MSM1528	Msp_0608	predicted transcriptional regulator	1.9E-04	MTH_700	conserved protein	1.1E-04
MSM1529	NONE			NONE		
MSM1530	NONE			NONE		
MSM1531	Msp_0691	predicted Na+-dependent transporter	1.3E-131	NONE		
MSM1532	Msp_0691	predicted Na+-dependent transporter	2.0E-137	NONE		
MSM1533	Msp_1465	member of asn/thr-rich large protein family	7.2E-12	MTH_1074	putative membrane protein	3.7E-06
MSM1534	Msp_0590	member of asn/thr-rich large protein family	2.0E-24	MTH_1074	putative membrane protein	3.0E-123
MSM1535	Msp_1114	predicted dTDP-D-glucose 4,6-dehydratase	1.3E-10	NONE	dTDP-glucose 4,6-dehydratase	1.2E-06
MSM1536	Msp_0290	predicted pyridoxal phosphate-dependent enzyme	6.9E-71	MTH_1188	pleiotropic regulatory protein DegT	6.6E-71
MSM1537	Msp_0310	predicted GTP:adenosylcobinamide-phosphate guanylyltransferase	4.2E-04	NONE		
MSM1538	Msp_1202	predicted acetyltransferase	1.9E-08	NONE	N-terminal acetyltransferase complex, subunit ARD1	3.5E-06
MSM1539	NONE			NONE		
MSM1540	NONE			MTH_368	glycerol-3-phosphate dehydrogenase (NAD)	6.5E-48
MSM1541	NONE			NONE		
MSM1542	Msp_0310	predicted GTP:adenosylcobinamide-phosphate guanylyltransferase	4.6E-06	MTH_1152	conserved protein	1.4E-04
MSM1543	NONE			NONE		
MSM1544	Msp_0060	putative lipooligosaccharide cholinephosphotransferase	3.9E-22	NONE		
MSM1545	Msp_0495	predicted glycosyltransferase	1.3E-31	MTH_136	dolichyl-phosphate mannose synthase	1.4E-08
MSM1546	NONE			NONE		
MSM1547	Msp_1195	PurC	3.9E-77	MTH_170	phosphoribosylaminoimidazolesuccinocarboxamide synthase	6.8E-69
MSM1548	Msp_1194	predicted phosphoribosylformylglycinamidine synthase	1.2E-25	MTH_169	conserved protein	4.5E-24
MSM1549	Msp_1193	PurQ	2.4E-75	MTH_168	phosphoribosylformylglycinamidine synthase I	6.8E-85
MSM1550	Msp_1192	CobA	6.2E-86	MTH_167	S-adenosyl-L-methionine uroporphyrinogen methyltransferase	7.1E-90
MSM1551	Msp_1196	GlmS	1.5E-201	MTH_171	glutamine-fructose-6-phosphate transaminase	1.5E-208
MSM1552	NONE			NONE		
MSM1553	NONE			NONE		
MSM1554	Msp_0141	member of asn/thr-rich large protein family	1.1E-09	NONE		
MSM1555	Msp_0076	conserved hypothetical protein	3.5E-60	MTH_175	conserved protein	4.7E-77
MSM1556	Msp_1344	conserved hypothetical membrane-spanning protein	6.5E-75	NONE		
MSM1557	Msp_0520	predicted queuine/archaeosine tRNA-ribosyltransferase	5.0E-219	MTH_176	tRNA-guanine transglycosylase	1.2E-206
MSM1558	NONE			MTH_1329	methyltransferase related protein	3.1E-04
MSM1559	Msp_0063	predicted polysaccharide biosynthesis protein	9.5E-74	MTH_379	O-antigen transporter related protein	1.7E-72
MSM1560	Msp_0448	predicted polysaccharide biosynthesis protein	1.3E-78	MTH_379	O-antigen transporter related protein	4.9E-75
MSM1561	Msp_0117	predicted 3-hydroxy-3-methylglutaryl CoA synthase	3.6E-145	MTH_792	3-hydroxy-3-methylglutaryl-CoA-synthase	3.4E-145
MSM1562	Msp_0116	predicted thiolase	2.1E-156	MTH_793	lipid-transfer protein (sterol or nonspecific)	3.5E-168
MSM1563	NONE			NONE		
MSM1564	Msp_0087	CbiT	4.6E-05	NONE		
MSM1565	Msp_1226	CobQ	9.4E-154	MTH_787	cobyric acid synthase	1.1E-162
MSM1566	Msp_0233	conserved hypothetical protein	2.3E-22	NONE		
MSM1567	Msp_0762	member of asn/thr-rich large protein family	7.2E-35	MTH_1485	serine/threonine protein kinase related protein	5.1E-13
MSM1568	NONE			NONE		
MSM1569	Msp_1227	predicted ATP-dependent protease	2.4E-226	MTH_785	ATP-dependent protease LA	9.0E-241
MSM1570	Msp_0557	hypothetical protein	1.1E-127	MTH_530	UDP-N-acetylmuramyl tripeptide synthetase related protein	2.6E-25
MSM1571	NONE			NONE		
MSM1572	Msp_0683	hypothetical protein	4.9E-61	NONE		

MSM1573	NONE			NONE		
MSM1574	Msp_0797	predicted nitroreductase	6.3E-10	MTH_120	NADPH-oxidoreductase	4.2E-11
MSM1575	Msp_1055	hypothetical membrane-spanning protein	7.8E-04	MTH_521	unknown	8.2E-05
MSM1576	NONE			NONE		
MSM1577	Msp_1229	ribose-phosphate pyrophosphokinase	1.2E-84	MTH_784	ribose-phosphate pyrophosphokinase	1.0E-88
MSM1578	NONE			NONE		
MSM1579	Msp_0573	UvrB	1.2E-247	MTH_442	excinuclease ABC subunit B	1.2E-261
MSM1580	NONE			NONE		
MSM1581	Msp_0574	UvrA	0.0E+00	MTH_443	excinuclease ABC subunit A	0.0E+00
MSM1582	Msp_0603	conserved hypothetical membrane-spanning protein	5.6E-85	MTH_465	unknown	4.8E-84
MSM1583	Msp_1178	predicted helicase	7.4E-193	MTH_656	ATP-dependent RNA helicase related protein	2.1E-232
MSM1584	Msp_1119	conserved hypothetical protein	1.0E-37	MTH_641	conserved protein	2.9E-29
MSM1585	Msp_0983	member of asn/thr-rich large protein family	5.5E-38	MTH_911	probable surface protein	9.9E-06
MSM1586	Msp_0713	member of asn/thr-rich large protein family	1.8E-52	MTH_911	probable surface protein	3.7E-14
MSM1587	Msp_0590	member of asn/thr-rich large protein family	6.0E-44	MTH_716	cell surface glycoprotein (s-layer protein)	1.2E-06
MSM1588	NONE			NONE		
MSM1589	NONE			NONE		
MSM1590	Msp_0619	member of asn/thr-rich large protein family	2.5E-48	MTH_716	cell surface glycoprotein (s-layer protein)	1.3E-07
MSM1591	Msp_1118	conserved hypothetical protein	1.0E-37	MTH_639	conserved protein	5.6E-42
MSM1592	Msp_0205	predicted ABC-type polysaccharide/polyol phosphate export system, ATP-binding protein	9.8E-72	MTH_1370	ABC transporter (ATP-binding protein)	1.5E-20
MSM1593	Msp_0204	predicted ABC-type polysaccharide/polyol phosphate export system, permease protein	1.3E-53	MTH_1092	putative membrane protein	5.7E-11
MSM1594	Msp_0442	predicted glycosyltransferase	4.4E-60	MTH_884	teichoic acid biosynthesis related protein	1.5E-07
MSM1595	Msp_0929	predicted helicase	6.7E-04	NONE		
MSM1596	Msp_0017	conserved hypothetical protein	1.7E-28	NONE		
MSM1597	NONE			NONE		
MSM1598	NONE			NONE		
MSM1599	NONE			NONE		
MSM1600	NONE			NONE		
MSM1601	Msp_0692	hypothetical membrane-spanning protein	1.3E-07	NONE		
MSM1602	Msp_0220	predicted glycosyltransferase	6.9E-20	MTH_361	teichoic acid biosynthesis protein RodC related protein	1.7E-04
MSM1603	NONE			MTH_637	conserved protein	1.1E-20
MSM1604	Msp_1101	predicted UDP-glucose pyrophosphorylase	1.2E-103	MTH_634	UTP--glucose-1-phosphate uridylyltransferase	7.6E-109
MSM1605	NONE			NONE		
MSM1606	Msp_0612	predicted arylsulfatase regulatory protein	4.8E-102	MTH_114	arylsulfatase regulatory protein	1.9E-64
MSM1607	Msp_1060	hypothetical protein	2.4E-13	MTH_121	unknown	1.2E-05
MSM1608	Msp_1350	putative oxidoreductase	5.9E-97	MTH_907	conserved protein	8.1E-50
MSM1609	NONE			MTH_924	molybdate-binding periplasmic protein	6.6E-23
MSM1610	Msp_0342	PstC	1.1E-15	MTH_921	anion transport system permease protein	6.4E-25
MSM1611	Msp_1000	predicted ABC-type nitrate/sulfonate/bicarbonate transport system, ATB-binding protein	1.7E-28	MTH_920	anion permease	2.4E-34
MSM1612	Msp_0210	predicted UDP-glucose 6-dehydrogenase	6.3E-93	MTH_836	UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase	5.4E-24
MSM1613	NONE			NONE		
MSM1614	Msp_0394	predicted transcriptional regulator	1.3E-74	MTH_126	inosine-5'-monophosphate dehydrogenase related protein VII	2.1E-97
MSM1615	Msp_0395	putative deoxyhypusine synthase	7.4E-106	MTH_127	deoxyhypusine synthase	4.6E-95
MSM1616	Msp_0396	hypothetical membrane-spanning protein	4.0E-27	MTH_128	unknown	6.2E-27
MSM1617	Msp_0397	PyrF	1.9E-66	MTH_129	orotidine 5' monophosphate decarboxylase	4.3E-67
MSM1618	Msp_0398	CbiM1	6.0E-72	MTH_130	cobalamin biosynthesis protein M	9.5E-79
MSM1619	Msp_0399	CbiN	3.0E-31	MTH_131	cobalt transport protein N	7.2E-26

MSM1620	Msp_0400	CbiQ1	3.0E-38	MTH_132	cobalt transport protein Q	3.4E-42
MSM1621	Msp_0401	CbiO1	6.0E-88	MTH_133	cobalt transport ATP-binding protein O	9.3E-88
MSM1622	Msp_1239	RibC	6.9E-55	MTH_134	riboflavin synthase	2.3E-61
MSM1623	Msp_0541	predicted glycosyltransferase	2.1E-46	MTH_136	dolichyl-phosphate mannose synthase	6.1E-52
MSM1624	Msp_0542	hypothetical membrane-spanning protein	9.4E-19	MTH_137	unknown	1.2E-18
MSM1625	Msp_1044	TfrB	3.2E-34	MTH_1850	fumarate reductase	7.6E-33
MSM1626	Msp_1044	TfrB	3.0E-07	MTH_140	conserved protein	4.8E-107
MSM1627	Msp_0989	predicted glycosyltransferase	9.5E-11	MTH_377	dolichyl-phosphate mannose synthase related protein	2.0E-11
MSM1628	Msp_0430	conserved hypothetical protein	1.9E-75	MTH_141	conserved protein	7.0E-99
MSM1629	Msp_0431	GuaB	2.1E-163	MTH_142	inosine-5'-monophosphate dehydrogenase	1.5E-174
MSM1630	Msp_1253	50S ribosomal protein L37Ae	6.0E-33	MTH_681	ribosomal protein L37a	1.1E-36
MSM1631	NONE			NONE		
MSM1632	Msp_1254	partially conserved hypothetical protein	1.0E-21	MTH_680	conserved protein	1.4E-15
MSM1633	Msp_1255	conserved hypothetical protein	1.0E-12	MTH_679	unknown	5.3E-14
MSM1634	Msp_1256	partially conserved hypothetical protein	2.5E-27	MTH_678	conserved protein	2.1E-35
MSM1635	NONE			MTH_677	unknown	1.7E-10
MSM1636	Msp_1257	conserved hypothetical protein	2.6E-39	MTH_669	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase related protein	1.3E-58
MSM1637	Msp_0173	hypothetical membrane-spanning protein	9.9E-08	NONE		
MSM1638	Msp_1259	hypothetical membrane-spanning protein	1.6E-09	MTH_667	unknown	3.0E-11
MSM1639	Msp_0519	predicted Co/Zn/Cd cation transporter	4.1E-16	MTH_1893	cation efflux system protein (zinc/cadmium)	3.7E-17
MSM1640	Msp_0482	hypothetical membrane-spanning protein	1.8E-38	NONE		
MSM1641	NONE			NONE		
MSM1642	NONE			NONE		
MSM1643	NONE			NONE		
MSM1644	NONE			NONE		
MSM1645	NONE			NONE		
MSM1646	NONE			NONE		
MSM1647	NONE			NONE		
MSM1648	NONE			NONE		
MSM1649	NONE			NONE		
MSM1650	Msp_0260	hypothetical protein	7.9E-04	NONE		
MSM1651	NONE			NONE		
MSM1652	NONE			NONE		
MSM1653	NONE			NONE		
MSM1654	NONE			NONE		
MSM1655	Msp_1059	hypothetical protein	1.3E-05	NONE		
MSM1656	NONE			NONE		
MSM1657	Msp_0793	hypothetical protein	4.9E-06	NONE		
MSM1658	NONE			NONE		
MSM1659	NONE			NONE		
MSM1660	NONE			NONE		
MSM1661	NONE			NONE		
MSM1662	NONE			NONE		
MSM1663	NONE			NONE		
MSM1664	NONE			NONE		
MSM1665	NONE			NONE		
MSM1666	Msp_0946	conserved hypothetical protein	1.2E-05	NONE		
MSM1667	NONE			NONE		
MSM1668	NONE			NONE		
MSM1669	NONE			NONE		

MSM1670	Msp_0113	conserved hypothetical protein	1.8E-04	NONE		
MSM1671	NONE			NONE		
MSM1672	NONE			NONE		
MSM1673	Msp_0474	hypothetical protein	4.6E-04	NONE		
MSM1674	Msp_0822	hypothetical protein	2.5E-04	NONE		
MSM1675	NONE			NONE		
MSM1676	NONE			NONE		
MSM1677	NONE			NONE		
MSM1678	NONE			NONE		
MSM1679	NONE			NONE		
MSM1680	NONE			NONE		
MSM1681	NONE			NONE		
MSM1682	NONE			NONE		
MSM1683	NONE			NONE		
MSM1684	Msp_0912	member of asn/thr-rich large protein family	2.1E-06	MTH_412	conserved protein	4.7E-04
MSM1685	NONE			NONE		
MSM1686	NONE			NONE		
MSM1687	Msp_0658	hypothetical membrane-spanning protein	8.1E-07	MTH_1459	unknown	3.6E-07
MSM1688	NONE			NONE		
MSM1689	NONE			NONE		
MSM1690	NONE			NONE		
MSM1691	Msp_1039	partially conserved hypothetical membrane-spanning protein	1.5E-07	MTH_357	conserved protein	5.3E-08
MSM1692	NONE			NONE		
MSM1693	Msp_1258	predicted ribokinase	6.9E-39	MTH_668	unknown	1.8E-20
MSM1694	Msp_0929	predicted helicase	3.6E-193	MTH_487	DNA helicase related protein	4.9E-304
MSM1695	Msp_0572	UvrC	6.3E-164	MTH_441	excinuclease ABC subunit C	5.6E-161
MSM1696	Msp_1548	hypothetical protein	1.7E-08	NONE		
MSM1697	NONE			NONE		
MSM1698	Msp_0439	methyl-coenzyme M reductase, component A2-like protein	2.7E-147	NONE	methyl coenzyme M reductase system, component A2 homolog	5.4E-179
MSM1699	Msp_0438	predicted universal stress protein	2.1E-14	MTH_153	conserved protein	5.4E-21
MSM1700	Msp_1061	hypothetical protein	7.3E-12	MTH_278	ferredoxin	1.4E-20
MSM1701	Msp_1062	predicted dehydrogenase	4.0E-130	MTH_277	bacteriochlorophyll synthase 43 kDa subunit	8.8E-147
MSM1702	Msp_1088	ExoB	7.9E-102	MTH_631	UDP-glucose 4-epimerase	3.5E-97
MSM1703	NONE			MTH_647	unknown	5.0E-25
MSM1704	Msp_1122	PurF	1.4E-143	MTH_646	amidophosphoribosyltransferase	1.2E-156
MSM1705	Msp_1121	predicted peptidase	2.4E-100	MTH_645	collagenase	3.7E-100
MSM1706	Msp_1513	hypothetical membrane-spanning protein	2.9E-24	NONE		
MSM1707	Msp_1120	NifH	2.6E-96	MTH_643	nitrogenase NifH subunit	5.5E-99
MSM1708	NONE			NONE		
MSM1709	Msp_0440	member of asn/thr-rich large protein family	1.3E-35	MTH_716	cell surface glycoprotein (s-layer protein)	2.4E-04
MSM1710	Msp_1277	SerS	1.9E-187	MTH_1455	threonyl-tRNA synthetase	5.3E-06
MSM1711	Msp_0725	hypothetical protein	1.0E-08	NONE		
MSM1712	Msp_0852	predicted ferritin	8.4E-50	MTH_158	ferritin like protein (RsgA)	2.3E-59
MSM1713	Msp_1008	predicted regulatory protein	5.4E-32	MTH_162	unknown	1.5E-41
MSM1714	Msp_1040	coenzyme F390 synthetase II	6.3E-164	MTH_161	coenzyme F390 synthetase III	3.7E-164
MSM1715	Msp_1110	CobN	1.7E-68	MTH_714	magnesium chelatase subunit	0.0E+00
MSM1716	Msp_0590	member of asn/thr-rich large protein family	2.5E-16	MTH_717	unknown	3.9E-25
MSM1717	Msp_1105	predicted transporter	1.9E-52	MTH_672	unknown	2.3E-52
MSM1718	Msp_1106	conserved hypothetical membrane-spanning protein	2.0E-50	MTH_671	unknown	3.7E-61

MSM1719	Msp_1107	conserved hypothetical membrane-spanning protein	4.1E-25	MTH_670	unknown	1.2E-32
MSM1720	Msp_1533	RpoM1	7.3E-28	MTH_1314	transcription elongation factor TFIIS	8.6E-30
MSM1721	NONE			NONE		
MSM1722	Msp_0965	predicted nitroreductase	6.9E-16	MTH_120	NADPH-oxidoreductase	7.3E-33
MSM1723	Msp_1238	N(5),N(10)-methenyltetrahydromethanopterin cyclohydrolase	6.7E-105	NONE	N5,N10-methenyl-tetrahydromethanopterin cyclohydrolase	2.1E-138
MSM1724	Msp_0961	hypothetical membrane-spanning protein	3.1E-36	MTH_1192	conserved protein	9.2E-25
MSM1725	Msp_0961	hypothetical membrane-spanning protein	5.7E-28	MTH_1192	conserved protein	1.6E-30
MSM1726	Msp_0879	hypothetical membrane-spanning protein	9.0E-30	MTH_1192	conserved protein	1.3E-25
MSM1727	Msp_0844	predicted multimeric flavodoxin	1.2E-18	MTH_135	conserved protein	1.9E-18
MSM1728	NONE			NONE		
MSM1729	Msp_0587	hypothetical membrane-spanning protein	5.0E-29	MTH_520	unknown	3.9E-10
MSM1730	Msp_0607	hypothetical membrane-spanning protein	6.5E-20	MTH_1192	conserved protein	1.2E-26
MSM1731	Msp_0714	predicted short chain dehydrogenase	1.7E-115	NONE		
MSM1732	Msp_1548	hypothetical protein	8.2E-07	NONE		
MSM1733	Msp_0789	rubrerythrin	1.6E-39	MTH_756	rubrerythrin	3.3E-43
MSM1734	Msp_1237	ThyA	8.9E-28	MTH_774	thymidylate synthase	7.2E-26
MSM1735	Msp_0777	member of asn/thr-rich large protein family	7.4E-116	MTH_716	cell surface glycoprotein (s-layer protein)	1.4E-06
MSM1736	NONE			NONE		
MSM1737	NONE			NONE		
MSM1738	Msp_0154	member of asn/thr-rich large protein family	2.3E-06	NONE		
MSM1739	Msp_0987	hypothetical membrane-spanning protein	2.7E-07	MTH_521	unknown	1.4E-05
MSM1740	Msp_1323	conserved hypothetical protein	1.1E-16	MTH_83	O-linked GlcNAc transferase	4.7E-38
MSM1741	Msp_0113	conserved hypothetical protein	5.0E-05	NONE		
MSM1742	Msp_0482	hypothetical membrane-spanning protein	2.7E-76	NONE		
MSM1743	Msp_0113	conserved hypothetical protein	4.1E-06	NONE		
MSM1744	NONE			NONE		
MSM1745	Msp_0344	predicted phosphate uptake regulator	2.0E-04	NONE		
MSM1746	NONE			NONE		
MSM1747	Msp_0911	member of asn/thr-rich large protein family	8.1E-06	NONE		
MSM1748	NONE			NONE		
MSM1749	NONE			NONE		
MSM1750	NONE			NONE		
MSM1751	Msp_0113	conserved hypothetical protein	6.3E-15	NONE		
MSM1752	Msp_0702	conserved hypothetical protein	1.2E-59	MTH_1210	mrr restriction system related protein	3.4E-42
MSM1753	Msp_0465	conserved hypothetical membrane-spanning protein	6.7E-04	NONE		
MSM1754	Msp_1328	putative ATP-dependent protease La	3.6E-06	NONE		
MSM1755	Msp_0219	conserved hypothetical protein	6.7E-04	NONE		
MSM1756	Msp_0976	hypothetical protein	2.8E-05	NONE		
MSM1757	NONE			NONE		
MSM1758	NONE			NONE		
MSM1759	NONE			NONE		
MSM1760	NONE			NONE		
MSM1761	Msp_0113	conserved hypothetical protein	7.6E-07	MTH_540	intracellular protein transport protein	2.7E-05
MSM1762	NONE			NONE		
MSM1763	Msp_1533	RpoM1	4.6E-10	MTH_1314	transcription elongation factor TFIIS	3.1E-09
MSM1764	Msp_0226	hypothetical protein	8.9E-04	NONE		
MSM1765	NONE			NONE		
MSM1766	Msp_1323	conserved hypothetical protein	4.8E-15	MTH_83	O-linked GlcNAc transferase	3.4E-35
MSM1767	Msp_1548	hypothetical protein	1.3E-04	NONE		
MSM1768	NONE			NONE		

MSM1769	Msp_0724	hypothetical membrane-spanning protein	2.1E-08	MTH_1277	unknown	8.9E-05
MSM1770	Msp_0934	conserved hypothetical membrane-spanning protein	1.4E-17	MTH_518	conserved protein	3.4E-19
MSM1771	Msp_0128	predicted helicase	5.0E-19	MTH_511	DNA helicase II	1.1E-26
MSM1772	Msp_0725	hypothetical protein	4.0E-11	MTH_470	conserved protein	1.2E-04
MSM1773	Msp_1548	hypothetical protein	4.3E-07	MTH_521	unknown	7.7E-05
MSM1774	NONE			NONE		
MSM1775	NONE			NONE		
MSM1776	NONE			NONE		
MSM1777	Msp_0799	predicted transcriptional regulator	3.3E-05	MTH_671	unknown	2.6E-04
MSM1778	Msp_0726	hypothetical protein	2.7E-69	NONE		
MSM1779	Msp_0725	hypothetical protein	2.6E-119	NONE		
MSM1780	Msp_1055	hypothetical membrane-spanning protein	1.1E-10	MTH_1277	unknown	2.7E-06
MSM1781	Msp_0725	hypothetical protein	2.4E-13	MTH_470	conserved protein	1.4E-05
MSM1782	NONE			NONE		
MSM1783	NONE			NONE		
MSM1784	NONE			NONE		
MSM1785	NONE			NONE		
MSM1786	Msp_1323	conserved hypothetical protein	4.1E-07	MTH_83	O-linked GlcNAc transferase	6.9E-12
MSM1787	Msp_1323	conserved hypothetical protein	5.6E-09	MTH_72	O-linked GlcNAc transferase	3.6E-16
MSM1788	Msp_1323	conserved hypothetical protein	7.3E-11	MTH_83	O-linked GlcNAc transferase	2.0E-20
MSM1789	Msp_0757	predicted ATPase	2.5E-08	NONE		
MSM1790	Msp_0757	predicted ATPase	4.9E-08	NONE		
MSM1791	NONE			MTH_512	unknown	1.1E-25
MSM1792	Msp_0764	predicted nicotinate phosphoribosyltransferase	1.7E-193	NONE		
MSM1793	NONE			NONE		
MSM1794	Msp_1103	member of asn/thr-rich large protein family	1.5E-04	MTH_512	unknown	1.2E-24
MSM1795	Msp_0757	predicted ATPase	1.7E-99	NONE		

Table 9. Cluster of Orthologous Groups (COG) represented in the *M. smithii* proteome

A. Summary

Number of *M. smithii* genes in COG	Code	Functional Category
136	J	Translation
60	K	Transcription
78	L	Replication, Recombination and Repair
3	B	Chromatin Structure and Dynamics
6	D	Cell Cycle Control
26	V	Defense Mechanisms
8	T	Signal Transduction Mechanisms
59	M	Cell Wall/Membrane Biogenesis
3	N	Cell Motility
1	Z	Cytoskeleton
17	U	Intracellular Trafficking and Secretion
41	O	Post-translational Modification, Protein Turnover, Chaperones
121	C	Energy Production and Conversion
30	G	Carbohydrate Transport and Metabolism
82	E	Amino Acid Transport and Metabolism
42	F	Nucleic Acid Transport and Metabolism
92	H	Coenzyme Transport and Metabolism
18	I	Lipid Transport and Metabolism
57	P	Inorganic Ion Transport and Metabolism
1	Q	Secondary Metabolites Biosynthesis, Transport and Catabolism
201	R	General Function Prediction Only
171	S	Function Unknown
491	-	Not in COGs

B. *M. smithii* genes in each COG

# in COG	COG	Description	*M. smithii* gene(s)
Translation (J)			
1	COG0008	Glutamyl- and glutaminyl-tRNA synthetases	MSM1452
1	COG0009	Putative translation factor (SUA5)	MSM0612
1	COG0012	Predicted GTPase, probable translation factor	MSM1164
1	COG0013	Alanyl-tRNA synthetase	MSM0619
1	COG0016	Phenylalanyl-tRNA synthetase alpha subunit	MSM1478
1	COG0017	Aspartyl/asparaginyl-tRNA synthetases	MSM1236
1	COG0018	Arginyl-tRNA synthetase	MSM1231
1	COG0023	Translation initiation factor 1 (eIF-1/SUI1) and related proteins	MSM0754
1	COG0024	Methionine aminopeptidase	MSM1120
1	COG0030	Dimethyladenosine transferase (rRNA methylation)	MSM1374
1	COG0042	tRNA-dihydrouridine synthase	MSM0972
1	COG0048	Ribosomal protein S12	MSM0901
1	COG0049	Ribosomal protein S7	MSM0900
1	COG0051	Ribosomal protein S10	MSM0897
1	COG0060	Isoleucyl-tRNA synthetase	MSM1341
1	COG0064	Asp-tRNAAsn/Glu-tRNAGln amidotransferase B subunit (PET112 homolog)	MSM1101
1	COG0072	Phenylalanyl-tRNA synthetase beta subunit	MSM0277
1	COG0080	Ribosomal protein L11	MSM0623
1	COG0081	Ribosomal protein L1	MSM0622
1	COG0087	Ribosomal protein L3	MSM0762
1	COG0088	Ribosomal protein L4	MSM0761
1	COG0089	Ribosomal protein L23	MSM0760
1	COG0090	Ribosomal protein L2	MSM0759
1	COG0091	Ribosomal protein L22	MSM0757
1	COG0092	Ribosomal protein S3	MSM0756
1	COG0093	Ribosomal protein L14	MSM0751
1	COG0094	Ribosomal protein L5	MSM0748
1	COG0096	Ribosomal protein S8	MSM0746
1	COG0097	Ribosomal protein L6P/L9E	MSM0745
1	COG0098	Ribosomal protein S5	MSM0741
1	COG0099	Ribosomal protein S13	MSM1425
1	COG0100	Ribosomal protein S11	MSM1427
1	COG0101	Pseudouridylate synthase	MSM0855
1	COG0102	Ribosomal protein L13	MSM1430
1	COG0103	Ribosomal protein S9	MSM1431
1	COG0124	Histidyl-tRNA synthetase	MSM1181
1	COG0130	Pseudouridine synthase	MSM0732
1	COG0143	Methionyl-tRNA synthetase	MSM0071
1	COG0154	Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases	MSM1253
1	COG0162	Tyrosyl-tRNA synthetase	MSM0513
1	COG0172	Seryl-tRNA synthetase	MSM1710
1	COG0180	Tryptophanyl-tRNA synthetase	MSM0216
1	COG0182	Predicted translation initiation factor 2B subunit, eIF-2B alpha/beta/delta family	MSM0804
1	COG0184	Ribosomal protein S15P/S13E	MSM1194
1	COG0185	Ribosomal protein S19	MSM0758
1	COG0186	Ribosomal protein S17	MSM0752
1	COG0197	Ribosomal protein L16/L10E	MSM0989
1	COG0198	Ribosomal protein L24	MSM0750
1	COG0199	Ribosomal protein S14	MSM0747
1	COG0200	Ribosomal protein L15	MSM0739
1	COG0215	Cysteinyl-tRNA synthetase	MSM0268
1	COG0231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)	MSM0877
1	COG0244	Ribosomal protein L10	MSM0621
1	COG0255	Ribosomal protein L29	MSM0755
1	COG0256	Ribosomal protein L18	MSM0742
1	COG0293	23S rRNA methylase	MSM0508
1	COG0343	Queuine/archaeosine tRNA-ribosyltransferase	MSM1557
1	COG0423	Glycyl-tRNA synthetase (class II)	MSM0403
1	COG0441	Threonyl-tRNA synthetase	MSM1214
1	COG0442	Prolyl-tRNA synthetase	MSM0287
1	COG0480	Translation elongation factors (GTPases)	MSM0899
1	COG0495	Leucyl-tRNA synthetase	MSM1172
1	COG0522	Ribosomal protein S4 and related proteins	MSM1426
1	COG0525	Valyl-tRNA synthetase	MSM0275
1	COG0532	Translation initiation factor 2 (IF-2; GTPase)	MSM0202
1	COG0565	rRNA methylase	MSM0394
1	COG0621	2-methylthioadenine synthetase	MSM0845
1	COG0689	RNase PH	MSM0242
1	COG1093	Translation initiation factor 2, alpha subunit (eIF-2alpha)	MSM1133

1	COG1096	Predicted RNA-binding protein (consists of S1 domain and a Zn-ribbon domain)	MSM1357
1	COG1097	RNA-binding protein Rrp4 and related proteins (contain S1 domain and KH domain)	MSM0243
1	COG1258	Predicted pseudouridylate synthase	MSM1361
1	COG1325	Predicted exosome subunit	MSM0297
1	COG1358	Ribosomal protein HS6-type (S12/L30/L7a)	MSM0206
1	COG1369	RNase P/RNase MRP subunit POP5	MSM0246
1	COG1383	Ribosomal protein S17E	MSM0833
1	COG1384	Lysyl-tRNA synthetase (class I)	MSM1387
1	COG1471	Ribosomal protein S4E	MSM0749
1	COG1491	Predicted RNA-binding protein	MSM1375
1	COG1498	Protein implicated in ribosomal biogenesis, Nop56p homolog	MSM1046
1	COG1500	Predicted exosome subunit	MSM0244
1	COG1503	Peptide chain release factor 1 (eRF1)	MSM0891
1	COG1514	2'-5' RNA ligase	MSM0054
1	COG1534	Predicted RNA-binding protein containing KH domain, possibly ribosomal protein	MSM0710
2	COG1549	Queuine tRNA-ribosyltransferases, contain PUA domain	MSM0633, MSM0797
1	COG1552	Ribosomal protein L40E	MSM0125
1	COG1588	RNase P/RNase MRP subunit p29	MSM0753
1	COG1601	Translation initiation factor 2, beta subunit (eIF-2beta)/eIF-5 N-terminal domain	MSM0511
1	COG1603	RNase P/RNase MRP subunit p30	MSM0247
1	COG1631	Ribosomal protein L44E	MSM1135
1	COG1632	Ribosomal protein L15E	MSM0298
1	COG1670	Acetyltransferases, including N-acetylases of ribosomal proteins	MSM1573
1	COG1676	tRNA splicing endonuclease	MSM0217
1	COG1717	Ribosomal protein L32E	MSM0744
1	COG1727	Ribosomal protein L18E	MSM1429
1	COG1736	Diphthamide synthase subunit DPH2	MSM1358
1	COG1746	tRNA nucleotidyltransferase (CCA-adding enzyme)	MSM0053
1	COG1798	Diphthamide biosynthesis methyltransferase	MSM0801
1	COG1841	Ribosomal protein L30/L7E	MSM0740
1	COG1867	N2,N2-dimethylguanosine tRNA methyltransferase	MSM1031
1	COG1889	Fibrillarin-like rRNA methylase	MSM1047
1	COG1890	Ribosomal protein S3AE	MSM0661
1	COG1911	Ribosomal protein L30E	MSM0907
1	COG1976	Translation initiation factor 6 (eIF-6)	MSM0704
1	COG1997	Ribosomal protein L37AE/L43A	MSM1630
1	COG1998	Ribosomal protein S27AE	MSM0193
1	COG2004	Ribosomal protein S24E	MSM0194
1	COG2007	Ribosomal protein S8E	MSM1486
1	COG2016	Predicted RNA-binding protein (contains PUA domain)	MSM0183
1	COG2023	RNase P subunit RPR2	MSM0711
1	COG2051	Ribosomal protein S27E	MSM1134
1	COG2053	Ribosomal protein S28E/S33	MSM0205
1	COG2075	Ribosomal protein L24E	MSM0204
1	COG2092	Translation elongation factor EF-1beta	MSM0602
1	COG2097	Ribosomal protein L31E	MSM0705
1	COG2117	Predicted subunit of tRNA(5-methylaminomethyl-2-thiouridylate) methyltransferase, contains the PP-loop ATPase domain	MSM0707
1	COG2123	RNase PH-related exoribonuclease	MSM0241
1	COG2125	Ribosomal protein S6E (S10)	MSM0201
1	COG2126	Ribosomal protein L37E	MSM0181
1	COG2139	Ribosomal protein L21E	MSM1377
1	COG2147	Ribosomal protein L19E	MSM0743
1	COG2157	Ribosomal protein L20A (L18A)	MSM0703
1	COG2163	Ribosomal protein L14E/L6E/L27E	MSM0733
1	COG2167	Ribosomal protein L39E	MSM0706
1	COG2174	Ribosomal protein L34E	MSM0735
1	COG2238	Ribosomal protein S19E (S16A)	MSM0709
1	COG2260	Predicted Zn-ribbon RNA-binding protein	MSM1132
1	COG2263	Predicted RNA methylase	MSM0764
1	COG2511	Archaeal Glu-tRNAGln amidotransferase subunit E (contains GAD domain)	MSM0335
1	COG2519	tRNA(1-methyladenosine) methyltransferase and related methyltransferases	MSM1173
1	COG2888	Predicted Zn-ribbon RNA-binding protein with a function in translation	MSM0603
1	COG2890	Methylase of polypeptide chain release factors	MSM1373
1	COG3277	RNA-binding protein involved in rRNA processing	MSM0425
1	COG5256	Translation elongation factor EF-1alpha (GTPase)	MSM0898
1	COG5257	Translation initiation factor 2, gamma subunit (eIF-2gamma; GTPase)	MSM0200
Transcription (K)			
2	COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	MSM0910, MSM0911
2	COG0086	DNA-directed RNA polymerase, beta' subunit/160 kD subunit	MSM0908, MSM0909
1	COG0195	Transcription elongation factor	MSM0906
1	COG0202	DNA-directed RNA polymerase, alpha subunit/40 kD subunit	MSM1428

1	COG0250	Transcription antiterminator	MSM0624
1	COG0571	dsRNA-specific ribonuclease	MSM0176
1	COG0583	Transcriptional regulator	MSM1390
3	COG0640	Predicted transcriptional regulators	MSM0819, MSM1126, MSM1350
1	COG0789	Predicted transcriptional regulators	MSM0949
1	COG0846	NAD-dependent protein deacetylases, SIR2 family	MSM1087
1	COG0864	Predicted transcriptional regulators containing the CopG/Arc/MetJ DNA-binding domain and a metal-binding domain	MSM0364
1	COG1095	DNA-directed RNA polymerase, subunit E'	MSM0197
1	COG1293	Predicted RNA-binding protein homologous to eukaryotic snRNP	MSM0778
1	COG1308	Transcription factor homologous to NACalpha-BTF3	MSM0384
2	COG1309	Transcriptional regulator	MSM0094, MSM0650
1	COG1321	Mn-dependent transcriptional regulator	MSM0218
1	COG1378	Predicted transcriptional regulators	MSM1445
1	COG1395	Predicted transcriptional regulator	MSM0453
3	COG1396	Predicted transcriptional regulators	MSM0026, MSM0329, MSM1528
1	COG1405	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB	MSM0424
1	COG1476	Predicted transcriptional regulators	MSM1150
1	COG1497	Predicted transcriptional regulator	MSM1499
1	COG1522	Transcriptional regulators	MSM1032
1	COG1581	Archaeal DNA-binding protein	MSM1245
3	COG1594	DNA-directed RNA polymerase, subunit M/Transcription elongation factor TFIIS	MSM1354, MSM1720, MSM1763
1	COG1644	DNA-directed RNA polymerase, subunit N (RpoN/RPB10)	MSM1432
1	COG1675	Transcription initiation factor IIE, alpha subunit	MSM0631
1	COG1695	Predicted transcriptional regulators	MSM1250
1	COG1733	Predicted transcriptional regulators	MSM0864
1	COG1758	DNA-directed RNA polymerase, subunit K/omega	MSM1433
1	COG1761	DNA-directed RNA polymerase, subunit L	MSM1356
1	COG1777	Predicted transcriptional regulators	MSM1107
1	COG1813	Predicted transcription factor, homolog of eukaryotic MBF1	MSM0355
3	COG1846	Transcriptional regulators	MSM0413, MSM0600, MSM1230
2	COG1958	Small nuclear ribonucleoprotein (snRNP) homolog	MSM0182, MSM1220
1	COG1996	DNA-directed RNA polymerase, subunit RPC10 (contains C4-type Zn-finger)	MSM1631
1	COG2012	DNA-directed RNA polymerase, subunit H, RpoH/RPB5	MSM0912
1	COG2093	DNA-directed RNA polymerase, subunit E''	MSM0196
1	COG2101	TATA-box binding protein (TBP), component of TFIID and TFIIIB	MSM0720
1	COG2183	Transcriptional accessory protein	MSM1292
1	COG2207	AraC-type DNA-binding domain-containing proteins	MSM0775
1	COG2524	Predicted transcriptional regulator, contains C-terminal CBS domains	MSM1614
2	COG2865	Predicted transcriptional regulator containing an HTH domain and an uncharacterized domain shared with the mammalian protein Schlafen	MSM0540, MSM1315
1	COG4008	Predicted metal-binding transcription factor	MSM0969
3	COG4742	Predicted transcriptional regulator	MSM0404, MSM0817, MSM0818
Replication, Recombination and Repair (L)			
2	COG0084	Mg-dependent DNase	MSM0097, MSM0416
1	COG0122	3-methyladenine DNA glycosylase/8-oxoguanine DNA glycosylase	MSM1365
1	COG0164	Ribonuclease HII	MSM0979
2	COG0177	Predicted EndoIII-related endonuclease	MSM0272, MSM1584
1	COG0178	Excinuclease ATPase subunit	MSM1581
2	COG0188	Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), A subunit	MSM1353, MSM1775
5	COG0210	Superfamily I DNA and RNA helicases	MSM0058, MSM0113, MSM0731, MSM1420, MSM1771
1	COG0258	5'-3' exonuclease (including N-terminal domain of PolI)	MSM0725
1	COG0270	Site-specific DNA methylase	MSM0531
1	COG0322	Nuclease subunit of the excinuclease complex	MSM1695
1	COG0350	Methylated DNA-protein cysteine methyltransferase	MSM1185
1	COG0358	DNA primase (bacterial type)	MSM0427
2	COG0417	DNA polymerase elongation subunit (family B)	MSM1041, MSM1481
3	COG0419	ATPase involved in DNA repair	MSM0120, MSM0693, MSM1761
1	COG0420	DNA repair exonuclease	MSM0121
2	COG0468	RecA/RadA recombinase	MSM0611, MSM1333
2	COG0470	ATPase involved in DNA replication	MSM1176, MSM1177
1	COG0550	Topoisomerase IA	MSM0717
1	COG0556	Helicase subunit of the DNA excision repair complex	MSM1579
3	COG0582	Integrase	MSM0428, MSM1640, MSM1742
1	COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)	MSM1137
2	COG0608	Single-stranded DNA-specific exonuclease	MSM1193, MSM1500
1	COG0648	Endonuclease IV	MSM0963
1	COG0708	Exonuclease III	MSM1479
1	COG1041	Predicted DNA modification methylase	MSM0352
1	COG1107	Archaea-specific RecJ-like exonuclease, contains DnaJ-type Zn finger domain	MSM0260
1	COG1111	ERCC4-like helicases	MSM1187

2	COG1112	Superfamily I DNA and RNA helicases and helicase subunits	MSM1081, MSM1694
1	COG1193	Mismatch repair ATPase (MutS family)	MSM0524
1	COG1241	Predicted ATPase involved in replication control, Cdc46/Mcm family	MSM0510
1	COG1311	Archaeal DNA polymerase II, small subunit/DNA polymerase delta, subunit B	MSM1271
1	COG1343	Uncharacterized protein predicted to be involved in DNA repair	MSM0163
1	COG1389	DNA topoisomerase VI, subunit B	MSM0955
2	COG1468	RecB family exonuclease	MSM0165, MSM1059
2	COG1518	Uncharacterized protein predicted to be involved in DNA repair	MSM0023, MSM0164
1	COG1525	Micrococcal nuclease (thermonuclease) homologs	MSM1495
1	COG1533	DNA repair photolyase	MSM0543
1	COG1570	Exonuclease VII, large subunit	MSM0001
1	COG1583	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	MSM0170
1	COG1591	Holliday junction resolvase - archaeal type	MSM1098
1	COG1599	Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit and related ssDNA-binding proteins	MSM1332
1	COG1637	Predicted nuclease of the RecB family	MSM0497
1	COG1688	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)	MSM0167
1	COG1697	DNA topoisomerase VI, subunit A	MSM0956
1	COG1793	ATP-dependent DNA ligase	MSM0645
1	COG1857	Uncharacterized protein predicted to be involved in DNA repair	MSM0168
1	COG1933	Archaeal DNA polymerase II, large subunit	MSM1384
1	COG2219	Eukaryotic-type DNA primase, large subunit	MSM0073
1	COG2231	Uncharacterized protein related to Endonuclease III	MSM1475
2	COG3335	Transposase and inactivated derivatives	MSM0460, MSM1589
1	COG3359	Predicted exonuclease	MSM0138
2	COG3415	Transposase and inactivated derivatives	MSM0458, MSM1588
5	COG3464	Transposase and inactivated derivatives	MSM0087, MSM0230, MSM0396, MSM1093, MSM1566
1	COG3666	Transposase and inactivated derivatives	MSM1523
Chromatin Structure and Dynamics (B)			
3	COG2036	Histones H3 and H4	MSM0213, MSM0844, MSM1260
Cell Cycle Control (D)			
3	COG0037	Predicted ATPase of the PP-loop superfamily implicated in cell cycle control	MSM0553, MSM1028, MSM1178
1	COG0489	ATPases involved in chromosome partitioning	MSM0045
1	COG1077	Actin-like ATPase involved in cell morphogenesis	MSM0980
1	COG1192	ATPases involved in chromosome partitioning	MSM1241
Defense Mechanisms (V)			
5	COG0534	Na+-driven multidrug efflux pump	MSM0152, MSM0252, MSM0414, MSM1228, MSM1229
2	COG0577	ABC-type antimicrobial peptide transport system, permease component	MSM0856, MSM1400
2	COG0732	Restriction endonuclease S subunits	MSM0157, MSM0158
2	COG0842	ABC-type multidrug transport system, permease component	MSM1248, MSM1484
6	COG1002	Type II restriction enzyme, methylase subunits	MSM1743, MSM1744, MSM1745, MSM1746, MSM1747, MSM1748
3	COG1131	ABC-type multidrug transport system, ATPase component	MSM0593, MSM1249, MSM1483
2	COG1132	ABC-type multidrug transport system, ATPase and permease components	MSM0773, MSM0774
1	COG1136	ABC-type antimicrobial peptide transport system, ATPase component	MSM0857
1	COG1715	Restriction endonuclease	MSM1752
1	COG1968	Uncharacterized bacitracin resistance protein	MSM1201
1	COG4845	Chloramphenicol O-acetyltransferase	MSM0047
Signal Transduction Mechanisms (T)			
3	COG0589	Universal stress protein UspA and related nucleotide-binding proteins	MSM0485, MSM0887, MSM1699
5	COG3448	CBS-domain-containing membrane protein	MSM0305, MSM0484, MSM0790, MSM1053, MSM1054
Cell Wall/Membrane Biogenesis (M)			
1	COG0381	UDP-N-acetylglucosamine 2-epimerase	MSM0853
3	COG0399	Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis	MSM0347, MSM1030, MSM1536
4	COG0438	Glycosyltransferase	MSM0836, MSM1313, MSM1317, MSM1322
1	COG0449	Glucosamine 6-phosphate synthetase, contains amidotransferase and phosphosugar isomerase domains	MSM1551
14	COG0463	Glycosyltransferases involved in cell wall biogenesis	MSM0423, MSM1290, MSM1294, MSM1297, MSM1310, MSM1311, MSM1312, MSM1316, MSM1323, MSM1324, MSM1328, MSM1545, MSM1623, MSM1627
2	COG0472	UDP-N-acetylmuramyl pentapeptide phosphotransferase/UDP-N-acetylglucosamine-1-phosphate transferase	MSM0066, MSM0360
1	COG0562	UDP-galactopyranose mutase	MSM1502
1	COG0668	Small-conductance mechanosensitive channel	MSM0493
1	COG0677	UDP-N-acetyl-D-mannosaminuronate dehydrogenase	MSM1303

1	COG0707	pfam match to MurG; not predicted to be a carbohydrate active enzyme by CAZy	MSM0638
1	COG0750	Predicted membrane-associated Zn-dependent proteases 1	MSM1344
3	COG0769	UDP-N-acetylmuramyl tripeptide synthase	MSM0359, MSM1139, MSM1570
1	COG0770	UDP-N-acetylmuramyl pentapeptide synthase	MSM0880
1	COG0771	UDP-N-acetylmuramoylalanine-D-glutamate ligase	MSM0118
1	COG0773	UDP-N-acetylmuramate-alanine ligase	MSM1190
1	COG0794	Predicted sugar phosphate isomerase involved in capsule formation	MSM1391
1	COG1004	Predicted UDP-glucose 6-dehydrogenase	MSM1612
1	COG1083	CMP-N-acetylneuraminic acid synthetase	MSM0944
1	COG1087	UDP-glucose 4-epimerase	MSM1702
1	COG1088	dTDP-D-glucose 4,6-dehydratase	MSM1309
1	COG1091	dTDP-4-dehydrorhamnose reductase	MSM1304
1	COG1209	dTDP-glucose pyrophosphorylase	MSM1307
1	COG1210	UDP-glucose pyrophosphorylase	MSM1604
1	COG1861	Spore coat polysaccharide biosynthesis protein F, CMP-KDO synthetase homolog	MSM1537
1	COG1887	Putative glycosyl/glycerophosphate transferases involved in teichoic acid biosynthesis TagF/TagB/EpsJ/RodC	MSM1327
1	COG1898	dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes	MSM1308
1	COG2089	Sialic acid synthase	MSM1539
1	COG2148	Sugar transferases involved in lipopolysaccharide synthesis	MSM1331
1	COG2222	Predicted phosphosugar isomerases	MSM0872
2	COG2230	Cyclopropane fatty acid synthase and related methyltransferases	MSM0274, MSM0490
1	COG2843	Putative enzyme of poly-gamma-glutamate biosynthesis (capsule formation)	MSM0700
1	COG3049	Penicillin V acylase and related amidases	MSM0986
3	COG3475	LPS biosynthesis protein	MSM1512, MSM1515, MSM1544
1	COG3764	Sortase (surface protein transpeptidase)	MSM0984
1	COG3980	Spore coat polysaccharide biosynthesis protein, predicted glycosyltransferase	MSM1538
Cell Motility (N)			
1	COG3351	Putative archaeal flagellar protein D/E	MSM0137
2	COG5651	PPE-repeat proteins	MSM1586, MSM1590
Cytoskeleton (Z)			
1	COG5023	Tubulin	MSM1794
Intracellular Trafficking and Secretion (U)			
1	COG0201	Preprotein translocase subunit SecY	MSM0738
1	COG0541	Signal recognition particle GTPase	MSM1360
1	COG0552	Signal recognition particle GTPase	MSM0701
2	COG0681	Signal peptidase I	MSM0232, MSM1232
3	COG0811	Biopolymer transport proteins	MSM0978, MSM1401, MSM1718
1	COG0848	Biopolymer transport protein	MSM0977
1	COG1400	Signal recognition particle 19 kDa protein	MSM1501
1	COG2443	Preprotein translocase subunit Sss1	MSM0625
2	COG3210	Large exoproteins involved in heme utilization or adhesion	MSM0461, MSM1398
1	COG4023	Preprotein translocase subunit Sec61beta	MSM1363
1	COG4962	Flp pilus assembly protein, ATPase CpaF	MSM0597
2	COG4965	Flp pilus assembly protein TadB	MSM0471, MSM0596
Post-translational Modification, Protein Turnover, Chaperones (O)			
1	COG0068	Hydrogenase maturation factor	MSM1106
1	COG0071	Molecular chaperone (small heat shock protein)	MSM0870
1	COG0225	Peptide methionine sulfoxide reductase	MSM0582
1	COG0298	Hydrogenase maturation factor	MSM0636
1	COG0309	Hydrogenase maturation factor	MSM1492
1	COG0396	ABC-type transport system involved in Fe-S cluster assembly, ATPase component	MSM1003
1	COG0409	Hydrogenase maturation factor	MSM0945
1	COG0443	Molecular chaperone	MSM1109
3	COG0459	Chaperonin GroEL (HSP60 family)	MSM0220, MSM0826, MSM1533
1	COG0464	ATPases of the AAA+ class	MSM0642
1	COG0484	DnaJ-class molecular chaperone with C-terminal Zn finger domain	MSM1110
1	COG0492	Thioredoxin reductase	MSM0340
2	COG0501	Zn-dependent protease with chaperone function	MSM1174, MSM1203
1	COG0533	Metal-dependent proteases with possible chaperone activity	MSM1198
1	COG0576	Molecular chaperone GrpE (heat shock protein)	MSM1108
1	COG0602	Organic radical activating enzymes	MSM1055
2	COG0638	20S proteasome, alpha and beta subunits	MSM0245, MSM1037
1	COG0652	Peptidyl-prolyl cis-trans isomerase (rotamase) - cyclophilin family	MSM1367
1	COG0719	ABC-type transport system involved in Fe-S cluster assembly, permease component	MSM1002
1	COG0785	Cytochrome c biogenesis protein	MSM0549
3	COG0826	Collagenase and related proteases	MSM0522, MSM0523, MSM1705
1	COG1047	FKBP-type peptidyl-prolyl cis-trans isomerases 2	MSM0930
1	COG1067	Predicted ATP-dependent protease	MSM1569
3	COG1180	Pyruvate-formate lyase-activating enzyme	MSM0538, MSM0652, MSM1284

1	COG1222	ATP-dependent 26S proteasome regulatory subunit	MSM0354
1	COG1382	Prefoldin, chaperonin cofactor	MSM1634
1	COG1397	ADP-ribosylglycohydrolase	MSM1572
1	COG1730	Predicted prefoldin, molecular chaperone implicated in de novo protein folding	MSM0702
1	COG1899	Deoxyhypusine synthase	MSM1615
1	COG1973	Hydrogenase maturation factor	MSM1158
1	COG2143	Thioredoxin-related protein	MSM0550
1	COG4070	Predicted peptidyl-prolyl cis-trans isomerase (rotamase), cyclophilin family	MSM0813
1	COG4930	Predicted ATP-dependent Lon-type protease	MSM1754
Energy Production and Conversion (C)			
1	COG0045	Succinyl-CoA synthetase, beta subunit	MSM0924
1	COG0074	Succinyl-CoA synthetase, alpha subunit	MSM0228
1	COG0221	Inorganic pyrophosphatase	MSM0198
1	COG0240	Glycerol-3-phosphate dehydrogenase	MSM1540
2	COG0243	Anaerobic dehydrogenases, typically selenocysteine-containing	MSM1404, MSM1463
1	COG0247	Fe-S oxidoreductase	MSM1625
1	COG0371	Glycerol dehydrogenase and related enzymes	MSM0286
1	COG0372	Citrate synthase	MSM0446
2	COG0426	Uncharacterized flavoproteins	MSM0222, MSM1349
1	COG0479	Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit	MSM0393
1	COG0636	F0F1-type ATP synthase, subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K	MSM0439
1	COG0644	Dehydrogenases (flavoproteins)	MSM1701
2	COG0650	Formate hydrogenlyase subunit 4	MSM0317, MSM1062
3	COG0674	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, alpha subunit	MSM0332, MSM0559, MSM0927
1	COG0680	Ni,Fe-hydrogenase maturation factor	MSM1123
3	COG0716	Flavodoxins	MSM0062, MSM0503, MSM0861
1	COG0731	Fe-S oxidoreductases	MSM0922
4	COG0778	Nitroreductase	MSM0445, MSM1293, MSM1574, MSM1722
1	COG0822	NifU homolog involved in Fe-S cluster formation	MSM0263
1	COG1012	NAD-dependent aldehyde dehydrogenases	MSM0467
3	COG1013	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, beta subunit	MSM0333, MSM0560, MSM0926
3	COG1014	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, gamma subunit	MSM0391, MSM0557, MSM0925
1	COG1029	Formylmethanofuran dehydrogenase subunit B	MSM1412
2	COG1032	Fe-S oxidoreductase	MSM0696, MSM0787
4	COG1035	Coenzyme F420-reducing hydrogenase, beta subunit	MSM0135, MSM1121, MSM1405, MSM1462
1	COG1036	Archaeal flavoproteins	MSM1338
1	COG1042	Acyl-CoA synthetase (NDP forming)	MSM1471
1	COG1053	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit	MSM1258
1	COG1139	Uncharacterized conserved protein containing a ferredoxin-like domain	MSM1626
2	COG1142	Fe-S-cluster-containing hydrogenase components 2	MSM0561, MSM0562
2	COG1143	Formate hydrogenlyase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I)	MSM0998, MSM1065
1	COG1144	Pyruvate:ferredoxin oxidoreductase and related 2-oxoacid:ferredoxin oxidoreductases, delta subunit	MSM0558
12	COG1145	Ferredoxin	MSM0136, MSM0306, MSM0310, MSM0311, MSM0395, MSM0579, MSM0783, MSM0784, MSM1066, MSM1409, MSM1410, MSM1700
5	COG1146	Ferredoxin	MSM0085, MSM0209, MSM0331, MSM0928, MSM1408
2	COG1148	Heterodisulfide reductase, subunit A and related polyferredoxins	MSM0082, MSM1336
2	COG1150	Heterodisulfide reductase, subunit C	MSM0084, MSM0796
1	COG1151	6Fe-6S prismane cluster-containing protein	MSM1446
1	COG1153	Formylmethanofuran dehydrogenase subunit D	MSM1411
1	COG1155	Archaeal/vacuolar-type H+-ATPase subunit A	MSM0435
1	COG1156	Archaeal/vacuolar-type H+-ATPase subunit B	MSM0434
1	COG1229	Formylmethanofuran dehydrogenase subunit A	MSM1413
1	COG1249	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes	MSM0637
1	COG1269	Archaeal/vacuolar-type H+-ATPase subunit I	MSM0440
1	COG1304	L-lactate dehydrogenase (FMN-dependent) and related alpha-hydroxy acid dehydrogenases	MSM1441
1	COG1390	Archaeal/vacuolar-type H+-ATPase subunit E	MSM0438
1	COG1394	Archaeal/vacuolar-type H+-ATPase subunit D	MSM0433
2	COG1413	FOG: HEAT repeat	MSM0372, MSM0501
1	COG1436	Archaeal/vacuolar-type H+-ATPase subunit F	MSM0436
2	COG1526	Uncharacterized protein required for formate dehydrogenase activity	MSM0295, MSM1392
1	COG1527	Archaeal/vacuolar-type H+-ATPase subunit C	MSM0437
2	COG1592	Rubrerythrin	MSM1348, MSM1733

1	COG1600	Uncharacterized Fe-S protein	MSM0609
1	COG1625	Fe-S oxidoreductase, related to NifB/MoaA family	MSM1020
2	COG1773	Rubredoxin	MSM0187, MSM0188
2	COG1838	Tartrate dehydratase beta subunit/Fumarate hydratase class I, C-terminal domain	MSM0769, MSM0929
2	COG1908	Coenzyme F420-reducing hydrogenase, delta subunit	MSM1001, MSM1461
2	COG1941	Coenzyme F420-reducing hydrogenase, gamma subunit	MSM1000, MSM1122
2	COG1951	Tartrate dehydratase alpha subunit/Fumarate hydratase class I, N-terminal domain	MSM0447, MSM0563
1	COG2033	Desulfoferrodoxin	MSM0262
2	COG2037	Formylmethanofuran:tetrahydromethanopterin formyltransferase	MSM0308, MSM1092
2	COG2048	Heterodisulfide reductase, subunit B	MSM0083, MSM0795
1	COG2055	Malate/L-lactate dehydrogenases	MSM1040
1	COG2141	Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase and related flavin-dependent oxidoreductases	MSM0542
1	COG2191	Formylmethanofuran dehydrogenase subunit E	MSM1396
1	COG2218	Formylmethanofuran dehydrogenase subunit C	MSM1414
1	COG2710	Nitrogenase molybdenum-iron protein, alpha and beta chains	MSM1160
1	COG2811	Archaeal/vacuolar-type H+-ATPase subunit H	MSM0441
2	COG3259	Coenzyme F420-reducing hydrogenase, alpha subunit	MSM0999, MSM1124
1	COG3260	Ni,Fe-hydrogenase III small subunit	MSM1064
1	COG3261	Ni,Fe-hydrogenase III large subunit	MSM1063
2	COG4231	Indolepyruvate ferredoxin oxidoreductase, alpha and beta subunits	MSM0392, MSM1460
1	COG5016	Pyruvate/oxaloacetate carboxyltransferase	MSM0939
Carbohydrate Transport and Metabolism (G)			
1	COG0057	Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	MSM0962
1	COG0063	Predicted sugar kinase	MSM1091
1	COG0120	Ribose 5-phosphate isomerase	MSM0284
1	COG0126	3-phosphoglycerate kinase	MSM0918
1	COG0148	Enolase	MSM1435
1	COG0149	Triosephosphate isomerase	MSM0919
1	COG0235	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases	MSM1270
1	COG0483	Archaeal fructose-1,6-bisphosphatase and related enzymes of inositol monophosphatase family	MSM0879
3	COG0524	Sugar kinases, ribokinase family	MSM0307, MSM1389, MSM1693
2	COG0574	Phosphoenolpyruvate synthase/pyruvate phosphate dikinase	MSM0823, MSM0988
1	COG0580	Glycerol uptake facilitator and related permeases (Major Intrinsic Protein Family)	MSM1085
2	COG1082	Sugar phosphate isomerases/epimerases	MSM1184, MSM1251
2	COG1109	Phosphomannomutase	MSM0648, MSM0656
1	COG1363	Cellulase M and related proteins	MSM0134
1	COG1830	DhnA-type fructose-1,6-bisphosphate aldolase and related enzymes	MSM0056
1	COG1980	Archaeal fructose 1,6-bisphosphatase	MSM0615
2	COG2074	2-phosphoglycerate kinase	MSM0408, MSM0791
2	COG2730	Endoglucanase	MSM1051, MSM1125
2	COG2814	Arabinose efflux permease	MSM1459, MSM1465
2	COG3635	Predicted phosphoglycerate mutase, AP superfamily	MSM0153, MSM0657
1	COG5297	Cellobiohydrolase A (1,4-beta-cellobiosidase A)	MSM0958
Amino Acid Transport and Metabolism (E)			
1	COG0002	Acetylglutamate semialdehyde dehydrogenase	MSM0860
1	COG0006	Xaa-Pro aminopeptidase	MSM0472
1	COG0019	Diaminopimelate decarboxylase	MSM1371
1	COG0031	Cysteine synthase	MSM0271
1	COG0040	ATP phosphoribosyltransferase	MSM1261
2	COG0065	3-isopropylmalate dehydratase large subunit	MSM0723, MSM1300
2	COG0066	3-isopropylmalate dehydratase small subunit	MSM0847, MSM1299
1	COG0067	Glutamate synthase domain 1	MSM0370
2	COG0069	Glutamate synthase domain 2	MSM0027, MSM0368
1	COG0070	Glutamate synthase domain 3	MSM0369
2	COG0075	Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	MSM0677, MSM1513
1	COG0076	Glutamate decarboxylase and related PLP-dependent proteins	MSM0987
1	COG0077	Prephenate dehydratase	MSM1052
1	COG0078	Ornithine carbamoyltransferase	MSM1226
2	COG0079	Histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase	MSM0653, MSM1516
1	COG0082	Chorismate synthase	MSM1474
1	COG0106	Phosphoribosylformimino-5-aminoimidazole carboxamide ribonucleotide (ProFAR) isomerase	MSM0858
1	COG0107	Imidazoleglycerol-phosphate synthase	MSM1364
1	COG0112	Glycine/serine hydroxymethyltransferase	MSM1337
1	COG0118	Glutamine amidotransferase	MSM1159
3	COG0119	Isopropylmalate/homocitrate/citramalate synthases	MSM0350, MSM0722, MSM1246
1	COG0128	5-enolpyruvylshikimate-3-phosphate synthase	MSM0273
1	COG0131	Imidazoleglycerol-phosphate dehydratase	MSM1206
1	COG0133	Tryptophan synthase beta chain	MSM1142

1	COG0134	Indole-3-glycerol phosphate synthase	MSM1143
1	COG0136	Aspartate-semialdehyde dehydrogenase	MSM0829
1	COG0137	Argininosuccinate synthase	MSM1084
1	COG0139	Phosphoribosyl-AMP cyclohydrolase	MSM1182
1	COG0140	Phosphoribosyl-ATP pyrophosphohydrolase	MSM1103
1	COG0141	Histidinol dehydrogenase	MSM1238
1	COG0165	Argininosuccinate lyase	MSM0192
1	COG0169	Shikimate 5-dehydrogenase	MSM1179
1	COG0174	Glutamine synthetase	MSM1418
1	COG0253	Diaminopimelate epimerase	MSM1372
1	COG0287	Prephenate dehydrogenase	MSM0641
1	COG0289	Dihydrodipicolinate reductase	MSM0830
1	COG0334	Glutamate dehydrogenase/leucine dehydrogenase	MSM0888
1	COG0345	Pyrroline-5-carboxylate reductase	MSM0089
1	COG0346	Lactoylglutathione lyase and related lyases	MSM1366
1	COG0347	Nitrogen regulatory protein PII	MSM0233
1	COG0367	Asparagine synthase (glutamine-hydrolyzing)	MSM0160
3	COG0436	Aspartate/tyrosine/aromatic aminotransferase	MSM0610, MSM0788, MSM1455
1	COG0440	Acetolactate synthase, small (regulatory) subunit	MSM1224
1	COG0460	Homoserine dehydrogenase	MSM0154
1	COG0498	Threonine synthase	MSM0214
1	COG0527	Aspartokinases	MSM0832
1	COG0547	Anthranilate phosphoribosyltransferase	MSM1144
1	COG0548	Acetylglutamate kinase	MSM0375
1	COG0560	Phosphoserine phosphatase	MSM0719
1	COG0620	Methionine synthase II (cobalamin-independent)	MSM0102
1	COG0710	3-dehydroquinate dehydratase	MSM0231
1	COG0747	ABC-type dipeptide transport system, periplasmic component	MSM0300
1	COG0765	ABC-type amino acid transport system, permease component	MSM0806
1	COG1045	Serine acetyltransferase	MSM0270
1	COG1104	Cysteine sulfinate desulfinase/cysteine desulfurase and related enzymes	MSM0264
1	COG1125	ABC-type proline/glycine betaine transport systems, ATPase components	MSM0990
1	COG1126	ABC-type polar amino acid transport system, ATPase component	MSM0805
1	COG1168	Bifunctional PLP-dependent enzyme with beta-cystathionase and maltose regulon repressor activities	MSM0044
1	COG1174	ABC-type proline/glycine betaine transport systems, permease component	MSM0991
2	COG1305	Transglutaminase-like enzymes, putative cysteine proteases	MSM0219, MSM0786
1	COG1465	Predicted alternative 3-dehydroquinate synthase	MSM0055
1	COG1605	Chorismate mutase	MSM0834
1	COG1812	Archaeal S-adenosylmethionine synthetase	MSM1340
1	COG1921	Selenocysteine synthase [seryl-tRNASer selenium transferase]	MSM0767
1	COG2021	Homoserine acetyltransferase	MSM0496
1	COG2061	ACT-domain-containing protein, predicted allosteric regulator of homoserine dehydrogenase	MSM0155
1	COG2303	Choline dehydrogenase and related flavoproteins	MSM0865
1	COG2423	Predicted ornithine cyclodeaminase, mu-crystallin homolog	MSM1517
1	COG2856	Predicted Zn peptidase	MSM1529
2	COG2873	O-acetylhomoserine sulfhydrylase	MSM0174, MSM0265
1	COG4992	Ornithine/acetylornithine aminotransferase	MSM1368
Nucleic Acid Transport and Metabolism (F)			
1	COG0005	Purine nucleoside phosphorylase	MSM0665
1	COG0015	Adenylosuccinate lyase	MSM1151
1	COG0034	Glutamine phosphoribosylpyrophosphate amidotransferase	MSM1704
1	COG0035	Uracil phosphoribosyltransferase	MSM0398
1	COG0041	Phosphoribosylcarboxyaminoimidazole (NCAIR) mutase	MSM1287
1	COG0044	Dihydroorotase and related cyclic amidohydrolases	MSM0997
1	COG0046	Phosphoribosylformylglycinamidine (FGAM) synthase, synthetase domain	MSM1342
1	COG0047	Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain	MSM1549
1	COG0104	Adenylosuccinate synthase	MSM1468
1	COG0105	Nucleoside diphosphate kinase	MSM0203
2	COG0125	Thymidylate kinase	MSM0077, MSM0520
1	COG0127	Xanthosine triphosphate pyrophosphatase	MSM1195
1	COG0150	Phosphoribosylaminoimidazole (AIR) synthetase	MSM1039
1	COG0151	Phosphoribosylamine-glycine ligase	MSM1227
1	COG0152	Phosphoribosylaminoimidazolesuccinocarboxamide (SAICAR) synthase	MSM1547
1	COG0167	Dihydroorotate dehydrogenase	MSM1044
1	COG0207	Thymidylate synthase	MSM1734
1	COG0274	Deoxyribose-phosphate aldolase	MSM0843
1	COG0284	Orotidine-5'-phosphate decarboxylase	MSM1617
1	COG0461	Orotate phosphoribosyltransferase	MSM0821
1	COG0503	Adenine/guanine phosphoribosyltransferases and related PRPP-binding proteins	MSM1359
1	COG0504	CTP synthase (UTP-ammonia lyase)	MSM0147
1	COG0516	IMP dehydrogenase/GMP reductase	MSM1629
1	COG0518	GMP synthase - Glutamine amidotransferase domain	MSM0343

1	COG0519	GMP synthase, PP-ATPase domain/subunit	MSM0345
1	COG0528	Uridylate kinase	MSM0415
1	COG0540	Aspartate carbamoyltransferase, catalytic chain	MSM1263
2	COG0717	Deoxycytidine deaminase	MSM0402, MSM0687
1	COG0856	Orotate phosphoribosyltransferase homologs	MSM0883
1	COG1001	Adenine deaminase	MSM0874
1	COG1051	ADP-ribose pyrophosphatase	MSM1355
1	COG1102	Cytidylate kinase	MSM0734
1	COG1328	Oxygen-sensitive ribonucleoside-triphosphate reductase	MSM1383
1	COG1437	Adenylate cyclase, class 2 (thermophilic)	MSM0721
1	COG1781	Aspartate carbamoyltransferase, regulatory subunit	MSM0862
1	COG1828	Phosphoribosylformylglycinamidine (FGAM) synthase, PurS component	MSM1548
1	COG1936	Predicted nucleotide kinase (related to CMP and AMP kinases)	MSM0713
1	COG2019	Archaeal adenylate kinase	MSM0737
1	COG2233	Xanthine/uracil permeases	MSM0397
1	COG3363	Archaeal IMP cyclohydrolase	MSM0976
Coenzyme Transport and Metabolism (H)			
1	COG0001	Glutamate-1-semialdehyde aminotransferase	MSM1233
1	COG0007	Uroporphyrinogen-III methylase	MSM1550
1	COG0043	3-polyprenyl-4-hydroxybenzoate decarboxylase and related decarboxylases	MSM1286
1	COG0054	Riboflavin synthase beta-chain	MSM1296
1	COG0108	3,4-dihydroxy-2-butanone 4-phosphate synthase	MSM1256
1	COG0113	Delta-aminolevulinic acid dehydratase	MSM1476
1	COG0142	Geranylgeranyl pyrophosphate synthase	MSM1443
1	COG0157	Nicotinate-nucleotide pyrophosphorylase	MSM0491
1	COG0163	3-polyprenyl-4-hydroxybenzoate decarboxylase	MSM0237
1	COG0171	NAD synthase	MSM1171
1	COG0181	Porphobilinogen deaminase	MSM0881
1	COG0237	Dephospho-CoA kinase	MSM0141
1	COG0294	Dihydropteroate synthase and related enzymes	MSM0556
1	COG0301	Thiamine biosynthesis ATP pyrophosphatase	MSM0617
2	COG0303	Molybdopterin biosynthesis enzyme	MSM0950, MSM1343
1	COG0311	Predicted glutamine amidotransferase involved in pyridoxine biosynthesis	MSM0371
1	COG0314	Molybdopterin converting factor, large subunit	MSM0130
1	COG0315	Molybdenum cofactor biosynthesis enzyme	MSM1362
1	COG0340	Biotin-(acetyl-CoA carboxylase) ligase	MSM0766
1	COG0351	Hydroxymethylpyrimidine/phosphomethylpyrimidine kinase	MSM0289
1	COG0352	Thiamine monophosphate synthase	MSM0917
1	COG0373	Glutamyl-tRNA reductase	MSM0967
1	COG0379	Quinolinate synthase	MSM0494
1	COG0382	4-hydroxybenzoate polyprenyltransferase and related prenyltransferases	MSM0941
1	COG0407	Uroporphyrinogen-III decarboxylase	MSM0518
2	COG0422	Thiamine biosynthesis protein ThiC	MSM0644, MSM1388
2	COG0452	Phosphopantothenoylcysteine synthetase/decarboxylase	MSM1048, MSM1049
1	COG0476	Dinucleotide-utilizing enzymes involved in molybdopterin and thiamine biosynthesis family 2	MSM0729
1	COG0499	S-adenosylhomocysteine hydrolase	MSM0727
2	COG0502	Biotin synthase and related enzymes	MSM0573, MSM1099
1	COG0521	Molybdopterin biosynthesis enzymes	MSM0820
1	COG0611	Thiamine monophosphate kinase	MSM1283
1	COG0684	Demethylmenaquinone methyltransferase	MSM0426
1	COG0720	6-pyruvoyl-tetrahydropterin synthase	MSM1056
1	COG0746	Molybdopterin-guanine dinucleotide biosynthesis protein A	MSM0240
1	COG1010	Precorrin-3B methylase	MSM1273
1	COG1056	Nicotinamide mononucleotide adenylyltransferase	MSM0129
1	COG1270	Cobalamin biosynthesis protein CobD/CbiB	MSM1266
2	COG1429	Cobalamin biosynthesis protein CobN and related Mg-chelatases	MSM1117, MSM1715
1	COG1488	Nicotinic acid phosphoribosyltransferase	MSM1792
2	COG1492	Cobyric acid synthase	MSM1254, MSM1565
2	COG1541	Coenzyme F390 synthetase	MSM0387, MSM1714
1	COG1587	Uroporphyrinogen-III synthase	MSM1504
1	COG1648	Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)	MSM0968
1	COG1731	Archaeal riboflavin synthase	MSM1622
1	COG1763	Molybdopterin-guanine dinucleotide biosynthesis protein	MSM1407
1	COG1767	Triphosphoribosyl-dephospho-CoA synthetase	MSM1477
1	COG1797	Cobyrinic acid a,c-diamide synthase	MSM1215
1	COG1893	Ketopantoate reductase	MSM0033
2	COG1962	Tetrahydromethanopterin S-methyltransferase, subunit H	MSM0627, MSM1007
1	COG1985	Pyrimidine reductase, riboflavin biosynthesis	MSM0065
1	COG2038	NaMN:DMB phosphoribosyltransferase	MSM1200
1	COG2073	Cobalamin biosynthesis protein CbiG	MSM1267
1	COG2082	Precorrin isomerase	MSM1234
1	COG2099	Precorrin-6x reductase	MSM0896
1	COG2104	Sulfur transfer protein involved in thiamine biosynthesis	MSM0552
1	COG2145	Hydroxyethylthiazole kinase, sugar kinase family	MSM0916

3	COG2226	Methylase involved in ubiquinone/menaquinone biosynthesis	MSM1448, MSM1558, MSM1564
1	COG2241	Precorrin-6B methylase 1	MSM1167
1	COG2242	Precorrin-6B methylase 2	MSM0238
1	COG2243	Precorrin-2 methylase	MSM1351
1	COG2266	GTP:adenosylcobinamide-phosphate guanylyltransferase	MSM1005
1	COG2875	Precorrin-4 methylase	MSM0101
1	COG2896	Molybdenum cofactor biosynthesis enzyme	MSM1406
1	COG3161	4-hydroxybenzoate synthetase (chorismate lyase)	MSM0724
1	COG3252	Methenyltetrahydromethanopterin cyclohydrolase	MSM1723
2	COG4054	Methyl coenzyme M reductase, beta subunit	MSM0905, MSM1019
2	COG4055	Methyl coenzyme M reductase, subunit D	MSM0904, MSM1018
1	COG4056	Methyl coenzyme M reductase, subunit C	MSM1017
2	COG4057	Methyl coenzyme M reductase, gamma subunit	MSM0903, MSM1016
2	COG4058	Methyl coenzyme M reductase, alpha subunit	MSM0902, MSM1015
1	COG4059	Tetrahydromethanopterin S-methyltransferase, subunit E	MSM1014
1	COG4060	Tetrahydromethanopterin S-methyltransferase, subunit D	MSM1013
1	COG4061	Tetrahydromethanopterin S-methyltransferase, subunit C	MSM1012
1	COG4062	Tetrahydromethanopterin S-methyltransferase, subunit B	MSM1011
1	COG4063	Tetrahydromethanopterin S-methyltransferase, subunit A	MSM1010
1	COG4064	Tetrahydromethanopterin S-methyltransferase, subunit G	MSM1008
1	COG4218	Tetrahydromethanopterin S-methyltransferase, subunit F	MSM1009
Lipid Transport and Metabolism (I)			
1	COG0020	Undecaprenyl pyrophosphate synthase	MSM0096
1	COG0170	Dolichol kinase	MSM0078
1	COG0183	Acetyl-CoA acetyltransferase	MSM1562
1	COG0365	Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases	MSM0330
1	COG0439	Biotin carboxylase	MSM0765
2	COG0558	Phosphatidylglycerophosphate synthase	MSM0613, MSM1706
1	COG0575	CDP-diglyceride synthetase	MSM0850
1	COG1183	Phosphatidylserine synthase	MSM0982
2	COG1211	4-diphosphocytidyl-2-methyl-D-erithritol synthase	MSM0377, MSM1542
1	COG1250	3-hydroxyacyl-CoA dehydrogenase	MSM0965
1	COG1257	Hydroxymethylglutaryl-CoA reductase	MSM0227
1	COG1260	Myo-inositol-1-phosphate synthase	MSM0940
1	COG1267	Phosphatidylglycerophosphatase A and related proteins	MSM0934
1	COG1577	Mevalonate kinase	MSM1439
1	COG1924	Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)	MSM0810
1	COG2084		MSM0548
1	COG3425	3-hydroxy-3-methylglutaryl CoA synthase	MSM1561
Inorganic Ion Transport and Metabolism (P)			
1	COG0003	Oxyanion-translocating ATPase	MSM1170
1	COG0004	Ammonia permease	MSM0234
1	COG0038	Chloride channel protein EriC	MSM1721
1	COG0053	Predicted Co/Zn/Cd cation transporters	MSM0789
1	COG0168	Trk-type K+ transport systems, membrane components	MSM1095
1	COG0226	ABC-type phosphate transport system, periplasmic component	MSM0568
1	COG0288	Carbonic anhydrase	MSM1223
4	COG0310	ABC-type Co2+ transport system, permease component	MSM0583, MSM0584, MSM1488, MSM1618
1	COG0370	Fe2+ transport system protein B	MSM0589
1	COG0474	Cation transport ATPase	MSM0895
1	COG0475	Kef-type K+ transport systems, membrane components	MSM1186
1	COG0530	Ca2+/Na+ antiporter	MSM1027
1	COG0569	K+ transport systems, NAD-binding component	MSM1096
1	COG0573	ABC-type phosphate transport system, permease component	MSM0567
1	COG0581	ABC-type phosphate transport system, permease component	MSM0566
1	COG0600	ABC-type nitrate/sulfonate/bicarbonate transport system, permease component	MSM0291
1	COG0609	ABC-type Fe3+-siderophore transport system, permease component	MSM1394
1	COG0614	ABC-type Fe3+-hydroxamate transport system, periplasmic component	MSM1393
3	COG0619	ABC-type cobalt transport system, permease component CbiQ and related transporters	MSM0585, MSM0771, MSM1620
2	COG0704	Phosphate uptake regulator	MSM0564, MSM0569
1	COG0715	ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components	MSM1469
1	COG0725	ABC-type molybdate transport system, periplasmic component	MSM1609
1	COG0798	Arsenite efflux pump ACR3 and related permeases	MSM1078
1	COG0855	Polyphosphate kinase	MSM1424
1	COG1006	Multisubunit Na+/H+ antiporter, MnhC subunit	MSM1072
1	COG1116	ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	MSM0290
1	COG1117	ABC-type phosphate transport system, ATPase component	MSM0565
1	COG1118	ABC-type sulfate/molybdate transport systems, ATPase component	MSM1611
2	COG1122	ABC-type cobalt transport system, ATPase component	MSM0586, MSM1621
1	COG1230	Co/Zn/Cd efflux system component	MSM1639

1	COG1320	Multisubunit Na+/H+ antiporter, MnhG subunit	MSM1074
1	COG1348	Nitrogenase subunit NifH (ATPase)	MSM1707
1	COG1528	Ferritin-like protein	MSM1712
1	COG1563	Predicted subunit of the Multisubunit Na+/H+ antiporter	MSM1073
1	COG1824	Permease, similar to cation transporters	MSM1275
1	COG1863	Multisubunit Na+/H+ antiporter, MnhE subunit	MSM1076
1	COG1918	Fe2+ transport system protein A	MSM0588
1	COG1930	ABC-type cobalt transport system, periplasmic component	MSM1619
2	COG2111	Multisubunit Na+/H+ antiporter, MnhB subunit	MSM1068, MSM1069
1	COG2116	Formate/nitrite family of transporters	MSM1403
1	COG2212	Multisubunit Na+/H+ antiporter, MnhF subunit	MSM1075
4	COG2217	Cation transport ATPase	MSM0293, MSM0960, MSM1127, MSM1153
1	COG2608	Copper chaperone	MSM0961
1	COG3263	NhaP-type Na+/H+ and K+/H+ antiporters with a unique C-terminal domain	MSM0618
1	COG3420	Nitrous oxidase accessory protein	MSM1397
1	COG4149	ABC-type molybdate transport system, permease component	MSM1610
Secondary Metabolites Biosynthesis, Transport and Catabolism (Q)			
1	COG1228	Imidazolonepropionase and related amidohydrolases	MSM1154
General Function Prediction Only (R)			
2	COG0110	Acetyltransferase (isoleucine patch superfamily)	MSM0189, MSM1600
2	COG0312	Predicted Zn-dependent proteases and their inactivated homologs	MSM0866, MSM0947
1	COG0375	Zn finger protein HypA/HybF (possibly regulating hydrogenase expression)	MSM0108
1	COG0388	Predicted amidohydrolase	MSM0500
1	COG0433	Predicted ATPase	MSM0122
1	COG0446	Uncharacterized NAD(FAD)-dependent dehydrogenases	MSM0046
2	COG0456	Acetyltransferases	MSM0893, MSM1104
11	COG0457	FOG: TPR repeat	MSM0530, MSM0651, MSM0914, MSM1449, MSM1451, MSM1740, MSM1766, MSM1776, MSM1786, MSM1787, MSM1788
2	COG0491	Zn-dependent hydrolases, including glyoxylases	MSM0421, MSM1097
1	COG0496	Predicted acid phosphatase	MSM1218
2	COG0517	FOG: CBS domain	MSM0175, MSM1102
4	COG0535	Predicted Fe-S oxidoreductases	MSM0663, MSM0808, MSM1301, MSM1497
1	COG0561	Predicted hydrolases of the HAD superfamily	MSM0946
1	COG0595	Predicted hydrolase of the metallo-beta-lactamase superfamily	MSM1442
1	COG0603	Predicted PP-loop superfamily ATPase	MSM0936
1	COG0613	Predicted metal-dependent phosphoesterases (PHP family)	MSM1244
1	COG0622	Predicted phosphoesterase	MSM0507
1	COG0627	Predicted esterase	MSM0149
1	COG0628	Predicted permease	MSM1042
1	COG0641	Arylsulfatase regulator (Fe-S oxidoreductase)	MSM1606
5	COG0655	Multimeric flavodoxin WrbA	MSM0267, MSM0664, MSM0923, MSM1209, MSM1727
1	COG0661	Predicted unusual protein kinase	MSM0525
1	COG0663	Carbonic anhydrases/acetyltransferases, isoleucine patch superfamily	MSM0654
1	COG0666	FOG: Ankyrin repeat	MSM0266
1	COG0673	Predicted dehydrogenases and related proteins	MSM0882
1	COG0679	Predicted permeases	MSM1334
1	COG0714	MoxR-like ATPases	MSM0555
1	COG0730	Predicted permeases	MSM0420
3	COG0733	Na+-dependent transporters of the SNF family	MSM0699, MSM1531, MSM1532
1	COG0824	Predicted thioesterase	MSM0133
1	COG1011	Predicted hydrolase (HAD superfamily)	MSM1480
1	COG1019	Predicted nucleotidyltransferase	MSM0785
1	COG1078	HD superfamily phosphohydrolases	MSM0236
1	COG1084	Predicted GTPase	MSM0869
1	COG1094	Predicted RNA-binding protein (contains KH domains)	MSM0954
1	COG1099	Predicted metal-dependent hydrolases with the TIM-barrel fold	MSM0405
3	COG1123	ATPase components of various ABC-type transport systems, contain duplicated ATPase	MSM0770, MSM0971, MSM1698
2	COG1163	Predicted GTPase	MSM0714, MSM0715
1	COG1201	Lhr-like helicases	MSM0502
1	COG1202	Superfamily II helicase, archaea-specific	MSM1583
1	COG1203	Predicted helicases	MSM0166
1	COG1204	Superfamily II helicase	MSM0839
1	COG1205	Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster	MSM0112
5	COG1216	Predicted glycosyltransferases	MSM1321, MSM1329, MSM1330, MSM1503, MSM1507
1	COG1223	Predicted ATPase (AAA+ superfamily)	MSM0966
1	COG1234	Metal-dependent hydrolases of the beta-lactamase superfamily III	MSM0492
1	COG1235	Metal-dependent hydrolases of the beta-lactamase superfamily I	MSM1473
1	COG1244	Predicted Fe-S oxidoreductase	MSM0544

1	COG1245	Predicted ATPase, RNase L inhibitor (RLI) homolog	MSM0607
1	COG1253	Hemolysins and related proteins containing CBS domains	MSM1026
4	COG1266	Predicted metal-dependent membrane protease	MSM0292, MSM0803, MSM1148, MSM1180
1	COG1268	Uncharacterized conserved protein	MSM0429
2	COG1277	ABC-type transport system involved in multi-copper enzyme maturation, permease component	MSM0594, MSM0595
1	COG1287	Uncharacterized membrane protein, required for N-linked glycosylation	MSM0716
1	COG1310	Predicted metal-dependent protease of the PAD1/JAB1 superfamily	MSM0462
2	COG1323	Predicted nucleotidyltransferase	MSM0547, MSM0994
1	COG1326	Uncharacterized archaeal Zn-finger protein	MSM0846
2	COG1342	Predicted DNA-binding proteins	MSM0207, MSM0208
1	COG1350	Predicted alternative tryptophan synthase beta-subunit (paralog of TrpB)	MSM1242
1	COG1355	Predicted dioxygenase	MSM1438
1	COG1365	Predicted ATPase (PP-loop superfamily)	MSM0190
9	COG1373	Predicted ATPase (AAA+ superfamily)	MSM0061, MSM0280, MSM0680, MSM1197, MSM1278, MSM1527, MSM1789, MSM1790, MSM1795
1	COG1402	Uncharacterized protein, putative amidase	MSM0184
2	COG1408	Predicted phosphohydrolases	MSM0964, MSM1165
1	COG1409	Predicted phosphohydrolases	MSM0383
1	COG1411	Uncharacterized protein related to proFAR isomerase (HisA)	MSM1636
1	COG1412	Uncharacterized proteins of PilT N-term./Vapc superfamily	MSM0199
1	COG1418	Predicted HD superfamily hydrolase	MSM0632
1	COG1439	Predicted nucleic acid-binding protein, consists of a PIN domain and a Zn-ribbon module	MSM0816
4	COG1453	Predicted oxidoreductases of the aldo/keto reductase family	MSM0148, MSM0728, MSM1450, MSM1608
1	COG1489	DNA-binding protein, stimulates sugar fermentation	MSM1090
1	COG1537	Predicted RNA-binding proteins	MSM0640
1	COG1545	Predicted nucleic-acid-binding protein containing a Zn-ribbon	MSM1279
2	COG1571	Predicted DNA-binding protein containing a Zn-ribbon domain	MSM0452, MSM1295
1	COG1606	ATP-utilizing enzymes of the PP-loop superfamily	MSM0482
1	COG1608	Predicted archaeal kinase	MSM1440
1	COG1611	Predicted Rossmann fold nucleotide-binding protein	MSM0004
1	COG1634	Uncharacterized Rossmann fold enzyme	MSM0672
1	COG1646	Predicted phosphate-binding enzymes, TIM-barrel fold	MSM0124
2	COG1672	Predicted ATPase (AAA+ superfamily)	MSM1196, MSM1646
1	COG1691	NCAIR mutase (PurE)-related proteins	MSM1105
1	COG1707	ACT domain-containing protein	MSM1060
1	COG1759	ATP-utilizing enzymes of ATP-grasp superfamily (probably carboligases)	MSM0506
1	COG1779	C4-type Zn-finger protein	MSM0409
1	COG1782	Predicted metal-dependent RNase, consists of a metallo-beta-lactamase domain and an RNA-binding KH domain	MSM1038
1	COG1821	Predicted ATP-utilizing enzyme (ATP-grasp superfamily)	MSM0852
1	COG1829	Predicted archaeal kinase (sugar kinase superfamily)	MSM0060
1	COG1855	ATPase (PilT family)	MSM1183
1	COG1878	Predicted metal-dependent hydrolase	MSM0827
1	COG1907	Predicted archaeal sugar kinases	MSM0848
1	COG1942	Uncharacterized protein, 4-oxalocrotonate tautomerase homolog	MSM0688
1	COG1964	Predicted Fe-S oxidoreductases	MSM0849
1	COG1988	Predicted membrane-bound metal-dependent hydrolases	MSM1079
1	COG1994	Zn-dependent proteases	MSM0479
2	COG2005	N-terminal domain of molybdenum-binding protein	MSM0131, MSM1207
1	COG2047	Uncharacterized protein (ATP-grasp superfamily)	MSM1131
1	COG2054	Uncharacterized archaeal kinase related to aspartokinases, uridylate kinases	MSM0604
1	COG2068	Uncharacterized MobA-related protein	MSM0116
1	COG2079	Uncharacterized protein involved in propionate catabolism	MSM0449
1	COG2081	Predicted flavoproteins	MSM1235
1	COG2085	Predicted dinucleotide-binding enzymes	MSM0049
1	COG2102	Predicted ATPases of PP-loop superfamily	MSM0142
1	COG2118	DNA-binding protein	MSM0708
1	COG2129	Predicted phosphoesterases, related to the Icc protein	MSM0792
1	COG2150	Predicted regulator of amino acid metabolism, contains ACT domain	MSM0635
1	COG2151	Predicted metal-sulfur cluster biosynthetic enzyme	MSM0634
1	COG2220	Predicted Zn-dependent hydrolases of the beta-lactamase fold	MSM0779
1	COG2232	Predicted ATP-dependent carboligase related to biotin carboxylase	MSM0431
3	COG2244	Membrane protein involved in the export of O-antigen and teichoic acid	MSM1208, MSM1559, MSM1560
1	COG2252	Permeases	MSM1736
1	COG2403	Predicted GTPase	MSM0091
1	COG2405	Predicted nucleic acid-binding protein, contains PIN domain	MSM1530
1	COG2517	Predicted RNA-binding protein containing a C-terminal EMAP domain	MSM0466
2	COG2520	Predicted methyltransferase	MSM0802, MSM1036
1	COG2522	Predicted transcriptional regulator	MSM0269
3	COG3291	FOG: PKD repeat	MSM0281, MSM1716, MSM1735

1	COG3442	Predicted glutamine amidotransferase	MSM1138
1	COG3552	Protein containing von Willebrand factor type A (vWA) domain	MSM0554
1	COG3608	Predicted deacylase	MSM1080
1	COG3894	Uncharacterized metal-binding protein	MSM0517
1	COG3942	Surface antigen	MSM0921
1	COG3943	Virulence protein	MSM1645
1	COG4002	Predicted phosphotransacetylase	MSM0095
1	COG4015	Predicted dinucleotide-utilizing enzyme of the ThiF/HesA family	MSM0577
1	COG4026	Uncharacterized protein containing TOPRIM domain, potential nuclease	MSM1703
2	COG4032	Predicted thiamine-pyrophosphate-binding protein	MSM0080, MSM0081
1	COG4052	Uncharacterized protein related to methyl coenzyme M reductase subunit C	MSM1021
1	COG4076	Predicted RNA methylase	MSM0363
1	COG4085	Predicted RNA-binding protein, contains TRAM domain	MSM0647
1	COG4087	Soluble P-type ATPase	MSM1252
1	COG4277	Predicted DNA-binding protein with the Helix-hairpin-helix motif	MSM1239
2	COG4747	ACT domain-containing protein	MSM0388, MSM1713
1	COG4801	Predicted acyltransferase	MSM1385
1	COG4827	Predicted transporter	MSM1717
1	COG5012	Predicted cobalamin binding protein	MSM0516
3	COG5271	AAA ATPase containing von Willebrand factor type A (vWA) domain	MSM0993, MSM1240, MSM1454
1	COG5362	Phage-related terminase	MSM1671
1	COG5518	Bacteriophage capsid portal protein	MSM1672
2	COG5643	Protein containing a metal-binding domain shared with formylmethanofuran dehydrogenase subunit E	MSM1489, MSM1491
Function Unknown (S)			
1	COG0011	Uncharacterized conserved protein	MSM1029
2	COG0028		MSM0686, MSM1225
1	COG0059		MSM1222
1	COG0111		MSM0457
1	COG0147		MSM1146
1	COG0248		MSM1423
2	COG0318		MSM0025, MSM0374
1	COG0327	Uncharacterized conserved protein	MSM0576
1	COG0378		MSM0107
1	COG0391	Uncharacterized conserved protein	MSM0974
1	COG0392	Predicted integral membrane protein	MSM1094
2	COG0393	Uncharacterized conserved protein	MSM0418, MSM0456
1	COG0432	Uncharacterized conserved protein	MSM0279
1	COG0444		MSM0303
1	COG0451		MSM0327
2	COG0458		MSM0361, MSM0488
1	COG0462		MSM1577
2	COG0473		MSM0373, MSM1298
2	COG0477		MSM0772, MSM1210
2	COG0500		MSM0028, MSM1510
1	COG0505		MSM0489
1	COG0512		MSM1145
1	COG0513		MSM1498
1	COG0543		MSM1043
1	COG0585	Uncharacterized conserved protein	MSM1156
1	COG0591		MSM0386
1	COG0599	Uncharacterized homolog of gamma-carboxymuconolactone decarboxylase subunit	MSM0296
1	COG0601		MSM0301
2	COG0615		MSM0859, MSM1514
1	COG1028		MSM1731
2	COG1061		MSM0690, MSM0695
1	COG1063		MSM0376
1	COG1086		MSM1535
1	COG1120		MSM1395
1	COG1124		MSM0304
2	COG1134		MSM1326, MSM1592
1	COG1173		MSM0302
1	COG1199		MSM1352
1	COG1208		MSM0655
1	COG1243		MSM0842
1	COG1255	Uncharacterized protein conserved in archaea	MSM0894
2	COG1300	Uncharacterized membrane protein	MSM0215, MSM1526
1	COG1303	Uncharacterized protein conserved in archaea	MSM0932
1	COG1339		MSM1257
1	COG1359	Uncharacterized conserved protein	MSM1378
1	COG1371	Uncharacterized conserved protein	MSM0668
1	COG1379	Uncharacterized conserved protein	MSM1129
1	COG1387		MSM0063
1	COG1415	Uncharacterized conserved protein	MSM0931

1	COG1422	Predicted membrane protein	MSM0736
1	COG1430	Uncharacterized conserved protein	MSM1339
1	COG1460	Uncharacterized protein conserved in archaea	MSM1376
1	COG1469	Uncharacterized conserved protein	MSM1033
2	COG1474		MSM0671, MSM1264
1	COG1478	Uncharacterized conserved protein	MSM0975
1	COG1511	Predicted membrane protein	MSM0093
2	COG1520	FOG: WD40-like repeat	MSM1247, MSM1567
1	COG1548		MSM0851
1	COG1578	Uncharacterized conserved protein	MSM0551
1	COG1602	Uncharacterized conserved protein	MSM0346
2	COG1617	Uncharacterized conserved protein	MSM0348, MSM0349
1	COG1627	Uncharacterized protein conserved in archaea	MSM0983
1	COG1630	Uncharacterized protein conserved in archaea	MSM0123
1	COG1641	Uncharacterized conserved protein	MSM0935
1	COG1665	Uncharacterized protein conserved in archaea	MSM1058
1	COG1679	Uncharacterized conserved protein	MSM1192
1	COG1685		MSM0835
1	COG1690	Uncharacterized conserved protein	MSM0666
1	COG1693	Uncharacterized protein conserved in archaea	MSM1417
1	COG1698	Uncharacterized protein conserved in archaea	MSM1268
1	COG1701	Uncharacterized protein conserved in archaea	MSM0140
2	COG1704	Uncharacterized conserved protein	MSM0660, MSM1422
1	COG1710	Uncharacterized protein conserved in archaea	MSM0069
1	COG1711	Uncharacterized protein conserved in archaea	MSM1136
1	COG1714	Predicted membrane protein/domain	MSM1493
1	COG1718		MSM0952
1	COG1720	Uncharacterized conserved protein	MSM0132
2	COG1738	Uncharacterized conserved protein	MSM0646, MSM1382
1	COG1739	Uncharacterized conserved protein	MSM0186
1	COG1751	Uncharacterized conserved protein	MSM0628
1	COG1771	Uncharacterized protein conserved in archaea	MSM0070
1	COG1784	Predicted membrane protein	MSM0599
1	COG1786	Uncharacterized conserved protein	MSM1155
1	COG1795	Uncharacterized conserved protein	MSM1213
1	COG1809	Uncharacterized conserved protein	MSM0086
1	COG1817	Uncharacterized protein conserved in archaea	MSM0106
2	COG1822	Predicted archaeal membrane protein	MSM0581, MSM1216
1	COG1836	Predicted membrane protein	MSM0659
1	COG1844	Uncharacterized protein conserved in archaea	MSM0356
1	COG1849	Uncharacterized protein conserved in archaea	MSM0614
2	COG1852	Uncharacterized conserved protein	MSM0225, MSM0649
1	COG1860	Uncharacterized protein conserved in archaea	MSM0285
1	COG1865	Uncharacterized conserved protein	MSM0825
1	COG1872	Uncharacterized conserved protein	MSM1603
4	COG1873	Uncharacterized conserved protein	MSM0465, MSM0822, MSM0841, MSM1004
1	COG1891	Uncharacterized protein conserved in archaea	MSM1628
1	COG1909	Uncharacterized protein conserved in archaea	MSM0195
1	COG1915	Uncharacterized conserved protein	MSM0875
1	COG1916	Uncharacterized homolog of PrgY (pheromone shutdown protein)	MSM1024
1	COG1917	Uncharacterized conserved protein, contains double-stranded beta-helix domain	MSM1447
1	COG1920	Uncharacterized conserved protein	MSM0288
1	COG1937	Uncharacterized protein conserved in bacteria	MSM0959
1	COG1944	Uncharacterized conserved protein	MSM0480
1	COG1945	Uncharacterized conserved protein	MSM0878
1	COG1950	Predicted membrane protein	MSM1166
1	COG1971	Predicted membrane protein	MSM0030
1	COG1990	Uncharacterized conserved protein	MSM0605
1	COG1991	Uncharacterized conserved protein	MSM0145
1	COG2029	Uncharacterized conserved protein	MSM1057
1	COG2035	Predicted membrane protein	MSM1582
1	COG2042	Uncharacterized conserved protein	MSM0126
1	COG2043	Uncharacterized protein conserved in archaea	MSM0115
1	COG2078	Uncharacterized conserved protein	MSM0867
1	COG2090	Uncharacterized protein conserved in archaea	MSM1591
1	COG2098	Uncharacterized protein conserved in archaea	MSM0985
1	COG2106	Uncharacterized conserved protein	MSM0763
1	COG2122	Uncharacterized conserved protein	MSM0088
1	COG2136		MSM1632
2	COG2138	Uncharacterized conserved protein	MSM1280, MSM1281
1	COG2246	Predicted membrane protein	MSM1289
2	COG2314	Predicted membrane protein	MSM0109, MSM1739
2	COG2364	Predicted membrane protein	MSM0673, MSM0676
1	COG2429	Uncharacterized conserved protein	MSM0973

1	COG2450	Uncharacterized conserved protein	MSM0406
1	COG2456	Uncharacterized conserved protein	MSM1624
1	COG2457	Uncharacterized conserved protein	MSM0873
1	COG2892	Uncharacterized protein conserved in archaea	MSM1633
1	COG3273	Uncharacterized conserved protein	MSM1274
2	COG3274	Uncharacterized protein conserved in bacteria	MSM1370, MSM1556
1	COG3356	Predicted membrane protein	MSM0776
1	COG3367	Uncharacterized conserved protein	MSM0407
1	COG3482	Uncharacterized conserved protein	MSM0481
1	COG3543	Uncharacterized conserved protein	MSM0430
3	COG3548	Predicted integral membrane protein	MSM0468, MSM0469, MSM1205
1	COG3586	Uncharacterized conserved protein	MSM1741
1	COG3815	Predicted membrane protein	MSM1770
1	COG3874	Uncharacterized conserved protein	MSM0683
1	COG3976	Uncharacterized protein conserved in bacteria	MSM1637
1	COG4009	Uncharacterized protein conserved in archaea	MSM0794
1	COG4010	Uncharacterized protein conserved in archaea	MSM0793
1	COG4012	Uncharacterized protein conserved in archaea	MSM1243
1	COG4014	Uncharacterized protein conserved in archaea	MSM0840
1	COG4016	Uncharacterized protein conserved in archaea	MSM0578
1	COG4017	Uncharacterized protein conserved in archaea	MSM0575
1	COG4018	Uncharacterized protein conserved in archaea	MSM0571
1	COG4019	Uncharacterized protein conserved in archaea	MSM0574
1	COG4020	Uncharacterized protein conserved in archaea	MSM1221
1	COG4021	Uncharacterized conserved protein	MSM0463
1	COG4022	Uncharacterized protein conserved in archaea	MSM0643
1	COG4029	Uncharacterized protein conserved in archaea	MSM0812
1	COG4030	Uncharacterized protein conserved in archaea	MSM0309
1	COG4033	Uncharacterized protein conserved in archaea	MSM0103
1	COG4035	Predicted membrane protein	MSM0315
1	COG4036	Predicted membrane protein	MSM0320
1	COG4037	Predicted membrane protein	MSM0321
1	COG4038	Predicted membrane protein	MSM0322
1	COG4039	Predicted membrane protein	MSM0323
1	COG4040	Predicted membrane protein	MSM0324
1	COG4041	Predicted membrane protein	MSM0325
1	COG4042	Predicted membrane protein	MSM0326
2	COG4050	Uncharacterized protein conserved in archaea	MSM0811, MSM1130
1	COG4051	Uncharacterized protein conserved in archaea	MSM0809
1	COG4053	Uncharacterized protein conserved in archaea	MSM0229
1	COG4065	Uncharacterized protein conserved in archaea	MSM1006
2	COG4066	Uncharacterized protein conserved in archaea	MSM0064, MSM0367
1	COG4068	Uncharacterized protein containing a Zn-ribbon	MSM0417
1	COG4069	Uncharacterized protein conserved in archaea	MSM0815
1	COG4071	Uncharacterized protein conserved in archaea	MSM0630
1	COG4073	Uncharacterized protein conserved in archaea	MSM0726
1	COG4077	Uncharacterized protein conserved in archaea	MSM1034
1	COG4078	Predicted membrane protein	MSM0319
1	COG4079	Uncharacterized protein conserved in archaea	MSM1472
1	COG4081	Uncharacterized protein conserved in archaea	MSM0104
1	COG4084	Uncharacterized protein conserved in archaea	MSM0314
1	COG4121	Uncharacterized conserved protein	MSM1555
1	COG4289	Uncharacterized protein conserved in bacteria	MSM1302
1	COG4635		MSM1262
3	COG4713	Predicted membrane protein	MSM0521, MSM1291, MSM1444
2	COG4744	Uncharacterized conserved protein	MSM1402, MSM1719
1	COG4883	Uncharacterized protein conserved in archaea	MSM1086
1	COG4907	Predicted membrane protein	MSM1421
1	COG5015	Uncharacterized conserved protein	MSM0863
1	COG5305	Predicted membrane protein	MSM1288
1	COG5423	Predicted metal-binding protein	MSM0050
1	COG5440	Uncharacterized conserved protein	MSM1265
4	COG5464	Uncharacterized conserved protein	MSM0067, MSM0681, MSM1765, MSM1785

Table 10. Glycosyltransferases (GT) in *M. smithii* and *M. stadtmanae* proteomes classified according to Carbohydrate Active enZyme (CAZy) database

	CAZy GT family	Protein	Annotation
M. smithii	**GT1**	MSM0423*	glycosyltransferase (modular protein with two domains distantly related to glycosyltransferases), GT2/GT1 families [CAZy]
	GT2	MSM0423*	glycosyltransferase (modular protein with two domains distantly related to glycosyltransferases), GT2/GT1 families [CAZy]
		MSM1290	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1294	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1297	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1310	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1311	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1312	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1316	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1321	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1323	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1324	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1328	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1329	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1330	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1503	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1507	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1545	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1594	glycosyltransferase (modular protein with two N-terminal beta-glycosyltransferase-related domains and C-terminal glycerophosphotransferase-related domain), GT2 families [CAZy]
		MSM1602	glycosyltransferase (modular protein with N-terminal beta-glycosyltransferase-related domain and C-terminal glycerophosphotransferase-related domain), GT2 family [CAZy]
		MSM1623	glycosyltransferase (related to beta-glycosyltransferases), GT2 family [CAZy]
		MSM1627	glycosyltransferase (related to bactoprenol beta-glucosyltransferase), GT2 family [CAZy]
	GT4	MSM0836	glycosyltransferase related to alpha-glycosyltransferases), GT4 family [CAZy]
		MSM1313	glycosyltransferase (distantly related to glycosyltransferases), GT4 family [CAZy]
		MSM1317	glycosyltransferase (distantly related to glycosyltransferases), GT4 family [CAZy]
		MSM1322	glycosyltransferase related to alpha-glycosyltransferases), GT4 family [CAZy]
	GT66	MSM0716	glycosyltransferase (distantly related to oligosaccharyltransferases), STT3 subunit, GT66 family [CAZy]
M. stadtmanae	**GT1**	Msp_0515	partially conserved hypothetical protein
		Msp_0645	predicted glycosyltransferase
	GT2	Msp_0042**	predicted glycosyltransferase
		Msp_0045	predicted glycosyltransferase
		Msp_0054	predicted glycosyltransferase
		Msp_0203	predicted glycosyltransferase
		Msp_0206	predicted glycosyltransferase
		Msp_0207	predicted glycosyltransferase
		Msp_0212	predicted glycosyltransferase
		Msp_0215	predicted glycosyltransferase
		Msp_0218	predicted glycosyltransferase
		Msp_0220	predicted glycosyltransferase
		Msp_0441	predicted glycosyltransferase
		Msp_0442	predicted glycosyltransferase
		Msp_0492	predicted glycosyltransferase
		Msp_0493	predicted glycosyltransferase
		Msp_0495	predicted glycosyltransferase
		Msp_0496	predicted glycosyltransferase
		Msp_0500	predicted glycosyltransferase
		Msp_0538	predicted glycosyltransferase
		Msp_0541	predicted glycosyltransferase
		Msp_0645	predicted glycosyltransferase
		Msp_0989	predicted glycosyltransferase
		Msp_1087	predicted glycosyltransferase
		Msp_1481	conserved hypothetical membrane-spanning protein
		Msp_1540	partially conserved hypothetical protein
	GT4	Msp_0039	predicted glycosyltransferase
		Msp_0044	predicted glycosyltransferase
		Msp_0049	predicted glycosyltransferase
		Msp_0051	predicted glycosyltransferase
		Msp_0052	predicted glycosyltransferase
		Msp_0053	predicted glycosyltransferase
		Msp_0055	predicted glycosyltransferase
		Msp_0056	predicted glycosyltransferase
		Msp_0057	predicted glycosyltransferase
		Msp_0101	predicted glycosyltransferase
		Msp_0492	predicted glycosyltransferase
		Msp_0991	predicted glycosyltransferase
	GT66	Msp_0368	conserved hypothetical membrane-spanning protein

---

* - modular protein

** - probable fragment

Table 12. InterPro-based classification of adhesin-like proteins (ALPs) in the *M. smithii* and *M. stadtmanae* proteomes

				Glycosylation Sites[1]		InterPro Domain Content[2]									
						Adhesin-like				Repeats			Activity		
	GENE	ANNOTATION	Number of Amino Acids in Protein	N-linked	O-linked	Intimin	Bacterial Ig-like	Pectate Lyase	GAG/ Chondroitinase	Chlamydia PMP Repeats	Parallel Beta-Helix Repeats	Other Repeats	Peptidase/ Protein Binding	Glycosidase/ Sugar Binding	Secreted Glucosidase
M. smithii	MSM0031	adhesin-like protein	354	2	1										
	MSM0051	adhesin-like protein	1022	10		■		■			■				
	MSM0052	adhesin-like protein	2530	13	2					■					■
	MSM0057	adhesin-like protein	251	8	23										
	MSM0092	putative adhesin-like protein	203		2	■									
	MSM0159	adhesin-like protein	153	2	1										
	MSM0173	adhesin-like protein	2879	35	1	■		■	■			BA	Pep	■	
	MSM0221	adhesin-like protein	336	5	2	■							tPep		
	MSM0266	adhesin-like protein	2065	10		■		■	■		■				■
	MSM0281	putative adhesin-like protein	592	1		■		■	■		■				
	MSM0282	adhesin-like protein	1049	10		■		■	■		■				
	MSM0337	putative adhesin-like protein	71												
	MSM0411	adhesin-like protein	1414	10				■				BA			
	MSM0412	putative adhesin-like protein	376	2											
	MSM0461	adhesin-like protein	1026	7	1	■		■	■		■				
	MSM0580	putative adhesin-like protein	262	4											
	MSM0616	adhesin-like protein	3684		2	■		■	■		■				■
	MSM0884	adhesin-like protein	755		11										
	MSM0885	adhesin-like protein	1730		1	■		■					CPD		
	MSM0957	adhesin-like protein	620			■							CPD		
	MSM0995	adhesin-like protein	951			■	■	■				F			
	MSM0996	adhesin-like protein	796					■			■				
	MSM1111	adhesin-like protein	1941		3	■		■		■	■	F	Pep		
	MSM1112	adhesin-like protein	1909			■		■		■	■			Gal	
	MSM1113	adhesin-like protein	2101		1	■	■	■						Ogly	
	MSM1114	adhesin-like protein	1879		2			■	■		■				
	MSM1116	adhesin-like protein	1491			■	■	■			■		CPD		
	MSM1168	putative adhesin-like protein	1055		1										
	MSM1188	adhesin-like protein	4691		2	■		■		■	■			■	
	MSM1282	putative adhesin-like protein	532	10	5	■							CPD		
	MSM1305	adhesin-like protein	2036	29		■		■		■					
	MSM1306	adhesin-like protein	1430	11	1	■	■	■		■		PQ			
	MSM1397	adhesin-like protein	1370	8		■		■	■		■				
	MSM1398	adhesin-like protein	767	3		■		■			■		CPD		
	MSM1399	adhesin-like protein	649	4	6			■	■		■				
	MSM1485	putative adhesin-like protein	128			■									
	MSM1533	adhesin-like protein	1152	15								BA			
	MSM1534	adhesin-like protein	2709	29	3							BA	Col		
	MSM1554	putative adhesin-like protein	612	7	2	■									
	MSM1567	adhesin-like protein	2193	13		■	■			■		PQ			■
	MSM1585	adhesin-like protein	1262	3						■			tPep		
	MSM1586	adhesin-like protein	1831	14				■		■	■				
	MSM1587	adhesin-like protein	1702	15		■		■		■	■				
	MSM1590	adhesin-like protein	1884	20		■		■		■	■				■
	MSM1709	adhesin-like protein	748	15	1	■		■		■					
	MSM1716	adhesin-like protein	1007	7		■	■								
	MSM1735	adhesin-like protein	1434	12		■		■		■			Pep		
	MSM1738	putative adhesin-like protein	156			■									

M. stadtmanae	Msp_0043	member of asn/thr-rich large protein family	816	9												
	Msp_0046	member of asn/thr-rich large protein family	881	2	6										tGlyc	
	Msp_0048	member of asn/thr-rich large protein family	1909	19	1											
	Msp_0141	member of asn/thr-rich large protein family	2120	13										CPD		
	Msp_0145	member of asn/thr-rich large protein family	2774	17	1									Pep		
	Msp_0154	member of asn/thr-rich large protein family	2036	28	1									CPD		
	Msp_0432	member of asn/thr-rich large protein family	2044	23												
	Msp_0440	member of asn/thr-rich large protein family	1552	10	1									CPD		
	Msp_0590	member of asn/thr-rich large protein family	2979	30	3								F	tPep		
	Msp_0606	member of asn/thr-rich large protein family	2459	20	2									tPep		
	Msp_0619	member of asn/thr-rich large protein family	2468	20	1											
	Msp_0713	member of asn/thr-rich large protein family	2151	20	1											
	Msp_0747	member of asn/thr-rich large protein family	706	7												
	Msp_0762	member of asn/thr-rich large protein family	3356	25	2								F	tPep		
	Msp_0765	member of asn/thr-rich large protein family	1222	8	1											
	Msp_0776	member of asn/thr-rich large protein family	421	3												
	Msp_0777	member of asn/thr-rich large protein family	1875	19	1									Pep		
	Msp_0836	member of asn/thr-rich large protein family	1790	25												
	Msp_0911	member of asn/thr-rich large protein family	1965	24												
	Msp_0912	member of asn/thr-rich large protein family	2037	18												
	Msp_0913	member of asn/thr-rich large protein family	1519	27	3											
	Msp_0922	member of asn/thr-rich large protein family	973	7												
	Msp_0963	member of asn/thr-rich large protein family	1104	12										CPD		
	Msp_0964	member of asn/thr-rich large protein family	769	5												
	Msp_0980	member of asn/thr-rich large protein family	1063	12									BA			
	Msp_0983	member of asn/thr-rich large protein family	1311	18										Pep		
	Msp_1091	member of asn/thr-rich large protein family	735	11	2											
	Msp_1103	member of asn/thr-rich large protein family	857	11												
	Msp_1104	member of asn/thr-rich large protein family	1140	11												
	Msp_1108	member of asn/thr-rich large protein family	1921	18	1											
	Msp_1109	member of asn/thr-rich large protein family	853	13	1											
	Msp_1115	member of asn/thr-rich large protein family	1896	16	1											
	Msp_1348	member of asn/thr-rich large protein family	834	3											tGlyc	
	Msp_1399	member of asn/thr-rich large protein family	1447	21	3											
	Msp_1401	member of asn/thr-rich large protein family	1324	18												
	Msp_1465	member of asn/thr-rich large protein family	2214	18										CPD		
	Msp_1572	member of asn/thr-rich large protein family	843											tPep		

KEY						
BA - *B. anthrasis* OMP repeat			**CPD** - Carboxypeptidase D regulatory domain			
F - filamin repeat			**tGlyc** - transglycosidase		**GAG** - glycosaminoglycan	
PQ - PQQ cofactor binding repeat			**tPep** - trans-peptidase		**Pep** - peptidase	
Col - collagen			**Gal** - galactose		**Ogly** - O-glycan	
>20	>10	>0	- degree of glycosylation		- present	- absent

[1]Predictions completed using NetNGlyc and NetOGlyc (http://www.cbs.dtu.dk/services/).

[2]InterPro domains: Invasin/intimin cell-adhesion (IPR008964); Bacterial Ig-like (IPR003344); pectin lyase fold (IPR011050); GAG lyase, Chondroitinase B-type (IPR012333); Polymorphic membrane protein, Chlamydia (IPR003368); Parallel beta-helix repeat (IPR006626); Peptidase S8 and S53 (IPR000209); Penicillin-binding protein, transpeptidase fold (IPR012338); Carboxypeptidase regulatory region (IPR008969)

Table 13. *M. smithii* GeneChip

	Naming Prefix	Genes Represented	Probesets	Probe pairs	Average number of probe pairs per probeset
control sequences	AFFX	64	64	1024	16
protein coding genes	MSM	1778	2018	19967	11
tRNA genes (1-2 probesets/gene)	MSM-tRNAxx	34	74	450	11
rRNA genes[1]	MSMxx-rRNA	8	7	77	11
intergenic sequences	ig	--	1581	4931	3

[1]Note that the *M. smithii* genome contains three 5S rRNA genes, one 7S rRNA gene, two 16S rRNA genes, and two 23S rRNA genes. Due to the high nucleotide sequence identity among rRNA genes of a given type, each is represented by a single probeset (the 16S rRNA probeset is replicated four times on the GeneChip)

Table 15. BLAST analyses of the putative *M. smithii* prophage

***M. smithii* Protein**	**Phage Protein Sequence ID***	**Function**	**HMM Annotation**	**Phage HMM**	**E value**
MSM1640	5417	unknown	Phage_integrase: Phage integrase family	PF00589	2.30E-06
MSM1654	5721	Gp40	ERF: ERF superfamily	PF04404	6.90E-11
MSM1671	5397	large terminase subunit	psiM2_ORF9: phage uncharacterized protein, C-terminal domain	TIGR01630	0.0042
MSM1672	5398	portal protein	portal_PBSX: phage portal protein, PBSX family	TIGR01540	6.70E-12
MSM1675	6246	putative structural protein			
MSM1677	6247	putative structural protein			
MSM1684	20206	ORF001	TMP: TMP repeat	PF05017	0.0036
MSM1691	6262	PeiW			

* - from the Phage Sequence Databank

Chapter 3

The pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii* in twins

Chapter 3

Classification: Microbiology

The pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii* in twins

Elizabeth E. Hansen[1], Catherine A. Lozupone[2], Federico E. Rey[1], Janaki Guruge[1], Aneesha Narra[1], Jonathan Goodfellow[1], Meng Wu[1], Andrew C. Heath[3], Rob Knight[2,4], and Jeffrey I. Gordon[1]

[1]Center for Genome Sciences and Systems Biology, and [3]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO 63108

[2]Department of Chemistry and Biochemistry and [4]HHMI, University of Colorado, Boulder, CO 80309

Key words: human gut microbiota; metagenomics; hydrogen-consuming microbes; co-occurrence analyses; microbial genome evolution, microbial RNA-Seq, microbial metabolism

Author contributions: EEH, FER and JIG designed research; EEH, FER, JG, and JM, and performed research; EEH, CAL, JF, MW and AN contributed new reagents/analytic tools, EEH, CAL, FER, MW, RK, and JIG analyzed data; EEH, CAL and JIG wrote the paper.

Abstract

The human gut microbiota harbors three main groups of H_2-consuming microbes: methanogens including the dominant archaeon, *Methanobrevibacter smithii*, a polyphyletic group of acetogens, and sulfate-reducing bacteria. Defining their roles in the gut is important for understanding how hydrogen metabolism affects the efficiency of fermentation of dietary components. We quantified methanogens in fecal samples from 40 healthy adult female monozygotic(MZ) and 28 dizygotic(DZ) twin pairs, analyzed bacterial 16S rRNA datasets generated from their fecal samples to identify taxa that co-occur with methanogens, sequenced the genomes of 20 *M. smithii* strains isolated from families of MZ and DZ twins, and performed RNA-Seq of a subset of strains to identify their responses to varied formate concentrations. The concordance rate for methanogen carriage was significantly higher for MZ versus DZ twin pairs. Co-occurrence analysis revealed 22 bacterial species-level taxa positively correlated with methanogens: all but two were members of the Clostridiales, with several being, or related to, known hydrogen-producing and -consuming bacteria. The *M. smithii* pan-genome contains 987 genes conserved in all strains, and 1,860 variably represented genes. Strains from MZ and DZ twin pairs had a similar degree of shared genes and SNPs, and were significantly more similar than strains isolated from mothers or members of other families. The 101 adhesin-like proteins(ALPs) in the pan-genome (45 ± 6 per strain) exhibit strain-specific differences in expression and responsiveness to formate. We hypothesize that *M. smithii* strains use their different repertoires of ALPs to create diversity in their metabolic niches, by allowing them to establish syntrophic relationships with bacterial partners with differing metabolic capabilities and patterns of co-occurrence.

Introduction

Human microbiome projects seek to determine how microbial communities are assembled, maintained, and operate within our various body habitats as a function of our different cultural and socioeconomic conditions, family structures, stages of life, genotypes and physiologies. Culture-independent metagenomic surveys have revealed that microbial communities cluster according to body habitat but with considerable interpersonal variation in bacterial species content (1), although differences are smaller within rather than between families (2). The gut harbors our largest collection of microbes, spanning all three domains of life. Bacteria dominate, specifically members of the phyla Bacteroidetes and Firmicutes (2-5).

Monozygotic (MZ) and dizygotic (DZ) twin pairs provide an attractive study paradigm for dissecting the relative contributions of host genotype and environmental exposures to shaping the microbial and viral landscape of our gut microbiota (2, 6). To date, bacterial 16S rRNA datasets indicate that adult MZ co-twins share no more similarity in their fecal bacterial communities than DZ co-twins, suggesting that shared environmental exposures likely play key roles in determining gut microbial community composition (2).

Here, we extend these twin studies to examine *Methanobrevibacter smithii*, the dominant archaeon in the human gut microbiota (3). We address several general questions. *First,* to what extent is the representation of Archaea influenced by host genotype versus shared environmental exposures? *Second,* what bacterial species, if any, co-occur with this hydrogen-consuming methanogen? *Third*, how does the genome of *M. smithii* vary within a co-twin, between co-twins, and across families?

Answers to these questions may have therapeutic implications. Accumulation of hydrogen from microbial fermentation inhibits bacterial NADH dehydrogenases, reducing the yield of ATP obtained by primary fermenters, and the short chain fatty acid end-products of fermentation that can be absorbed by the host (7, 8). Thus, manipulation of the

abundance of hydrogen consumers or their metabolic activities could affect the efficiency of energy harvest by the host. The human gut contains three major groups of organisms capable of consuming hydrogen that could be targeted: methanogenic archaea, acetogenic bacteria, and sulfate-reducing bacteria (SRB). *M. smithii* is an attractive therapeutic target not only because of its prominence among human gut-associated Archaea, but because it has a lower H_2-utilization threshold than acetogens and, thus, is likely to be more efficient at depleting H_2 from the gut environment. Acetogens are a metabolically diverse group of microbes that are distributed among a number of bacterial phyla; many consume H_2 or formate and CO_2 to generate acetate and ATP via the Wood–Ljungdahl pathway. Sulfate reducers can also use H_2 as an electron donor to generate hydrogen sulfide (H_2S) through anaerobic sulfate respiration. *Desulfovibrio piger*, a member of the δ-Proteobacteria, appears to be the dominant SRB present in the human gut microbiota (9). To consider the guild of H_2 consumers as a therapeutic target, we must first identify which of their genetic and metabolic features are conserved within and between individual hosts and are therefore critical for fitness in the human gut. Unfortunately, in the case of *M. smithii,* tractable genetic systems are not yet available for genome-wide screens of fitness determinants.

Thirty to fifty percent of humans are positive for methanogens in the gut as measured either by breath tests (10,11) or PCR assays targeting *mcrA* (methyl coenzyme M reductase subunit A, conserved in the methanogenesis pathway) (12). However, whether and how host physiologic factors, such as gastrointestinal transit time (13,14), and BMI (15,16), influence interpersonal variation in the representation of methanogens remains unclear. Methane excretion phenotyping of 274 Australian families containing adolescent twin pairs indicated that environmental exposures play a deterministic role in methanogen carriage, with similar concordance and correlations between MZ and DZ co-twins, and less concordance between parents and their offspring (17). As in other studies (18), more females than males had positive methane breath tests. Studies of intergenerational transfer of a positive methane excretion phenotype in rats demonstrated the critical effects of en-

vironmental factors during the weaning period; however, colonization through adulthood varied between strains of rats (17) suggesting that host genetic factors affect carriage.

In this report, we have characterized the representation of *M. smithii* in adult rather than adolescent female MZ and DZ twin pairs living in the United States, identified bacteria that co-occur with this methanogen, compared and contrasted the genomes of 20 sequenced *M. smithii* isolates recovered from the frozen fecal microbiota of MZ and DZ co-twins, and used RNA-Seq to perform whole genome transcriptional profiling of a subset of sequenced isolates under different growth conditions. The results provide an expanded view of the diversity and adaptations of this dominant archaeon to life in the human gut.

Results and Discussion

MZ Twins Have Higher Concordance *for* Gut Methanogens *than DZ* Twins.

We used a quantitative PCR (qPCR) assay of the *mcrA* gene to measure methanogens present in single fecal samples collected from 40 female MZ and 28 adult female DZ twin pairs (age 21–31 y). All were born in Missouri, although at the time they provided samples, only 29% were living in the same home and some lived >800 km apart (2). Based on a health questionnaire, all were healthy and none had a history of gastrointestinal disease including irritable bowel syndrome. Sixty-one percent were obese (BMI ≥ 30) and 7% overweight (BMI 25–30) at the time of sampling (2).

Thirty-two of the 136 individuals (23%) had levels of methanogens above our threshold for confidently calling the fecal sample "positive" (i.e., ≥4 x 10^7 genome equivalents per mg of total fecal DNA), and this proportion did not vary significantly by zygosity group ($P = 0.59$). The MZ twin pair concordance rate for carriage of methanogens was 74%, a value significantly higher than the DZ pair concordance rate (15%; $P = 0.009$ by Breslow-Day test). In addition, there was a significantly higher degree of correlation of methanogen levels between MZ pairs by linear regression ($r^2 = 0.43$, $P < 0.0001$) than DZ

pairs ($r^2 = 0.04$, p=0.32), (**Figure 1 A and B**). Fecal samples were also collected from 23 of the MZ twin pairs and 12 of the DZ pairs 2 mo after the initial time point. Linear regression showed that time point 1 and time point 2 samples were highly correlated for both the presence of methanogens ($r^2 = 0.54$, $P < 0.0001$; **Figure 1C**) and their levels. Neither carriage nor levels of methanogens was significantly correlated with being overweight or obese in this study population ($P = 0.37$ and 0.38, respectively).

Thirteen samples from the initial timepoint representing 4 MZ twin pairs, 1 DZ twin pair, plus 3 other unrelated individuals that were positive for *mcrA* were chosen for sequencing of amplicons generated by using the *mcrA* primers and previously described archaeal 16S rRNA primers (n = 5-10 amplicon subclones/primer set/fecal DNA sample). In 12 of the 13 samples *M. smithii* was the only sequence detected by *mcrA* or 16S rRNA-directed PCR. In one MZ co-twin (TS17 in Dataset S1, Table S1), 2 of 6 16S rRNA amplicons and 2 of 8 *mcrA* amplicons matched to *Methanosphaera stadtmanae*, a mesophilic euryarchaeota known to be present in the gut microbiota of some humans (19); the remaining amplicons generated from her fecal DNA matched to *M. smithii*. Her co-twin (TS16) had no detectable methanogens.

We also examined fecal samples from 51 mothers in this study for presence of methanogens and found a similar overall degree of methanogen carriage in this population as found in their daughters (31% and 25%, respectively). Concordance for carriage of methanogens between mother and daughter (i.e., the probability that the daughter of a methanogen carrier was also a carrier, 32%) was nonsignificant ($P = 0.33$).

Co-Occurrence Between *M. smithii* and Bacterial Taxa

Our qPCR results suggest that host genetic factors, including factors that influence the representation of potential syntrophic partners, may play a role in carriage of methanogens. In contrast, the study of Florin *et al*. (17), which used methane breath tests, showed no

significant differences in concordance between young adolescent Australian MZ and DZ twin pairs. The difference could be explained if environmental factors play a dominant role in determining whether methanogens are acquired early in life, whereas persistent carriage in later life is determined by a variety of host factors. Such factors range from human genotype to the presence or absence of bacterial taxa that can collaborate or compete with the methanogens.

A role for host factors in determining carriage of methanogens is supported by previous studies of nonhuman primates. Methanogens were present in the gut microbiota of some primate phylogenetic lineages but not others; however, these patterns did not follow any identifiable features of gut physiology or morphology, nor behavior or diet (20). Another study that examined the distribution of methanogens within the guts of 253 vertebrate species found "methanogenic branches" of the host phylogenetic tree [i.e. branches containing ruminants (bovidae, cervidae, giraffidae) and "nonmethanogenic" branches (felidae, canidae, and ursidae)]. As with the primate study, the methane-producing groups could not be distinguished from the methane-negative groups based on their diets or features of their gut structure/physiology (21).

To understand whether methanogen carriage might be determined, in part, by the presence or absence of bacterial taxa that can collaborate or compete with the methanogens, we investigated co-occurrence patterns between methanogens and sulfate-reducing bacteria (SRB). SRB, which can use H_2 as an electron donor to generate hydrogen sulfide (H_2S) through anaerobic sulfate respiration, may show positive associations with methanogens if a hydrogen economy is more important in some individuals than others, or negative associations due to competition for H_2. Positive associations between SRB and methanogens might also occur because of syntrophy, because some methanogens and SRB can grow syntrophically on lactate, with the methanogen removing H_2 generated by the SRB (22; 23). Therefore, we determined whether SRB and methanogens had nonrandom codistribution patterns by SRB-directed qPCR assays of 87 fecal samples from the MZ and DZ

twin pairs. The *aps* gene encodes adenosine-*5'*-phosphosulfate reductase, a key enzyme that catalyzes activation and then reduction of sulfate to sulfite (24). We chose *aps* as a target for a qPCR assay that used previously described and validated primers (25). Forty-five percent of the samples were positive for SRB, above our threshold of detection defined as $\approx 4 \times 10^7$ genome equivalents per mg of fecal DNA. The concordance rate for sulfate reducers was not significant for either MZ or DZ co-twins (31% and 27%, Dataset S1, Table S1). A logistic regression was performed to determine whether a higher level of *mcrA* is predictive of the presence of *aps*, or vice versa. No statistically significant relationship was identified in either comparison ($P = 0.10$ and 0.07).

We also performed a general search for bacterial Operational Taxonomic Units (OTUs) that had positive or negative associations with *M. smithii*, using sequences generated from multiplex pyrosequencing of the V2 variable region of bacterial 16S rRNA genes from these same fecal samples (2). The raw sequences from this prior study were now processed using the PyroNoise algorithm to remove sequencing noise (26) as implemented in QIIME (27). Using UCLUST (28), the denoised sequences were further divided into OTUs that each shared ≥96% nucleotide sequence identity (a value slightly more permissive than the 97% ID threshold typically used to denote a microbial species). The most abundant sequence within each of the resulting 12,833 OTUs was then selected as a representative of that OTU. Because some of the individuals in the study were sampled multiple times, we randomly selected one sample per individual. For each of the 607 OTUs that were found in at least 10 of the samples for which we had *mcrA* qPCR data, an ANOVA was performed to determine whether the OTU relative abundance was significantly different in methanogen-positive and -negative individuals. We also checked for associated presence/absence patterns by using the G-test of independence (an OTU was scored as present if it was observed one or more times). The resulting p-values were corrected for multiple comparisons by using the Bonferroni correction (multiplied by 607; the number of comparisons) and the false discovery rate (FDR) method (multiplied by the number of comparisons divided by the *P* value rank).

Twenty-two OTUs had significantly different relative abundances in *mcrA*-positive versus negative individuals ($P < 0.05$ using ANOVA with the FDR correction). Of these 22 OTUs, 21 were more abundant in samples where methanogens were present, whereas one OTU was less abundant. The G-test-identified five significant OTUs ($P < 0.05$ with FDR correction), and 4 of these 5 were also significant as judged by ANOVA. All G-test identified associations were positive. Thus, the two statistical tests together identified 22 positively associated OTUs (Dataset S1, **Table S2**) and one negatively associated OTU.

To investigate the phylogenetic relationships of these OTUs to each other, and to bacterial isolates and lineages with known biological properties, we used parsimony insertion to add a representative sequence for each significant OTU into the Greengenes coreset tree (29) in the Arb software package (30). Because the closest relatives of the OTUs were mostly from other culture-independent metagenomic studies, we also inserted 16S rRNA sequences into the tree that were from well-characterized bacteria, including 16S rRNAs from fully sequenced genomes deposited in KEGG or sequenced through the Human Gut Microbiome Initiative (HGMI; http://genome.wustl.edu/genomes/list/human_gut_microbiome/), and 16S rRNA sequences from related organisms with known properties that were identified by using BLAST searches against the National Center for Biotechnology Information nonredundant database. To look for evidence of whether relatives of the OTUs were capable of growing in pure culture, we also BLASTed the 16S rRNA sequences against sequences in the RDP (31) that were marked as being from cultured bacterial isolates.

Remarkably, 20 of the 22 positively associated OTUs were members of the class Clostridiales (Firmicutes phylum). These 20 OTUs binned into five broad groups that were scattered throughout the class, including members of the three main clusters found in the human gut (clusters I, IV, and XIVa).

The group most positively associated with *M. smithii* was a lineage within Clostridia cluster IV that contains members of the genera Oscillospira and Sporobacter (Sporobacter,

(Dataset S1, **Table S2**; note that this group had the four most significant OTUs according to the ANOVA test). Two of these OTUs are highly related to *Oscillospira guilliermondii,* an as yet uncultured, large and morphologically conspicuous organism found in ruminants (32,33). The most closely related cultured isolate that we could find for any of these OTUs is *Sporobacter termitidis*, a hydrogen-consuming acetogen from the termite gut (34).

Two of the positively associated OTUs are members of Clostridia cluster XIVa. The closest isolate with a sequenced genome was *Blautia hydrogenotrophica*, a hydrogen-consuming homoacetogen from the human gut, although the percent identity across the lanemasked V2 region was low (89–93%) and more closely related organisms to *B. hydrogenotrophica* are known not to be acetogens. Whether the Sporobacter and *B. hydrogenotrophica*-related OTUs are acetogens cannot be determined by using 16S rRNA sequences alone, because acetogenesis is only inconsistently associated with 16S rRNA-defined phylotypes (35). However, the relationship suggests that some OTUs may co-occur with methanogens because they are homoacetogens and have a shared preference for hydrogen. Nonetheless, the OTU most related to *B. hydrogenotrophica* in this analysis (99%ID) did not show significant co-occurrence with *M. smithii* (uncorrected *P*-value = 0.38), indicating that not all homoacetogens in the human co-occur with *M. smithii* because of this preference for hydrogen.

Because members of the SRB can produce and consume H_2, we were specifically interested in OTUs in the dataset that were in this group. Eighty-two of 281 fecal samples (29%) from the 16S rRNA analysis of these same twin pairs (including additional fecal samples for which we did not obtain *mcrA* data) (2) had OTUs that were within the SRB clade (Figure S1*B*). The actual prevalence of SRB is likely higher, because the samples were not exhaustively sequenced. Phylogenetic comparison indicated that these OTUs represented *Desulfovibrio piger* in 41 (14.6%) of the samples, *Desulfovibrio desulfuricans* in 10 samples (3.6%), and an additional taxon (1908) in 38 samples (13.5%) that was only distantly related to cultured isolates (Dataset S1, Table S2 and Figure S1). Although sig-

nificant associations were not detected with the SRB-specific qPCR, OTU 1908 showed a significantly positive association with methanogens (Dataset S1, Table S2). The abundant OTU representing *D. piger* (OTU 12050), did not have statistically significant co-occurrence with methanogens (**Figure S1**), and the three different types of SRB did not significantly co-occur with each other. The differing distribution patterns of the three different SRB species, coupled with the smaller number of fecal samples for which we had *aps* compared with *mcrA* qPCR data, likely contributed to our inability to detect a significant association between methanogens and SRB with the *aps* qPCR assay.

The concentration of H_2 in the gut lumen can vary over a wide range in healthy individuals (from 0.17 to 49% in a study of 11 subjects; 36). Levels of H_2 in the distal gut reflect the dynamic interplay between microbial production and consumption. One of the co-occurring groups within the Clostridiales may produce abundant amounts of hydrogen. Specifically, two of the positively associated OTUs in the Clostridiales family mapped to a clade that included isolate Rennanqilyf3, which was recovered from activated sludge by using a procedure designed to retrieve bacteria with particularly high yields of hydrogen (37). This isolate performs ethanol-type fermentation with glucose as an optimal carbon source for hydrogen production; however, its hydrogen production capacity varies with hydrogen concentration and pH. Thus, methanogen (*M. smithii*) abundance may be in part regulated by the presence of bacterial lineages that are efficient hydrogen producers. To our knowledge, no cultured isolates are available for members of this lineage from the gut.

Some of the OTUs that are positively associated with methanogens are quite distant from any cultured relatives (ribotypes): This observation is intriguing, because it suggests that syntrophic relationships may inhibit them from growing in monoculture. For example, four OTUs grouped in a clade of the Clostridiales family that is dominated by relatives identified in culture-independent studies of cellulose-degrading gut environments where methanogens also reside (e.g., termite gut and cow rumen) (Gut Clone Group; Dataset S1, **Table S2** and **Figure S1A**). The closest organism with a sequenced genome was only very

distantly related, with a 78–86%ID over the lanemasked V2 region of rRNA. A BLAST search against the cultured component of the RDP revealed one successful attempt to culture a relative of one of these four OTUs (95%ID) from the forestomach of the kangaroo (38). However, this cultured isolate was much more distant from the other three co-occurring OTUs in this clade, and there are no reported cultured relatives for any of these four OTUs from the human gut. Three co-occurring OTUs fell within the Catabacter lineage. The closest cultured isolate, *Catabacter* sp. YIT12065, is only 82–92% identical to these co-occurring OTUs; very little is known about this isolate's biology. The presence of obligate syntrophs for methanogens in the human gut would not be surprising, because they are known to exist in other environments, such as sludge (39,40).

Unfortunately, the lack of cultured relatives for these OTUs limits our ability to more fully interpret the co-occurrence results, because we lack knowledge about their biological properties. Targeted attempts to culture gut bacteria in the presence of *M. smithii* as well as targeted attempts to obtain and sequence their genomes from mixed populations should help to elucidate their functional relationships with human gut methanogens.

Analysis of the Pan-Genome of *M. smithii*.

We reasoned that one approach for further characterizing factors that affect *M. smithii* colonization of the human gut would be to develop a method for isolating strains from frozen fecal samples obtained from twins and their mothers, sequencing their genomes, and performing RNA-Seq to evaluate strain-level variations in patterns of gene expression during growth under varying levels of hydrogen and formate.

The method we developed for recovering *M. smithii* from frozen fecal samples is described in *SI Methods*. A total of 20 strains were isolated from two families: one consisting of a MZ twin pair and their mother and the other a DZ twin pair and their mother (n = 2–5 strains isolated and sequenced per individual). Deep draft genome assemblies were generated by using reads produced by Illumina GA-IIx and 454 sequencers. Dataset

S1, **Table S3** describes the details of genome coverage and of the assembly statistics. Assembled genomes were aligned by using Mauve (41), which iteratively reordered contigs based on the finished genome sequence of the *M. smithii* type strain PS (42). Dataset S1, **Table S3** also provides information about deep draft assemblies previously generated of the genomes of two other *M. smithii* type strains obtained from culture collections (42).

On average, any two strains shared 92.96 ± 6.5% of their single nucleotide polymorphisms (SNPs) [129,112 ± 6,322 (mean ± SD)]. A binary table of the presence or absence of a SNP was subsequently generated, a distance matrix was calculated, and a principal components analysis (PCA) was performed (Figure S2*A* and *C*). The PCA showed that strains from the same individual and strains from within co-twins clustered together. Both MZ and DZ co-twins shared significantly more SNPs in their strains, than with strains from their mothers or unrelated individuals (Fig S2*B*).

Genes were identified by using Glimmer (v3.02) trained on contigs >500 bp in each of the 20 sequenced *M. smithii* isolate genomes, plus the PS type strain and the two other *M. smithii* isolates we had previously sequenced. Genes in all 23 genomes were binned by using the program CD-HIT and its default parameters (>90% nucleotide sequence identity over of the length of the shorter gene in each pairwise comparison; Figure S3) into "operational gene units" (OGUs), a term we use in a way that is analogous to OTUs. If any predicted gene from an assembled genome was present in a given OGU bin, that OGU was called "present" within that genome (43). Functions were assigned to predicted proteins encoded by each gene by using the KEGG and STRING databases; Pfam and TIGRFAM annotations were also made. Note that all predicted protein-coding sequences <300 nt were filtered out and not considered in the analyses reported below.

Rarefaction analysis to determine the rate at which sequencing the genes of the new strains revealed new OGUs showed that the number of new or unique OGUs identified begins to plateau by the time ≈6 strains were sequenced (≈10,000 genes) (**Figure S4 A** and **B**). A total of 987 OGUs were present in all 23 strains (34.7% of 2,847 identified OGUs),

whereas 1,532 (53.8%) were found in more than one strain but not all, and 328 (11.5%) in only a single strain (**Figure S3A** and **B**).

PCA of OGU assignments showed clustering of strains based on family of origin: Strains from MZ family members (TS94-96) generally clustered together, whereas strains from the DZ family (TS145-147) split into two groups (**Figure S3C**). Further pairwise comparisons of the degree of sharing of OGUs in strains showed that strains within an individual and within MZ and DZ co-twins shared significantly more OGUs than strains from the co-twin's mother or from unrelated individuals. Moreover, the degree of sharing of OGUs was not significantly different between MZ and DZ twin pairs (Figure S3*D*). As noted above, MZ twins have greater concordance for carriage and levels of methanogens in their fecal microbiota than DZ twins. The fact that the sequenced strains are no more similar between MZ co-twins than DZ twins suggests that although shared environmental exposures to methanogens direct which strains are found in an individual's gut, long-term persistence is influenced by a combination of host and microbial genetic factors.

We used KEGG to assign enzyme commission (EC) numbers to genes in all of the isolates' genomes. A total of 412 ECs were identified: 349 were shared by all strains, 63 were variably represented and 18 had significant differences in their representation between strains as judged by binomial test (**Figure S5D**). These discriminatory ECs include (i) several restriction enzymes, (ii) two peptidases [a serine protease known as Do, HtrA or DegP (44) that may protect against heat-stress and unfolded proteins, and endopeptidase La (45)], both of which may be related to quality-control in protein folding, and (iii) tRNA-guanine transglycosylases (involved in the anti-codon modification of tRNAs specific for Asn, Asp, His and Tyr) (**Figure S5B–D**).

Genes assigned to COG M (cell envelope biogenesis/outer membrane) were prominently represented in the variable component of the pan-genome (**Figure S5A**). Variability in surface proteins may directly impact the fitness of *M. smithii* strains *in vivo*, including their ability to adhere to host structures, or to interact with syntrophic partners. For exam-

ple, all of the *M. smithii* strains contain the six genes involved in synthesis of pseudaminic acid structures related to sialic acid molecules expressed on host cell surfaces. The resulting surface epitopes are thought to play a role in the adaptation of *M. smithii*'s to the gut environment by mimicking the sialic acids that decorate the surfaces of host epithelial cells (46). Adhesin-like proteins (ALPs) are a novel class of proteins with homology to bacterial adhesins that were first identified in the *M. smithii* type strain. They are also hypothesized to play a role in adaptation to the gut environment (42). The 23 sequenced strains contain a total of 101 ALP OGUs (average 45 ± 6 ALP genes per strain): Only six were present in all strains. ALP sequences are quite divergent in terms of their domain structure: e.g., many have intimin domains, which in *Escherichia coli* mediate binding to intestinal epithelial cells; others have pectate lyase domains and/or parallel β-helix repeats that are often found in enzymes with polysaccharide substrates.

To better understand genomic differences among *M. smithii* strains, we searched the *M. smithii* pan-genome for evidence of horizontal gene transfer (HGT) (see *SI Methods* for details of compositional- and phylogenetic-based analyses of HGT). The results, described in *SI Results* and summarized in Dataset S1, **Table S4C** show that HGT has contributed to both the core and variable elements of the *M. smithii* pan-genome. They include core genes involved in methanogenesis and folate biosynthesis; e.g., both compositional- and phylogenetic-based methods revealed transfer of genes encoding THMP methyltransferase C subunit (EC 2.1.1.86), formate dehydrogenase (EC 1.2.1.2), and formylmethanofuran dehydrogenase subunit F (E.C. 1.2.99.5) (Dataset S1, **Table S4B)**. Note that the early steps in synthesis of methanopterin, a C1 carrier coenzyme involved in the methanogenesis pathway (Figure S6), are the same as those used for generation of folate (Dataset S1, **Table S4B**). In addition, between 52% and 65% ALPs show evidence of transfer: Large-scale HGT of ALPs would be consistent with their variability among strains (Dataset S1, **Table S6A**).

Expression Profiling of *M. smithii* Strains by RNA-Seq.

We used RNA-Seq to profile the transcriptomes of five of the *M. smithii* isolates: One from each member of the MZ family, one from each of the DZ co-twins, plus the PS type strain. The five strains from the two families were chosen because SNP, OGU, and EC analyses indicated that these isolates were representative of the strains from their human hosts, and because they exhibited consistent patterns of growth on MBC medium containing 2.8 or 44.1 mM formate, a substrate for the first enzyme involved in the methanogenesis pathway, formate dehydrogenase (EC 1.2.1.2 in **Figure S6B**). Triplicate cultures were grown to midlog phase in defined medium with either low or high formate concentrations under an atmosphere that contained 80% hydrogen. Total RNA was extracted, structural RNAs were depleted (*SI Methods*), and double-stranded cDNA was synthesized and sequenced with an Illumina GA-IIx instrument (36-nt reads; 3–4 million reads per sample, with each biological triplicate sequenced twice as technical replicates). Reads were normalized to reads per kilobase per million (RPKM) and mapped back to each strain's own reference genome. At midlog phase, the number of protein-coding genes with ≥10 mapped mRNA-derived reads varied from 1,594-1,782 (89–97% of all CDS) among the 5 strains (Dataset S1, **Table S5A**). When we compared the 987 OGUs that comprise the conserved core of the *M. smithii* pan-genome to 31 sequenced methanogens associated with the human gut (*M. stadmanae*), cow rumen (*M. ruminantium*) or various environmental habitats, 55 OGUs were identified as unique to *M. smithii* (Blastp threshold $E < 10^{-10}$), of which 42 encoded predicted conserved hypothetical or hypothetical proteins (Dataset S1, **Table S5C)**. At the depth of sequencing achieved, RNA-Seq indicated that 34 of these 42 hypothetical genes were expressed in mid-log phase in the PS type strain (Dataset S1, **Table S5C**).

We subsequently compared the phenotypes of strains based on normalized expression of each gene encoding each EC. Examining the gene expression data across functional groups allowed us to compare strains: The results revealed that no gene family was consis-

tently regulated by formate across all strains. To identify genes significantly regulated by formate in each strain, we first analyzed normalized reads with CyberT. We used two criteria for determining significance in regulation: a posterior probability of differential expression (PPDE) threshold ≥0.97 and a ≥2-fold difference in expression (either direction) when a given strain was incubated in low versus high levels of formate (Dataset S1, **Table S7**).

All of the genes in the methanogenesis pathway illustrated in Figure S6*B* were expressed in all six strains. Nonetheless, several of the genes in this pathway exhibited strain-specific differences in their levels of expression including EC 1.5.99.9 (F420-dependent methylene tetrahydromethanopterin dehydrogenase) and EC 1.5.99.11 (5,10-methylenetetrahydromethanopterin reductase). Cobalt, an important cofactor for some of the enzymes in the methanogenesis pathway, is translocated by an ABC transporter: Components of the transporter exhibited formate-responsive behavior in the PS type strain and in the strain from one of the DZ co-twins (TS145) but not in the strains from her sister or mother (Dataset S1, **Table S7**).

Looking beyond the methanogenesis pathway, none of the genes encoding ECs in the *M. smithii* pan-genome satisfied our criteria for being responsive to differences in formate levels in the medium at midlog phase in all strains. However, as with components of the methanogenesis pathway, some exhibited strain-specific differences in formate sensitivity e.g., in strain METSMITS145B (from DZ co-twin 1) genes encoding the subunits of MtrH (EC 2.1.1.86; tetrahydromethanopterin S-methyltransferase) were up-regulated in high formate, whereas in strain METSMITS146E (from the sister of DZ co-twin 1) they were down-regulated (see Dataset S1, **Table S7** for additional examples).

M. smithii uses ammonia as a nitrogen source via an energy-dependent glutamine synthetase-glutamate synthase pathway, which has high affinity for ammonia, and a ATP-independent pathway with lower affinity (Figure 2*A*). Both pathways are expressed in all strains, with 0.4–1.21% of reads mapping to enzymes involved in assimilation of ammo-

nia. The energy-dependent GlnA pathway is generally expressed at a much higher level than the low affinity pathway, although strain-specific differences in levels expression were noted. With few exceptions, such as the genes encoding EC 1.4.1.4 and EC 1.4.1.13 in strains METSMITS145B and METSMITS96A, components of both pathways failed to exhibit a significant difference in their levels of expression in any of the strains as a function of formate concentration. Another exception was the ammonium transporter (AmtB) (Figure 2 *B* and *C* and Dataset S1, **Table S7**).

Using our threshold criteria for formate-responsive expression, four of the six strains were defined as having genes that were sensitive to levels of this compound. Dataset S1, **Table S7** lists the 9 genes present in type strain PS, the 340 genes in the strain recovered from the mother of the DZ co-twins (TS145), the 23 genes in the strain isolated from one of her daughters (TS146), and the 81 genes in the strain from the mother of the MZ twins (TS96). Intriguingly, no genes were identified in strains from MZ twins of this mother (TS94, TS95) that exhibited significant formate responsiveness. The core component of *M. smithii*'s pan-genome contained no genes that met our criteria for formate-responsive behavior in every isolate.

The utility of using formate to identify strain-specific phenotypes is best illustrated by ALPs. As noted above, each sequenced strain contained a distinctive repertoire of genes encoding ALPs, with only 6 ALP OGUs shared by all isolates. ALP OGUs 112, 208, 412, and 827 are encoded by genes present in 4–6 of the strains: None of the genes are formate-responsive but members of each OGU exhibit strain-specific differences in their levels of expression (levels of expression are also notably different between ALP OGUs). OGUs 18, 37, 133, and 226 show strain-specific differences in their representation, strain-specific differences in their levels of expression, plus within-OGU differences in their formate sensitivity (**Figure 3**).

Prospectus

These results lead us to hypothesize that *M. smithii* strains use their different repertoires of ALPs and the different sensitivities of ALP genes to formate to create diversity in their physical locations and/or their metabolic niches within the gut. Stated another way, these variations in expressed ALP repertoires could have important effects on the ability of different strains to establish syntrophic relationships with bacterial partners that have different abilities to generate formate or other substrates, or that have differing patterns of co-occurrence within an individual over time and between individuals. To further explore this notion, it will be important to define the structures of representative members of different ALP clusters through an *M. smithii*-directed structural genomics effort: Selection of ALPs could be guided by a number of criteria, including their strain distribution and their patterns of expression, both *in vitro* in monoculture in the presence of a variety of potential substrates for their metabolic networks, and *in vivo* in gnotobiotic mice containing various collections of sequenced *M. smithii* isolates and available cultured co-occurring bacterial taxa. The interactions between isolates and co-occurring bacterial species can also be explored *in vitro* if cocolonization of gnotobiotic mice proves to be problematic either because of difficulty in identifying suitable host diets or strains that are fit in the mouse gut (e.g., we have not yet been able to achieve persistent colonization of gnotobiotic mice with any of the five strains characterized *in vitro* by RNA-Seq after inoculating all of them together with a consortium of human gut-derived members of the Firmicutes, Bacteroidetes and Proteobacteria that include saccharolytic bacteria, and hydrogen producers and consumers). A complementary approach will be to select taxa for these *in vitro* and *in vivo* studies by predicting potential syntrophic relationships through *in silico* metabolic reconstructions of the metabolic networks of sequenced co-occurring species and *M. smithii* isolates, using methods described by Borenstein *et al*. (47).

Methods

Genome Sequencing

M. smithii strains are isolated and grown by using the procedure detailed in *SI Methods*. Genomic DNA was sequenced with an Illumina Genome Analyzer IIx instrument (36 base read lengths; 3.5–29 million reads per strain), and a 454 pyrosequencer (Titanium chemistry; 27,844–449,545 reads per strain). Reads were assembled using Velvet (48) for Illumina reads, and the Newbler v2.3 (Roche) for 454 reads. Hybrid assemblies were generated by using AMOS minimus2 (http://sourceforge.net/apps/mediawiki/amos/index.php?title=Minimus2), producing draft assemblies with on average 55 contigs, N50 contig lengths of 103,633 nucleotides and total genome size of 1.9 Mb (Dataset S1, Table S3). Procedures used for genome annotation are described in *SI Methods*.

Other Methods

Details of the experimental and computational approaches used for qPCR, co-occurrence, comparative genomic, HGT (Figure S7), and microbial RNA-Seq analyses, plus how *M. smithii* RNA-Seq datasets compare with custom *M. smithii* GeneChip datasets are provided in *SI Methods* and Figs. S8 and S9. Analyses of familial concordance or correlation for methanogen carriage or levels, and of their associations with overweight/obesity, were conducted by using logistic or linear regression, a robust variance estimator to adjust for the nonindependence of observations on family members.

Acknowledgements

We thank Sabrina Wagoner and Jill Manchester for superb technical assistance, Daniela Puiu and Steven Salzberg for generous help with genome assembly software, Stacey Marion and Deborah Hopper for assistance in obtaining fecal samples, plus Nicholas Griffin, Jeremiah Faith, Nathan McNulty, Ansel Hsiao, and Alejandro Reyes for many very

helpful suggestions during the course of the study. This work was supported in part by National Institute of Health Grants DK78669, DK30292, DK70977, and AA09022. The Crohn's and Colitis Foundation of America, and Howard Hughes Medical Insitute. E.E.H. is a member of Washington University's Medical Scientist Training Program (NIH T32 GM07200-31).

Footnotes

The authors declare no conflict of interest.

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Microbes and Health" held November 2–3, 2009, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. The complete program and audio files of most presentations are available on the NAS Web site athttp://www.nasonline.org/SACKLER_Microbes_and_Health.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported for *M. smithii* strains in this paper have been deposited in the GenBank database [accession nos. AEKU00000000 (for strain TS145 A); AELL00000000(TS145B); AELM00000000 (TS146A); AELN00000000 (TS146B); AELO00000000 (TS146C);AELP00000000 (TS146D); AELQ00000000 (TS146E); AELR 00000000 (TS147A); AELS00000000(TS147B); AELT00000000 (TS147C); AELU0000 0000 (TS94A); AELV00000000 (TS94B);AELW00000000 (TS94C); AELX00000000 (T S95A); AELY00000000 (TS95B); AELZ00000000(TS95C); AEMA00000000 (TS95D); A EMB00000000 (TS96A); AEMC00000000 (TS96B); andAEMD0000000 (TS96C)]. RNA-Seq and GeneChip data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos.GSE25408 (RNA-Seq) and GSE25535 (GeneChip)].
This article contains supporting information online atwww.pnas.org/lookup/suppl/ doi:10.1073/pnas.1000071108/-/DCSupplemental.

References

1. Costello EK, et al. Bacterial community variation in human body habitats across space and time.*Science*. 2009;**326**:1694–1697.

2. Turnbaugh PJ, et al. A core gut microbiome in obese and lean twins. *Nature*. 2009;**457**:480–484.

3. Eckburg PB, et al. Diversity of the human intestinal microbial flora. *Science*. 2005;**308**:1635–1638.

4. Dethlefsen L, Huse S, Sogin ML, Relman DA. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol*. 2008;**6**:e280

5. Qin J, et al. MetaHIT Consortium A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;**464**:59–65.

6. Reyes A, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*.2010;**466**:334–338

7. Wolin MJ, Miller TL. Interactions of microbial populations in cellulose fermentation. *Fed Proc*.1983;**42**:109–113.

8. McNeil NI. The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr*.1984;**39**:338–342.

9. Scanlan PD, Shanahan F, Marchesi JR. Culture-independent analysis of desulfovibrios in the human distal colon of healthy, colorectal cancer and polypectomized individuals. *FEMS Microbiol Ecol*. 2009;**69**:213–221.

10. Bond JH, Jr, Engel RR, Levitt MD. Factors influencing pulmonary methane excretion in man. An indirect method of studying the in situ metabolism of the methane-producing colonic bacteria. *J Exp Med*. 1971;**133**:572–588.

11. Levitt MD, Furne JK, Kuskowski M, Ruddy J. Stability of human methanogenic flora over 35 years and a review of insights obtained from breath methane measurements. *Clin Gastroenterol Hepatol*.2006;**4**:123–129.

12. Scanlan PD, Shanahan F, Marchesi JR. Human methanogen diversity and incidence in healthy and diseased colonic groups using mcrA gene analysis. *BMC Microbiol*. 2008;**8**:79.

13. Attaluri A, Jackson M, Valestin J, Rao SSC. Methanogenic flora is associated with altered colonic transit but not stool characteristics in constipation without IBS. *Am J Gastroenterol*. 2010;**105**:1407–1411.

14. Pimentel M, et al. Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. *Am J Physiol Gastrointest Liver Physiol*.2006;**290**:G1089–G1095.

15. Armougom F, Henry M, Vialettes B, Raccah D, Raoult D. Monitoring bacterial community of human gut microbiota reveals an increase in Lactobacillus in obese patients and methanogens in anorexic patients. *PLoS ONE*. 2009;**4**:e7125.

16. Zhang H, et al. Human gut microbiota in obesity and after gastric bypass. *Proc Natl Acad Sci USA*.2009;**106**:2365–2370.

17. Florin TH, Zhu G, Kirk KM, Martin NG. Shared and unique environmental factors determine the ecology of methanogens in humans and rats. *Am J Gastroenterol*. 2000;**95**:2872–2879.

18. Pitt P, de Bruijn KM, Beeching MF, Goldberg E, Blendis LM. Studies on breath methane: The effect of ethnic origins and lactulose. *Gut*. 1980;**21**:951–954.

19. Fricke WF, et al. The genome sequence of Methanosphaera stadtmanae reveals why this human intestinal archaeon is restricted to methanol and H_2 for methane formation and ATP synthesis. *J Bacteriol*. 2006;**188**:642–658.

20. Hackstein JHP, Van Alen TA, Op Den Camp H, Smits A, Mariman E. Intestinal methanogenesis in primates—a genetic and evolutionary approach. *Dtsch Tierarztl Wochenschr.* 1995;**102**:152–154.

21. Hackstein JHP, et al. Fecal methanogens and vertebrate evolution. *Evolution*. 1996;**50**:559–572.

22. Scholten JC, Culley DE, Brockman FJ, Wu G, Zhang W. Evolution of the syntrophic interaction between Desulfovibrio vulgaris and Methanosarcina barkeri: Involvement of an ancient horizontal gene transfer. *Biochem Biophys Res Commun*. 2007;**352**:48–54.

23. Plugge CM, et al. Global transcriptomics analysis of the Desulfovibrio vulgaris change from syntrophic growth with Methanosarcina barkeri to sulfidogenic metabolism. *Microbiology*.2010;**156**:2746–2756.

24. Friedrich MW. Phylogenetic analysis reveals multiple lateral transfers of adenosine-5´-phosphosulfate reductase genes among sulfate-reducing microorganisms. *J Bacteriol*.2002;**184**:278–289.

25. Stewart JA, Chadwick VS, Murray A. Carriage, quantification, and predominance of methanogens and sulfate-reducing bacteria in faecal samples. *Lett Appl Microbiol*. 2006;**43**:58–63.

26. Quince C, et al. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods*. 2009;**6**:639–641.

27. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;**7**:335–336.

28. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*.2010;**26**:2460–2461.

29. DeSantis TZ, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol*. 2006;**72**:5069–5072.

30. Ludwig W, et al. ARB: a software environment for sequence data. *Nucleic Acids Res*.2004;**32**:1363–1371.

31. Cole JR, et al. The Ribosomal Database Project (RDP-II): Sequences and tools for high-throughput rRNA analysis. *Nucleic Acids Res*. 2005;**33**(Database issue):D294–D296.

32. Mackie RI, et al. Ecology of uncultivated Oscillospira species in the rumen of cattle, sheep, and reindeer as assessed by microscopy and molecular approaches. *Appl Environ Microbiol*.2003;**69**:6808–6815.

33. Yanagita K, et al. Flow cytometric sorting, phylogenetic analysis and in situ detection of Oscillospira guillermondii, a large, morphologically conspicuous but uncultured ruminal bacterium. *Int J Syst Evol Microbiol*. 2003;**53**:1609–1614.

34. Grech-Mora I, et al. Isolation and characterization of Sporobacter termitidis gen nov sp nov, from the digestive tract of the wood-feeding termite Nasutitermes lujae. *Int J Syst Bacteriol*. 1996;**46**:512–518.

35. Drake HL, Gössner AS, Daniel SL. Old acetogens, new light. *Ann N Y Acad Sci*. 2008;**1125**:100–128.

36. Levitt MD. Volume and composition of human intestinal gas determined by means of an intestinal washout technic. *N Engl J Med*. 1971;**284**:1394–1398.

37. Li YF, et al. Molecular characterization and hydrogen production of a new species of anaerobe. *J Environ Sci Health A Tox Hazard Subst Environ Eng*. 2005;**40**:1929–1938.

38. Ouwerkerk D, Klieve AV, Forster RJ, Templeton JM, Maguire AJ. Characterization of culturable anaerobic bacteria from the forestomach of an eastern grey kangaroo, Macropus giganteus. *Lett Appl Microbiol*. 2005;**41**:327–333.

39. Kosaka T, et al. The genome of Pelotomaculum thermopropionicum reveals niche-associated evolution in anaerobic microbiota. *Genome Res*. 2008;**18**:442–448.

40. McInerney MJ, et al. The genome of Syntrophus aciditrophicus: Life at the thermodynamic limit of microbial growth. *Proc Natl Acad Sci USA*. 2007;**104**:7600–7605.

41. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 2004;**14**:1394–1403.

42. Samuel BS, et al. Genomic and metabolic adaptations of Methanobrevibacter smithii to the human gut. *Proc Natl Acad Sci USA*. 2007;**104**:10643–10648.

43. Giannakis M, et al. Response of gastric epithelial progenitors to Helicobacter pylori Isolates obtained from Swedish patients with chronic atrophic gastritis. *J Biol Chem*. 2009;**284**:30383–30394.

44. Lipinska B, Zylicz M, Georgopoulos C. The HtrA (DegP) protein, essential for Escherichia coli survival at high temperatures, is an endopeptidase. *J Bacteriol*. 1990;**172**:1791–1797.

45. Lee I, Berdis AJ, Suzuki CK. Recent developments in the mechanistic enzymology of the ATP-dependent Lon protease from Escherichia coli: Highlights from kinetic studies. *Mol Biosyst*.2006;**2**:477–483.

46. Lewis AL, et al. Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. *Proc Natl Acad Sci USA*. 2009;**106**:13552–13557.

47. Borenstein E, Kupiec M, Feldman MW, Ruppin E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci USA*. 2008;**105**:14482–14487.

48. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs.*Genome Res*. 2008;**18**:821–829.

Figure Legends

Figure 1. Correlation of methanogen levels in the fecal microbiota of MZ and DZ co-twins. The presence and levels of fecal methanogens were defined by qPCR assay that targeted the *mcrA* gene in fecal samples obtained from (**A**) MZ twin pairs (n=40) and (**B**) DZ twin pairs (n=28). All axes are $\log_{10}$(genome equivalents/ng total DNA +1). Dashed lines represent 95% confidence intervals for linear regression. (**C**) Correlation between *mcrA* levels in fecal samples collected at two time points per individual (2-month interval between sampling).

Figure 2. Normalized RNA-Seq reads assigned to the gene encoding an ammonium transporter (AmtB) and ECs involved in ammonia assimilation. (**A**) Overview of *M. smithii*'s two pathways for assimilating ammonia: the energy-dependent glutamine synthetase-glutamate synthase pathway has high affinity for ammonia (red arrow); an ATP-independent pathway has lower affinity (orange). (**B**) Strain-specific differences in the relative expression of components of the high affinity Gln pathway and the energy-independent low affinity pathway for ammonia assimilation. Mean values ±S.E.M. are plotted. Colors represent components of the two pathways shown in panel A; color codes are coordinated between panels A and B. (**C**) Strain-specific differences in levels of expression of the ammonium transporter, AmtB. $p<0.0001$ by one-way ANOVA.

Figure 3. Differential expression of *M. smithii* adhesin-like proteins (ALPs). Members of selected ALP OGUs with (**A**) strain-specific differences in their expression profiles and (**B**) strain-specific, as well as OGU-associated, differences their sensitivity to levels of formate during mid-log phase growth. OGUs 112, 412, 827 and 208 exhibit strain-specific differences in their expression irrespective of formate concentration (one-way ANOVA, $p<0.0001$), while OGUs 226, 287, 18, 133 and 37 contain at least one representative that is significantly regulated by formate concentration. Mean values ±S.E.M. are plotted (n=6 replicates/condition). * indicates a ≥2-fold difference, PPDE≥0.97 (see **Table S7**).

Figures

Figure 1.

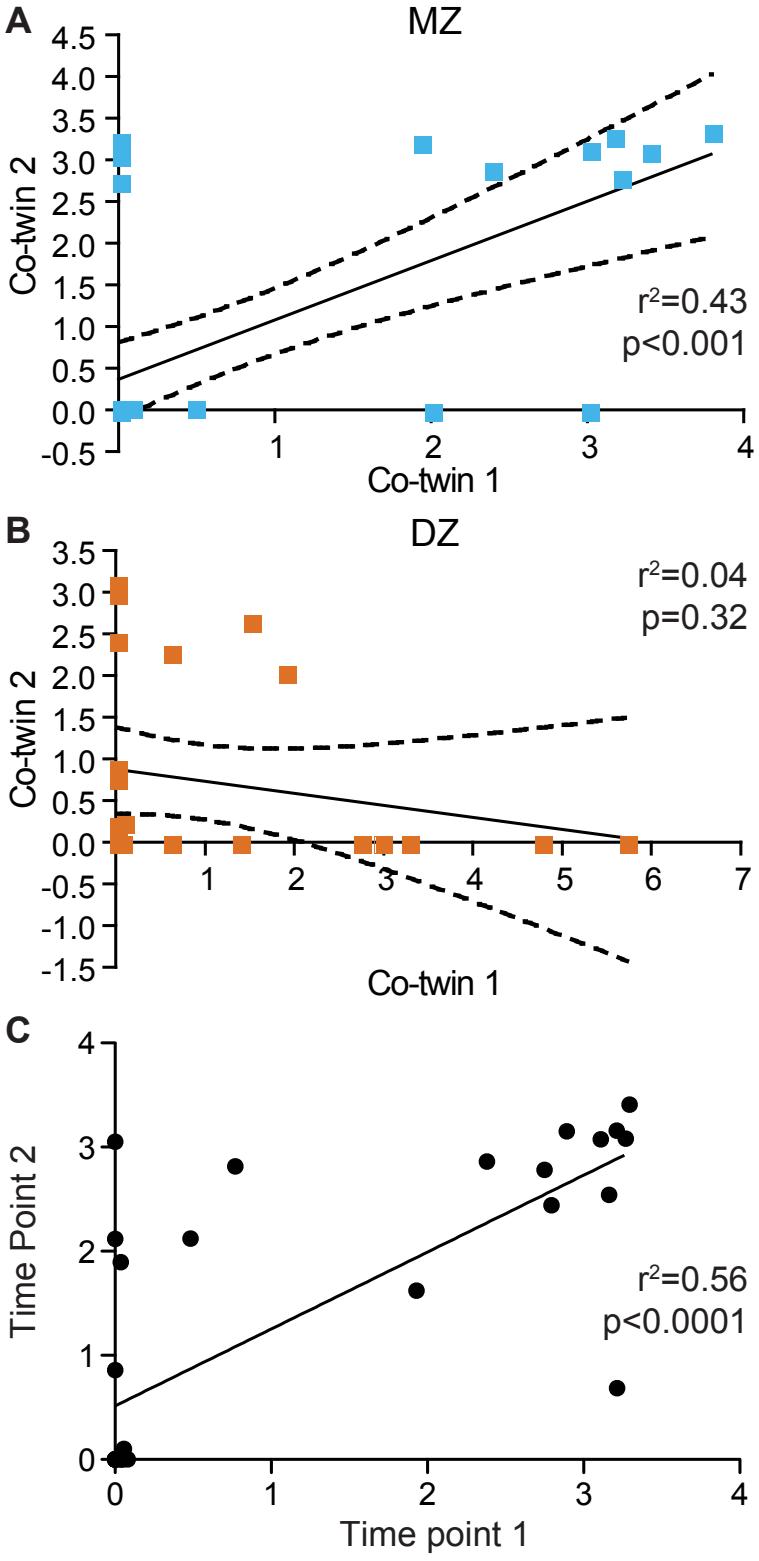


Figure 2.

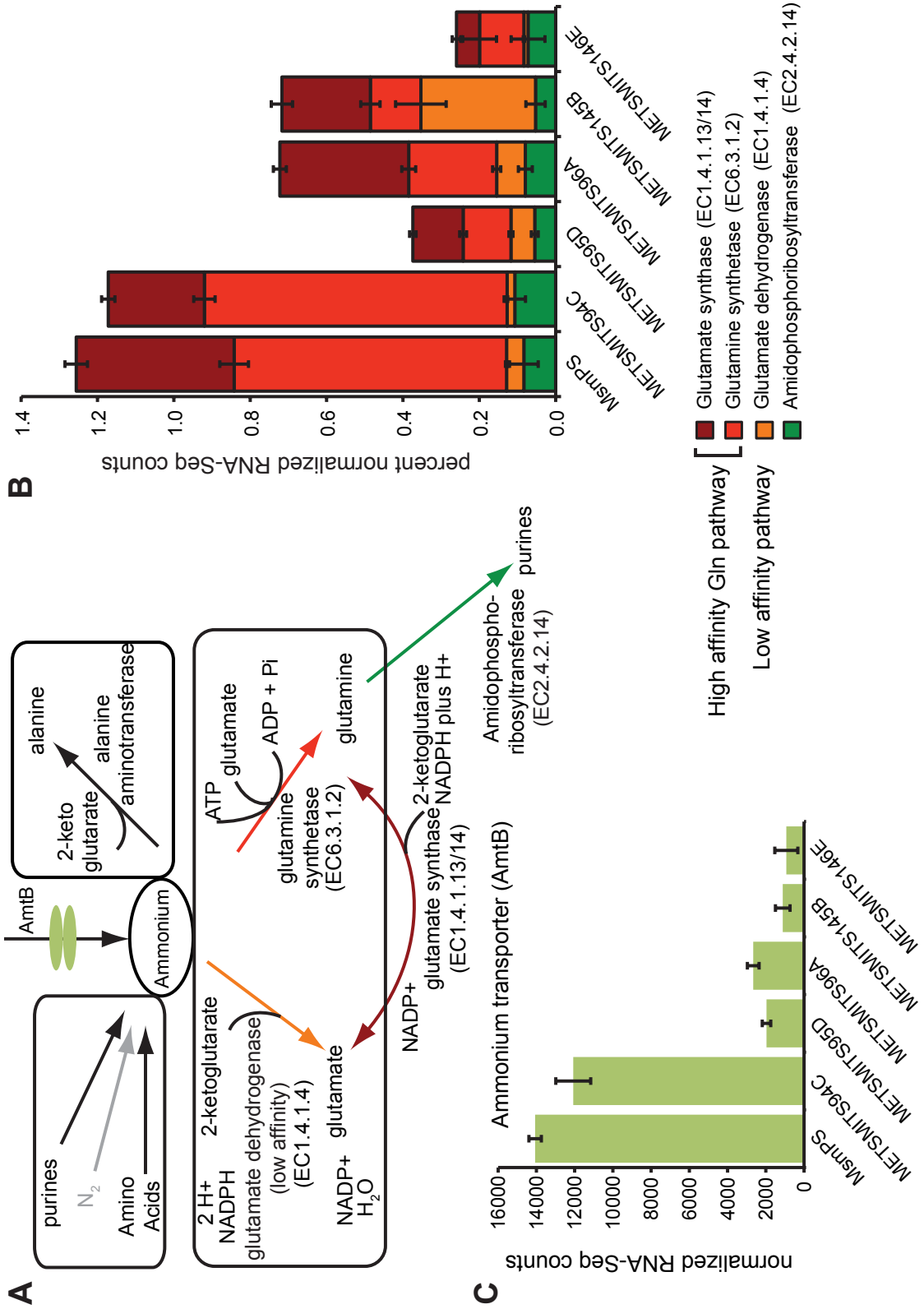

A
purines
N2
Amino Acids
AmtB
Ammonium
alanine
2-keto glutarate
alanine aminotransferase
2 H+
NADPH
2-ketoglutarate
glutamate dehydrogenase (low affinity) (EC1.4.1.4)
NADP+
H2O
glutamate
ATP
glutamate
ADP + Pi
glutamine synthetase (EC6.3.1.2)
glutamine
NADP+
glutamate synthase (EC1.4.1.13/14)
2-ketoglutarate
NADPH plus H+
Amidophospho-ribosyltransferase (EC2.4.2.14)
purines
B
percent normalized RNA-Seq counts
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
MsmPS
MET5MIT5S4C
MET5MIT5S5D
MET5MIT5S9A
MET5MIT5S45B
MET5MIT5S46E
High affinity Gln pathway
Low affinity pathway
Glutamate synthase (EC1.4.1.13/14)
Glutamine synthetase (EC6.3.1.2)
Glutamate dehydrogenase (EC1.4.1.4)
Amidophosphoribosyltransferase (EC2.4.2.14)
C
Ammonium transporter (AmtB)
normalized RNA-Seq counts
16000
14000
12000
10000
8000
6000
4000
2000
0
MsmPS
MET5MIT5S4C
MET5MIT5S5D
MET5MIT5S9A
MET5MIT5S45B
MET5MIT5S46E

Figure 3.

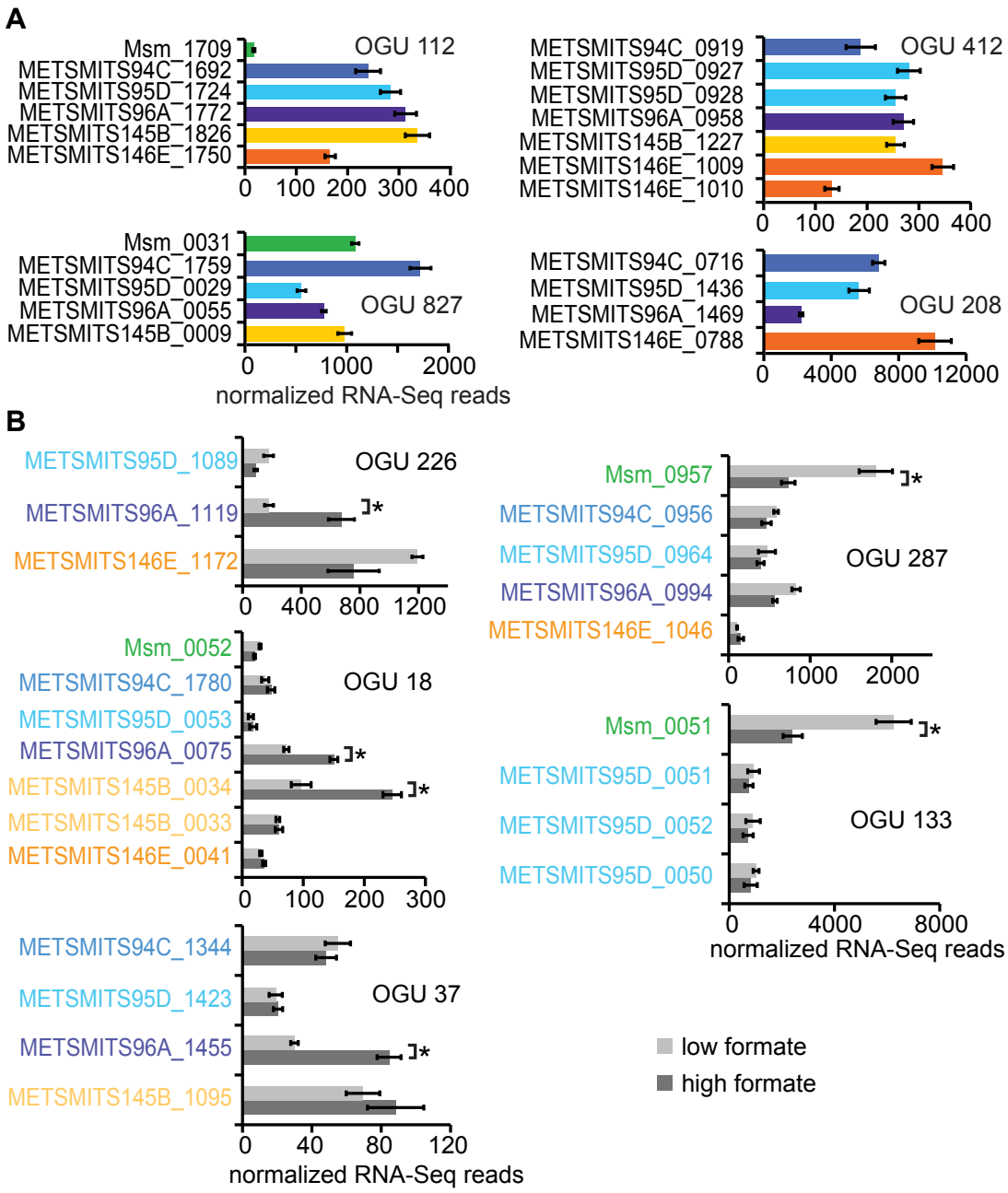

A
Msm_1709
METSMITS94C_1692
METSMITS95D_1724
METSMITS96A_1772
METSMITS145B_1826
METSMITS146E_1750
OGU 112
0 100 200 300 400
METSMITS94C_0919
METSMITS95D_0927
METSMITS95D_0928
METSMITS96A_0958
METSMITS145B_1227
METSMITS146E_1009
METSMITS146E_1010
OGU 412
0 100 200 300 400
Msm_0031
METSMITS94C_1759
METSMITS95D_0029
METSMITS96A_0055
METSMITS145B_0009
OGU 827
0 1000 2000
METSMITS94C_0716
METSMITS95D_1436
METSMITS96A_1469
METSMITS146E_0788
OGU 208
0 4000 8000 12000
normalized RNA-Seq reads
B
METSMITS95D_1089
METSMITS96A_1119
METSMITS146E_1172
OGU 226
*
0 400 800 1200
Msm_0957
METSMITS94C_0956
METSMITS95D_0964
METSMITS96A_0994
METSMITS146E_1046
OGU 287
*
0 1000 2000
Msm_0052
METSMITS94C_1780
METSMITS95D_0053
METSMITS96A_0075
METSMITS145B_0034
METSMITS145B_0033
METSMITS146E_0041
OGU 18
* *
0 100 200 300
Msm_0051
METSMITS95D_0051
METSMITS95D_0052
METSMITS95D_0050
OGU 133
*
0 4000 8000
normalized RNA-Seq reads
METSMITS94C_1344
METSMITS95D_1423
METSMITS96A_1455
METSMITS145B_1095
OGU 37
*
0 40 80 120
normalized RNA-Seq reads
low formate
high formate

Table Legend

Table 1. Bacterial taxa that co-occur with methanogens. OTUs found to be significantly co-occurring with methanogens are shown, together with information about their phylogeny, the percent identity of the V2 regions of their 16S rRNA gene sequence with previously described related bacterial taxa, a p-value for co-occurrence as defined by ANOVA and corrected for multiple hypothesis testing (false discovery rate correction). Significant p-values are noted in red while insignificant values are shown in black or denoted with "NS." The rank is for the ANOVA p-values. G-test p-values are only given for the ones that were significant after applying the fdr correction. Related isolates are followed by their percent nucleotide sequence identity (%ID) to the listed organism over the lanemasked V2 region of their 16S rRNA genes.

Table

Table 1

OTU #	Related bacteria (% identity)	ANOVA p-value			G-test p-value			rank
		Raw	Bonferroni	fdr	Raw	Bonferroni	fdr	
Delta Proteobacteria; Desulfovibrio;								
1908	*D. piger* (87.43) *D. desulfuricans* (90.05)	4.07E-04	2.47E-01	2.24E-01		NS	NS	11
Bacteroidetes;Alistipes;								
4544	*Alistipes putridinis* (91.62)	7.10E-04	4.31E-01	2.87E-02		NS	NS	15
Firmicutes; Clostridiales; Cluster IV; Sporobacter/Oscillospira;								
994	*Oscillospira guilliermondi* (94.38) *Sporobacter termitidis* (89.29)	3.07E-06	1.86E-03	1.86E-03	7.48E-05	4.54E-02	2.27E-02	1
7178	*Oscillospira guilliermondi* (95.58) *Sporobacter termitidis* (89.5)	1.80E-05	1.09E-02	5.45E-03		NS	NS	2
11076	*Oscillospira guilliermondi* (96.02) *Sporobacter termitidis* (89.73)	4.12E-05	2.50E-02	8.33E-03		NS	NS	3
12187	*Oscillospira guilliermondi* (93.02) *Sporobacter termitidis* (93.02)	5.46E-05	3.32E-02	8.29E-03	7.55E-05	4.58E-02	1.53E-02	4
10817	*Oscillospira guilliermondi* (89.08) *Sporobacter termitidis* (88.51)	9.70E-04	5.89E-01	3.10E-02		NS	NS	19
10188	*Oscillospira guilliermondi* (92.61) *Sporobacter termitidis* (88.07)	1.06E-03	6.43E-01	3.22E-02	2.44E-04	1.48E-01	2.96E-02	20
Firmicutes; Clostridiales; Cluster IV; Rennanqily;								
10297	Rennanqilyf3_AY363375 (91.86)	2.82E-04	1.71E-01	2.14E-02		NS	NS	8
10741	Rennanqilyf3_AY363375 (87.01)	3.03E-04	1.84E-01	2.05E-02	2.28E-04	1.39E-01	3.46E-02	9
Firmicutes; Clostridiales; Cluster IV; Anaerotruncus;								
10014	*Clostridium methylpentosum* (92.27) *Anaerotruncus colihominis* (91.16)	2.07E-04	1.26E-01	1.79E-02		NS	NS	7
8310	*Clostridium methylpentosum* (92.61) *Anaerotruncus colihominis* (92.05)	6.13E-04	3.72E-01	2.86E-02		NS	NS	13
Firmicutes; Clostridiales; Catabacter;								
3231	Catabacter sp. YIT12065 AB490809 (85.20)	1.48E-04	8.98E-02	1.80E-02		NS	NS	5
6560	Catabacter sp. YIT12065 AB490809 (92.06)	1.56E-04	9.46E-02	0.016		NS	NS	6
4838	Catabacter sp. YIT12065 AB490809 (81.87)	9.07E-03	5.50E+00	1.34E-01	6.71E-05	4.07E-02	4.07E-02	41
Firmicutes; Clostridiales; Cluster I; Gut Clone Group;								
3247	*Clostridium cellulovorans* (83.25) Kangaroo forestomach isolate YE57 AY442821 (86.46)	3.13E-04	1.90E-01	1.90E-02		NS	NS	10
7622	*Clostridium cellulovorans* (85.71) Kangaroo forestomach isolate YE57 AY442821 (83.52)	7.32E-04	4.44E-01	2.78E-02		NS	NS	16
9347	*Clostridium cellulovorans* (81.71) Kangaroo forestomach isolate YE57 AY442821 (82.95)	9.21E-04	5.59E-01	3.11E-02		NS	NS	18
8770	*Clostridium cellulovorans* (78.41) Kangaroo forestomach isolate YE57 AY442821 (94.92)	1.41E-03	8.59E-01	4.10E-02		NS	NS	21
Firmicutes; Clostridiales; Cluster XIVa;								
2502	*Blautia hydrogenotrophica* (92.51)	6.37E-04	3.87E-01	2.76E-02		NS	NS	14
4531	*Blautia hydrogenotrophica* (89.01)	1.75E-03	1.06E+00	4.82E-02		NS	NS	22
4683	*Coprococcus eutactus* (98.92)	7.72E-04	4.69E-01	2.76E-02		NS	NS	17

Supporting Information

SI Results

Detection of insertion sequence (IS) elements and prophages. A putative rearrangement was discovered in the *M. smithii* PS type strain by aligning draft assemblies of other strains using Mauve (1). This putative rearrangement is further evidenced by flanking transposases (Msm1419, Msm0730). When we first sequenced the type strain (2), we noted a large number of genes predicted to be involved in genome evolution: restriction modification systems, transposases, recombinases and insertion sequence (IS) elements. IS finder (www-is.biotoul.fr) was able to detect matches to a known *M. smithii* IS element, ISM1 which is a member of the ISNCY family, and no other significant matches. However, the number of matches varied between strains quite considerably (Dataset S1, Table S3).

Our recent metagenomic study of the fecal viromes of adult female MZ twins showed that viromes are unique to individuals regardless of their degree of genetic relatedness. Intrapersonal diversity is very low with >95% of virotypes retained over a 1-yr period. Moreover, an individual's virome is dominated by a few temperate phage that exhibit remarkable genetic stability. These results indicated that a predatory Lotka-Volterra (LV)/Kill-the-Winner dynamics manifest in a number of other characterized environmental ecosystems, is notably absent in the distal intestine where a more temperate phage lifestyle is evident (3). Therefore, it was of interest to characterize phage diversity in *M. smithii* as a function of host and family.

Prophages were detected by PhageFinder (4) in 7 of the 20 strains, including 4 of the 5 strains isolating one of the dizygotic twins (TS146), one strain from her co-twin (TS145), and two strains from their mother (TS147). When prophage sequences were blasted against the other strains, prophages were identified in two more strains, one from the mother of the MZ twins (METSMITS96C), and another from TS145 (METSMITS145A) (Dataset S1, Table S3).

To identify regions of variation within these prophage, raw 454 Titanium reads for each strain were aligned (nucmer; ref. 5) to the prophage sequence of the PS type strain (coordinates 1705364:1736208). The results were plotted with Mummer (5) and overlayed to create a single plot with the PS type strain prophage gene calls displayed (**Figure S9).** Regions of greatest variation in the prophage were in genes encoding the phage's tail protein (Msm1684), a putative PeiW-related protein (Msm1691, a predicted pseudomurein endoisopeptidase, see ref. 6) and several hypothetical proteins (Msm1674 and Msm1688).

Horizontal Gene Transfer (HGT). To better understand genomic differences among *M. smithii* strains, HGT was detected by using both compositional and phylogenetic methods. Compositional HGT detection was performed by examining the typicality of dinucleotides, codons, and k-words of lengths 4 and 6 (see *SI Methods* below). Because highly expressed genes are known to contain unusual compositions, genes were scored for typicality against both a whole-genome compositional model and a model built using ribosomal proteins (7, 8). Only genes found to be below the significance threshold when compared against both models were annotated as transferred**.** To select significance thresholds for transfer, genes in each genome were ordered from most to least atypical. As reported (9), gene typicality was observed to increase rapidly for the most extreme genes, and then to rise only gradually for the rest of the genome (Figure S7A). In this case, thresholds were set at the point where the change among overlapping 30 gene windows was less than 0.1% of the score of the previous window.

Among the compositional measures analyzed, the proportion of genes defined as horizontally transferred ranged from 3.3 to 10.1% in the dataset as a whole. However, because the absolute number of horizontally transferred genes predicted can depend on the compositional measure chosen, the stringency of the thresholds selected, the amount of time that has passed since the transfer occurred, and the compositional distinctiveness of gene transfer donors (ref. 10; reviewed in ref. 8), we did not focus our analysis on the absolute magnitude of gene transfer in these lineages. Instead, we were primarily interested

in differences in the frequency of HGT events for different classes of genes, and how this process has contributed to the evolution and specialization of the characterized *M. smithii* strains.

HGT has contributed to both the core and variable components of the M. smithii pan-genome. When using compositional methods, we observe that gene transfer is more frequent in the variable genome than the core. For example, when examining 3-1 dinucleotide use (7) and using the rank order of G scores as the significance threshold, 5.7% of the core genes in the pan-genome show compositional evidence of transfer, as compared with fully 16.4% of the variably represented genes, suggesting an approximately threefold enrichment of gene transfer in the variable relative to the core components of the pan-genome.

However, others have observed that phylogenetic methods tend to detect more ancient transfer events than compositional methods (11). Consistent with these observations, 73% of the genes for which PhyloNet found evidence of HGT were part of *M. smithii*'s core genome, indicating transfer before the divergence of strains. By contrast, most putative HGT events predicted by compositional methods were part of the variable genome (59.3–68.0% of transfers, depending on the method) (Dataset S1, **Table S6)**. This difference may be due in part to the requirement of phylogenetic methods for orthologs of the gene under investigation: Compositional HGT predictions for the subset of genes that could be mapped to KEGG orthology groups were also biased toward the core genome. Genes with both compositional and phylogenetic evidence of transfer tend to be more evenly split between the core and variable genomes than transfers supported by either type of evidence alone (Dataset S1, **Table S6)**.

Taken together, these findings suggest that gene transfer has shaped both the core genome of *M. smithii* and differences between strains. External evidence further supports a role for HGT in shaping the core genome of *M. smithii*: 89.1% of genes within prophage (as detected by PhageFinder) are part of the core genome (Dataset S1, **Table S6)**.

Functional contribution of horizontally transferred genes.To test for differences in the functions contributed to the *M. smithii* pan-genome by the core genome, variable genome, or horizontally transferred genes, each of these three gene sets were annotated to KEGG pathways (level 2). The *M. smithii* core genome is enriched in genes involved in "translation" while being depleted in "membrane transporters" and "unclassified metabolic" genes (Bonferroni-corrected G-test for significance; $P < 0.001$). The variable genome is enriched in genes for membrane transporters, "glycan biosynthesis and metabolism," and genes whose functions are poorly characterized, while being depleted for genes involved in translation (Bonferroni-corrected G-test; $P < 0.001$). Horizontally transferred genes, regardless of the detection method used, are most divergent from the pan-genome in their functional profile than either the core or variable components of the *M.smithii* pan-genome. This finding suggests that gene transfer has contributed significant functional diversity to *M. smithii*.

To understand in more detail the specific categories of genes that have been most frequently transferred, significant HGT results for 3-1 dinucleotide use were pooled across genomes and categorized according to KEGG pathway and KEGG orthology group, weighting genes with multiple pathway annotations on a per gene (rather than per annotation) basis (Dataset S1, **Table S4A**). As previously observed for genomic islands (12), genes of unknown or poorly characterized function dominated the HGT pool. Among genes with known KEGG level 2 pathway annotations, those in the KEGG category for folate biosynthesis were the most frequently transferred (101.7 normalized annotations). Tetrahydromethanopterin (THMP) methyltransferase genes were the most frequently transferred KEGG orthology (KO) within this group (23 putative HGT events for the D subunit). THMP methytransferase (13) participates in both the methanogenesis and folate biosynthesis pathways by transferring a methyl group from 5-Methyl-THMP to coenzyme-M (**Figure S6**). Genes involved in coenzyme-M recycling during methanogenesis were similarly frequently transferred, including methyl-coenzyme M reductase α subunit (EC

2.8.4.1; 23 annotations), and heterodisulfide reductase subunit a (EC 1.8.98.1; 22 annotations). Other frequently exchanged KEGG pathway functions included PST-family polysaccharide transporters (50.5/52.5 normalized annotations were compositionally atypical, representing a 5.3-fold enrichment in the putative HGT pool).

Phylogenetic analysis of HGT revealed similar trends. Genes involved in the KEGG folate biosynthesis pathway are the second most frequently transferred functional class (after unclassified metabolic genes). Methanogenesis genes are also among the most abundant transferred functional classes (rank order 22/173 classes; 278 genes). As in the analysis of genes with atypical dinucleotide compositions, phylogenetic HGT detection found transfer in KO groups involved in methyl-coenzyme M recycling, including those for THMP methyltransferase A,B, and C subunits (EC 2.1.1.86), methyl-coenzyme M reductase system component A2, and heterodisulfide reductase (B and D subunits) (EC 1.8.98.1).

In addition to characterizing KEGG functional categories, we analyzed ALP gene transfer given their proposed importance in *M. smithii* niche specialization. Because the vast majority of ALP genes could not be assigned to KEGG orthology groups, only a small subset could be tested for gene transfer by using phylogenetic methods. Of the ALPs that could be assigned to KO groups, 6/49 (12.2%) were classified as being horizontally transferred using both phylogenetic techniques. When analyzed compositionally, 5 or 6 of 6 of these ALPs were compositionally atypical in dinucleotide use, codon use, and k-words of length 4 or 6.

Remarkably, we found that in the full pool of 854 ALP OGUs, between 52% and 65% show evidence of transfer across a variety of compositional measures, an enrichment of 6.4- to 9.3-fold when normalized to the overall levels of gene transfer predicted by the same methods. ALPs that could be mapped to KO groups were less compositionally atypical than ALPs as a whole (only 30.6%–36.7% were compositionally annotated as transferred for this subgroup). Despite the observation that these genes are highly expressed in

M. smithii strains, the ALPs annotated as possessing compositional evidence of transfer do not match the model for ribosomal proteins in their genome, meaning that their expression level alone does not account for their compositional atypicality. Large-scale HGT of ALPs would be consistent with their variability among strains.

SI Materials and Methods

qPCR Assay of mcrA in Human Fecal Samples. Frozen fecal samples were pulverized by manual grinding under liquid nitrogen, and crude DNA was isolated by bead beating and phenol/chloroform extraction. The Qiagen Blood and Tissue kit was used to clean up the crude DNA, and remove RNA and protein. Twenty nanograms of purified community DNA was amplified by using an Mx3000 real-time PCR system (Stratagene) in 25-μL reaction mixtures containing SYBR-green and 0.8 μM McrA_MLf/r primers (5'-GGT-GGTGTMGGATTCACACARTAYGCWACAGC-3' and 5'- TTCATTGCRTAGTTWG-GRTAGTT-3'; ref. 14) which amplified a ≈450-bp region of mcrA. Cycling conditions were as follows: 40 cycles of denaturing at 94 °C for 45 s, annealing 56 °C for 45 s, extension 72 °C for 30 s, with collections at 79–81 °C. A subsequent dissociation curve was used to examine the homogeneity of amplicons, to detect the presence of primer dimers, and to determine the appropriate collection temperature.

A standard curve was constructed using purified *M. smithii* gDNA at concentrations ranging from 0.01 ng to 10 ng and used to define the concentration of *mcrA* DNA in each of the fecal DNA samples. Based on the known genome size of *M. smithii* PS, we expressed the data as number of genome equivalents (GE) per ng of total fecal DNA. Samples that only produced detectable amplification after 37 cycles of PCR were scored as "negative," as were samples having <40 GE per ng of DNA. Data were not normally distributed; therefore, a log-base 10 transformation was performed.

A subset of samples was selected for amplicon sequencing to determine the identity

and diversity of *mcrA* sequences amplified by these primers, and whether archaeal DNA was present in these samples that was not found by our *mcrA*-based primers. The latter was determined by using PCR primers directed at archaeal 16S rRNA genes [571aF (5'-GCY-TAAAGSRICCGTAGC-3'; ref. 15) and 958aR (5'-YCCGGCGTTGAMTCCAATT-3'; ref. 16)] and the following cycling conditions; 30 cycles of denaturing at 94 °C for 2 min, annealing at 65 °C for 45 s, and extension at 72 °C for 2 min. Amplicons were sequenced using the method of Sanger (Retrogen).

Isolation and Culturing of *M. smithii* from Human Fecal Samples. Two-gallon stainless steel paint canisters (Binks; catalog number 83S-210) were modified for incubation of plates at 37 °C in an oxygen-free mixture of 20% CO_2/80% H_2 at a pressure of 15 psi. Canisters contained a heating element (Electro-Flex Pail Heaters) regulated by a custom designed controller consisting of a 16A2120 temperature/process control (Love Controls, Dwyer Instruments), a resistance temperature detector probe to measure the internal tank temperature, and several safety features to prevent overheating or burns. Pressure in each tank was measured and recorded with a digital manometer (LEO record, Omni Instruments). The apparatus was housed inside an anaerobic chamber (COY Labs).

All human fecal samples used in this study were obtained by using protocols approved by the Washington University Human Research Protection Office and its constituent review committees. All samples were deidentified and assigned codes as described in a previous publication (17): Information about the age and BMI of the donors can also be found in this publication. All samples were frozen at –20 °C within 30 min after they had been produced by donors; they were then placed in a standard –80 °C freezer no more than 24 h later, and stored at this temperature for at least 1 yr prior to their use in the present study. An ≈2 g aliquot of a given frozen fecal sample was thawed (inside of the Coy anaerobic Chamber) and serially diluted in modified MBC medium (18) within the anaerobic chamber. Aliquots of serial dilutions (10^{-2} to 10^{-8}) were transferred to 14 mL of MBC supplemented 5% rumen fluid with 10 μg/mL erythromycin, 1 μg/mL ampicillin, 10 μg/

mL vancomycin and 10 mg/mL amphotericin B. The mixture was introduced into 125-mL serum bottles (Bellco Glass). These enrichment cultures were incubated under a fully deoxygenated atmosphere of 20% CO_2/80% H_2 (30 psi of pressure) at 37°C. After at least 7 d, aliquots were plated onto MBC noble agar and were incubated in the custom pressurized tanks described above for colony isolation. In parallel, the same serial dilutions were spread directly onto MBC noble agar plates with antibiotics. All plates were incubated under an atmosphere of 20% CO_2/80% H_2 (15psi of pressure) in our custom PHAT (Pressurized Heated Anaerobic Tank) system at 37 °C. Colonies were picked and screened by PCR of their 16S rRNA genes using bacterial primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1391R (5'-GACGGGCGGTGWGTRCA-3') and archaeal primers 571aF (5'-GCYTAAAGSRICCGTAGC-3') and 958aR (5'-YCCGGCGTTGAMTCCAATT-3'). Amplicons generated from archaeal-directed primers were sequenced using the method of Sanger (Retrogen).

Pure isolates were then cultured anaerobically in MBC medium in a fully deoxygenated atmosphere of 20% CO_2/80% H_2 (30 psi of pressure) at 37°C. Cells were harvested by centrifugation, and DNA was isolated by phenol-chloroform and ethanol precipitation, as described (2). The purity of each DNA preparation was verified by gel electrophoresis.

HGT analysis. For each gene call, compositional statistics were calculated by using the PyCogent code base (19). The statistics included the GC content at each position, three versions of the dinucleotide use (overlapping, nonoverlapping, or "3-1"), all kwords ranging from length 1 through 6, and codon use (Dataset S1, **Table S6A** and **B**).

For each *M. smithii* strain, the composition of each gene was compared against (i) the composition of the genome as a whole and (ii) the composition of highly expressed genes. Genes that mapped to the KEGG orthology (KO) groups for ribosomal proteins were used to calculate the highly expressed test set. The gene and control vectors were compared using either the G-test statistic or Pearson correlation.

The significance of the results was calculated in two ways; first, the Bonferroni corrected P value was calculated for the G-test; second, because the distribution of compositional counts may violate normality, the method of picking significance thresholds based on the rank order of gene scores of Tsirigos *et al*. (9) was employed.

Because highly expressed genes frequently possess unusual gene compositions, gene transfer was predicted only in cases where the gene did not match the whole-genome model, and the gene also did not match the highly expressed model. Annotated tRNAs and rRNAs were also excluded from the analysis.

Phylogenetic confirmation of gene transfers predicted by compositional means was performed using the RIATA-HGT program of the PhyloNet package version 1.7 (20). We obtained all available gene sequences for all KO groups that contained one or more *M. smithii* genes. Annotations for gene family level KEGG assignments were obtained by blasting each protein sequence against version 54 of the KEGG database. The best hit with a KEGG assignment was taken. Multiple assignments were given if the best hit had more than one annotation.

Python scripts were used to generate separate FASTA files for each orthology group containing the amino acid sequences for *M. smithii* and KEGG proteins. All sequences for each orthology group were then separately aligned in MUSCLE (21) by using maxiters = 4, and gene trees for each group were constructed in FASTTREE (22).

PhyloNet requires that no paralogs be present on protein trees. Therefore, multiple members of a KO present in a single KEGG genome were reduced to a single copy by removing sequences that produced the longest branches on the resulting phylogenetic tree. However, for *M. smithii* genes, we wanted to ensure that the process of paralog resolution did not prevent detection of possible xenologs (extra gene copies introduced by gene transfer). Therefore, all *M. smithii* genes were retained in each gene tree in the analysis. The species tree used consisted of the KEGG 16S rRNA sequences for each lineage in the

tree, gathered by BLAST against the *E. coli rrsG* gene, and alignment in PyNAST. The location of "msi," the *M. smithii* strain present in KEGG was taken as the tree position for all *M. smithii*.

Because all multiple copies of gene family members were retained in *M. smithii* genomes, it was necessary to introduce an artificial polytomy into the species tree at the location of msi, with one tip for each paralog/strain combination. This approach is identical to separately running each gene copy, but is computationally more tractable because it avoids reinferring all transfers not involving *M. smithii* across the rest of the tree many times.

Microbial RNA-Seq. *M. smithii* strains were grown in standard MBC medium containing 2.8 or 44.1 mM formate. Medium was prepared anaerobically and aliquoted into 125-mL serum bottles, which were sealed and autoclaved. Triplicate cultures of each strain and condition were grown at 37°C with agitation (100 rpm), in serum bottles containing 21 mL of medium plus 0.5 mL 2.5% Na_2S, under an atmosphere of 80% H_2 and 20% CO_2 that was replenished every 6h to a pressure of 30 psi. Seven milliliters of the culture were harvested at 36 h (**Figure S6A**), placed directly into an equal volume of RNA-Protect (Qiagen), incubated for 5 min at room temperature, then centrifuged for 15 min at 3,220 × g at 4 °C. RNA was harvested by bead beating and phenol-chloroform extraction, and then treated with Turbo DNase (Ambion) and Baseline-ZERO DNase (Epicenter) to remove genomic DNA (23). RNA was then purified with the MEGAClear kit (Ambion) which also removes tRNAs and 5S rRNA. Ribosomal RNA was further depleted by using custom biotinylated oligos (Dataset S1, **Table S5B**) bound to magnetic Streptavidin Dynabeads (Invitrogen). Depleted RNA was reverse transcribed to doublestranded cDNA, then prepared for sequencing on an Illumina GAIIx instrument with 4 nucleotide barcoded adapters (23). Reads were assigned to barcodes, rRNA sequences were pruned, and the remaining reads were mapped to each strain's genome by using custom scripts (23) that use the ssaha mapping algorithm (24).

Comparison of RNA-Seq and Custom Affymetrix *M. smithii* GeneChips. RNA from four samples of *M. smithii* PS (two replicates at each formate concentration) were split into aliquots for subsequent GeneChip target preparation, or for rRNA depletion and RNASeq. Nearly 106 million 36-nt Illumina GA-IIx reads were generated from the four samples (each sample run on a single lane of the eight-lane flow cell): 7.2 million of these reads mapped to coding regions (6.9%), whereas the remaining reads mapped to rRNA genes, or other noncoding regions of the genome. Dataset S1, **Tables S1-S7** were also generated for each replicate sample by using custom *M. smithii* GeneChips that have been described in an earlier report (2). GeneChip data were processed (see ref. 2 for details), and the resulting datasets were compared with RNA-Seq data (counts per million reads, normalized for gene length). The results obtained with each type of data were highly similar: Pearson's correlation r^2 values ranged from 0.86-0.89 for each replicate ($P < 2e\text{-}16$; **Figure S8**).

SI References

1. Darling AC, Mau B, Blattner FR, Perna NT (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394-1403.

2. Samuel BS, *et al*. (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci USA* 104:10643-10648.

3. Reyes A, *et al*. (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466: 334-338.

4. Fouts DE (2006) Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Res*. 34:5839–5851.

5. Delcher AL, Phillippy A, Carlton J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparision. *Nucleic Acids Research* 30: 2478-2483.

6. Luo Y, Pfister P, Leisinger T, Wasserfallen A (2002) Pseudomurein endoisopeptidases PeiW and PeiP, two moderately related members of a novel family of proteases produced in Methanothermobacter strains. *FEMS Microbiol Lett* 208(1):47-51.

7. Karlin S, Mrázek J, Campbell AM (1998) Codon usages in different gene classes of the Escherichia coli genome. *Mol Microbiol* 29:1341-1355.

8. Zaneveld JR, Nemergut DR, Knight R (2008) Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology* 154:1-15.

9. Tsirigos A, Rigoutsos I. (2005) A new computational method for the detection of horizontal gene transfer events. *Nucleic Acids Res* 33:922-933

10. Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: Rates of change and exchange. *J Mol Evol* 44:383-397.

11. Ragan MA, Harlow TJ, Beiko RG (2006) Do different surrogate methods detect lateral genetic transfer events of different relative ages? *Trends Microbiol* 14:4-8

12. Hsiao WW, *et al* (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet* 1(5):e62.

13. Sauer FD (1986) Tetrahydromethanopterin methyltransferase, a component of the methane-synthesizing complex *of Methanobacterium thermoautotrophicum*. *Biochem Biophys Res Commun* 136:542-547.

14. Hales BA, *et al*. (1996) Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl Environ Microbiol* 62(2):668-675.

15. Eckburg PB, *et al*. (2005) Diversity of the human intestinal microbial flora. *Science*. 308(5728):1635-1638.

16. DeLong EF (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci USA* 89:5685-5689.

17. Turnbaugh PJ, *et al*. (2009) A core gut microbiome in obese and lean twins. *Nature* 457:480-484.

18. Kayar SR, Fahlman A, Lin WC, Whitman WB. (2001) Increasing activity of H_2-metabolizing microbes lowers decompression sickness risk in pigs during H_2 dives. *J. Appl. Physiol* **91,** 2713–2719.

19. Knight R, *et al*. (2007) PyCogent: A toolkit for making sense from sequence. *Genome Biol* 8:R171

20. Than C, Ruths D, Nakhleh L (2008) PhyloNet: A software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.

21. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*.32:1792-1797.

22. Price MN, Dehal PS, Arkin AP (2010) FastTree 2–approximately maximum-likelihood trees for large alignments *PLoS ONE* 5:e9490.

23. Rey FE, *et al*. (2010) Dissecting the *in vivo* metabolic potential of two human gut acetogens. *J Biol Chem* 285: 22082-22090.

24. Ning Z, Cox AJ, Mullikin JC (2001) SSAHA: A fast search method for large DNA databases. *Genome Res* 11:1725-1729.

SI Figure Legends

Figure S1. Bacterial taxa that co-occur with human gut methanogens (*M. smithii*) and their phylogenetic lineages. Sections of the Arb parsimony insertion trees for selected co-occurring lineages. Trees contain all OTUs found in >9 samples and their relatives with cultured representatives or with known biological properties for (**A**) Firmicutes; Clostridiales; Cluster I; Gut Clone Group; (**B**) Proteobacteria; Delta Proteobacteria; Desulfovibrio; and (**C**) Firmicutes; Clostridiales; Cluster IV; Sporobacter/Oscillospira. The Desulfovibrio tree has all OTUs within this group including two, OTU7973 and OTU12216, that were found in <10 samples. The branches of the tree are colored by the co-occurrence index (CI), which is calculated as the log fold difference in the average relative abundance in *M. smithii* positive versus negative samples. Red indicates a positive association with *M. smithii*, blue negative, and purple neutral. The CI scores are listed after the OTU name (the number following the colon). OTUs with a significantly higher relative abundance in *M. smithii*-positive versus -negative individuals (ANOVA, $p<0.05$ with FDR correction) are marked with a star. Internal branches are colored based on the average value for all of the OTUs descending from that node. The branches were colored across a red- blue spectrum using -1.8 and +1.8 as min/max values. These values were selected to represent the range of CI scores (which were between -1.71 and 1.8). OTUs always or never detected in *M. smithii*-positive individuals were assigned the maximum and minimum CI score respectively; a CI could not be calculated for these because it would require dividing by zero.

Figure S2. Comparison of strains based on their SNP content. Draft *M. smithii* genomes were aligned using Mauve, and SNPs were identified within localized collinear blocks (LCBs). (**A)** Pair-wise comparison of shared SNPs among all 20 strains plus the reference type strain (MsmPS). (**B**) Principal components analysis of SNP data reveals clustering by individual and by family. (**C**) Comparison of percent shared SNPs among *M. smithii* strains by familial relationship. The statistical analysis consisted of a one-way ANOVA followed by Tukey's post-hoc analysis.

Figure S3. Analysis of *M. smithii* strains based on their gene content. (**A**) Overview of *M. smithii* pan-genome as defined by operational gene units (OGUs) with >90% identity by CD-Hit. (**B**) Pairwise comparisons of strains for the presence of shared OGUs. Boxes are shaded from light gray to black to display the percent of total OGUs that are shared in a given comparison. The colored insert summarizes *M. smithii* strain nomenclature and relates the nomenclature to the human donor based on family and the zygosity of co-twins. (**C**) Principal components analysis of the OGU table shown in panel B. (**D**) Comparison of percent shared OGUs of *M. smithii* strains by familial relationship. Mean values ± S.E.M. are plotted. The statistical significance of observed differences between groups was determined one-way ANOVA followed by Tukey's post-hoc analysis, with red bars indicating $p<0.001$ and green bars indicating $p<0.01$.

Figure S4. Rarefaction analysis of gene discovery in the *M. smithii* pan-genome. (**A**) Rarefaction curve. Light blue and light orange lines indicate 95% confidence limits. (**B**) OGUs present in strains as shown by the cumulative number of strains containing the OGU. Just over 1,000 OGUs are present in all 10 strains of a family. The MZ family (blue) has a higher number of OGUs present in a greater number of strains (5-10) while the DZ family (orange) has more OGUs present in 2-4 strains.

Figure S5. Discrimination of *M. smithii* strains based on their content of genes encoding COGs and enzymes with assigned enzyme classification (EC) numbers. (**A**) COG assignments in core versus variable OGUs distributed over the various strains. COG assignments were given to all possible OGUs, both for core genes (i.e., OGUs containing genes from all strains) and variably represented genes (OGUs containing genes from one or more of the strains). The left column shows the distribution of COG categories in the defined 'core' component of the *M. smithii* pan-genome. COG categories represented in each strain are displayed as the percent of all OGUs in that strain that had an assigned COG annotation. **(B)** Distribution of strains based on their enzyme classification (EC) assignments. (**C**) ECs were assigned to protein coding genes in each strain using KEGG.

Canonical correspondence analysis was used to determine which ECs contributed to the variation seen between the strains. ECs located furthest from the origin contribute most to the variance of strains. (**D**) Results of a binomial test for enrichment or depletion of ECs in various strains after normalizing to the number of genes in that strain that could be assigned a KEGG annotation.

Figure S6. RNA-Seq of selected *M. smithii* strains. (A) Growth characteristics of *M. smithii* strains when cultured in modified MBC medium (see ***SI methods***) containing either low or high concentrations of formate. All strains were grown under an atmosphere of 80%H_2/20%CO_2 at 30 psi. Gases were replenished every 6h. Aliquots were taken at the time of re-pressurization for measurement of optical density (OD) at 600 nm to monitor growth. **(B)** Normalized RNA-Seq reads assigned to KEGG Gene Families involved in the methanogenesis pathway. For each EC, expression is displayed as mean percent normalized counts (normalized per million reads and to the length of the gene).

Figure S7. Analysis of Horizontal Gene Transfer. (A) Threshold for gene atypicality in strain METSMIALI against the whole-genome model. The vertical axis represents the compositional typicality of each gene in the genome of the METSMIALI type strain. Scores along the vertical axis represent the G-statistic [made negative so as to represent gene typicality following the convention of Tsirigos *et al*. (9)]. A threshold for the significance of atypical genes has been chosen in two ways: either using a rank order threshold (ref. 9, red points) or by naively assuming a normal distribution and applying the Bonferroni corrected G-test (red plus blue points). In this case, the two methods select similar significance thresholds. **(B)** Dinucleotide Atypicality in the METSMIALI genome. The colored trendlines indicate differences between gene dinucleotide composition and the composition of either the whole-genome (black line) or ribosomal proteins (blue lines). Each trendline represents a moving average over a 50-gene window. The gray lines show gene typicality for each gene against the whole genome model. In order for a gene to be scored as transferred, the individual gene typicality must be below the significance threshold (horizontal lines)

for both comparison sets. Tracks along the top of the graph represent gene annotations; from top to bottom, these are, core genome members (thin blue line), ribosomal proteins (blue squares), horizontally transferred genes (green circles), ALP genes (red triangles), degenerate prophage (pink bar), and members of the variable genome (thin black line).

Figure S8. Correlation between *M. smithii* transcriptional profiles generated from RNA-Seq versus GeneChip analyses. RNA samples were processed and analyzed (see ***SI methods***) by both RNA-Seq and by GeneChip. The two platforms yielded highly similar results (Pearson's correlation r^2 values: 0.86-0.89, $p < 2e^{-16}$).

Figure S9. Analysis of prophages present in *M. smithii* strains. Raw 454 titanium sequencing reads from those strains with predicted prophages (see **Table S3**) were mapped onto the *M. smithii* type strain prophage sequence (coordinates 1705364:1736208) using nucmer, and plotted with mummer (5). Axes are from 80-100% similarity.

SI Figures

Figure S1.

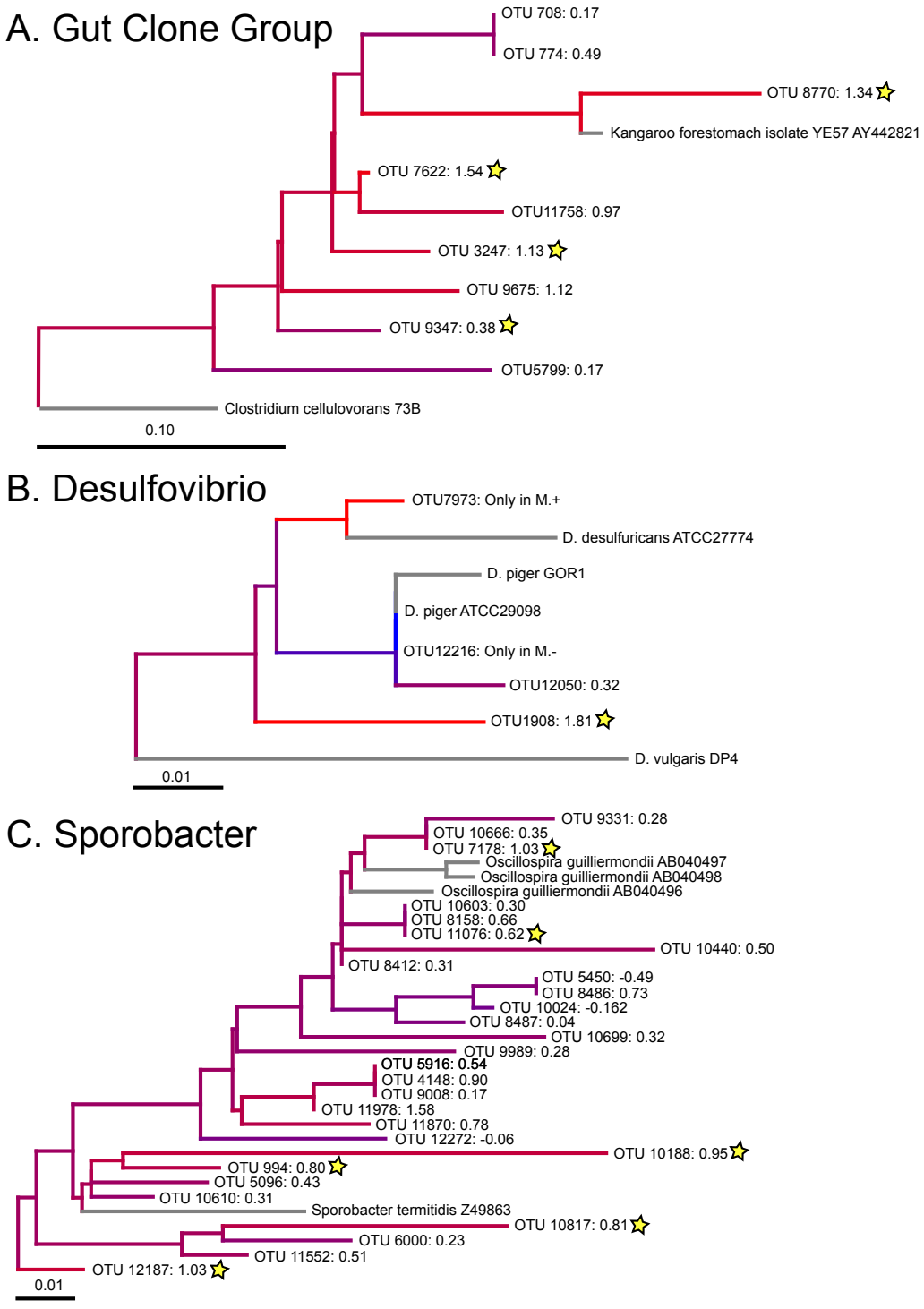


Figure S2.

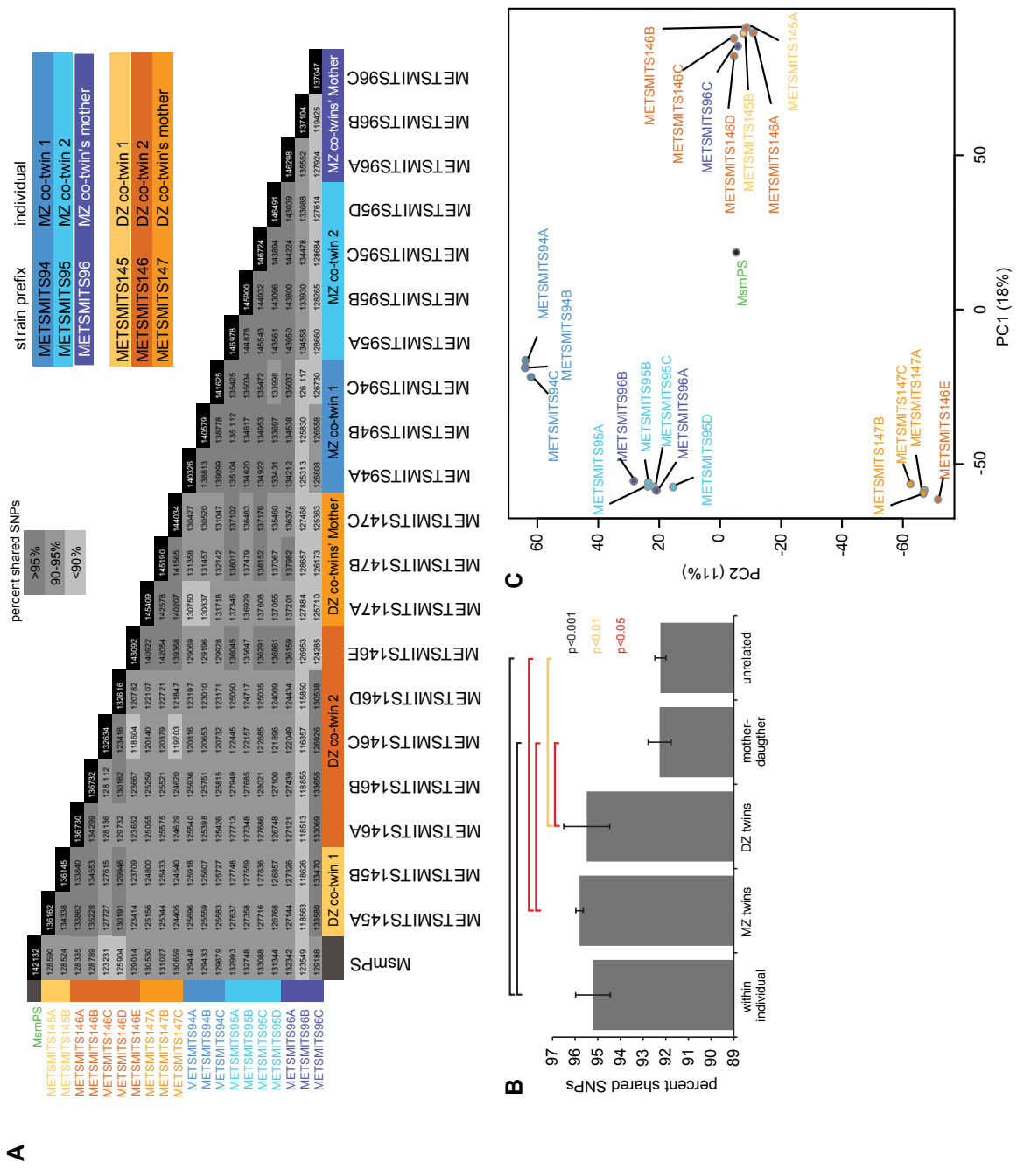

A
percent shared SNPs
>95%
90-95%
<90%
strain prefix
individual
METSMITS94
MZ co-twin 1
METSMITS95
MZ co-twin 2
METSMITS96
MZ co-twin's mother
METSMITS145
DZ co-twin 1
METSMITS146
DZ co-twin 2
METSMITS147
DZ co-twin's mother
MsmPS
METSMITS145A
METSMITS145B
METSMITS146A
METSMITS146B
METSMITS146C
METSMITS146D
METSMITS146E
METSMITS147A
METSMITS147B
METSMITS147C
METSMITS94A
METSMITS94B
METSMITS94C
METSMITS95A
METSMITS95B
METSMITS95C
METSMITS95D
METSMITS96A
METSMITS96B
METSMITS96C
DZ co-twin 1
DZ co-twin 2
DZ co-twins' Mother
MZ co-twin 1
MZ co-twin 2
MZ co-twins' Mother
B
percent shared SNPs
p<0.001
p<0.01
p<0.05
within individual
MZ twins
DZ twins
mother-daughter
unrelated
C
PC2 (11%)
PC1 (18%)
MsmPS
METSMITS145A
METSMITS145B
METSMITS146A
METSMITS146B
METSMITS146C
METSMITS146D
METSMITS146E
METSMITS147A
METSMITS147B
METSMITS147C
METSMITS94A
METSMITS94B
METSMITS94C
METSMITS95A
METSMITS95B
METSMITS95C
METSMITS95D
METSMITS96A
METSMITS96B
METSMITS96C

Figure S3.

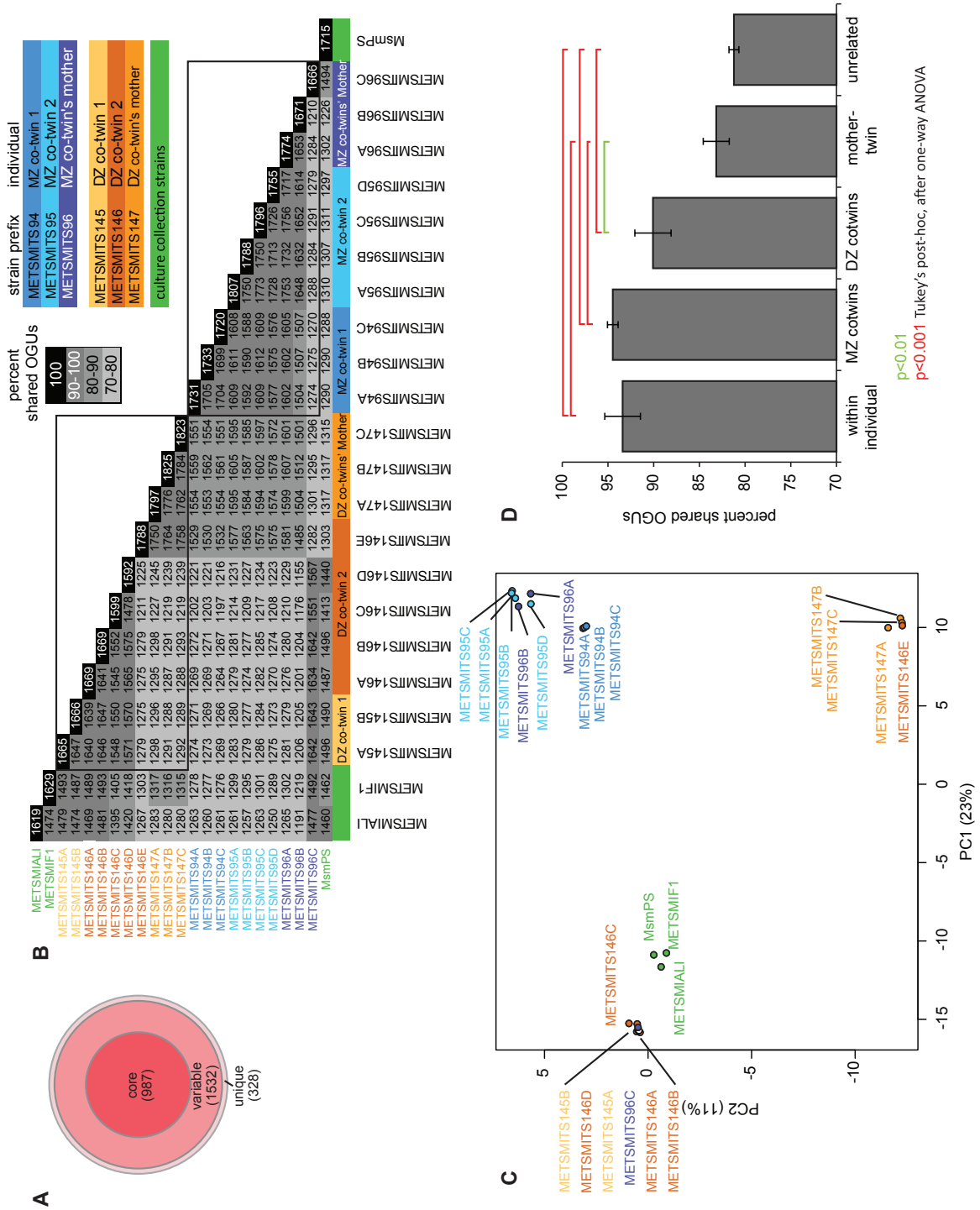

A
core (987)
variable (1532)
unique (328)
B
percent shared OGUs
100
90-100
80-90
70-80
strain prefix
individual
METSMITS94
METSMITS95
METSMITS96
METSMITS145
METSMITS146
METSMITS147
MZ co-twin 1
MZ co-twin 2
MZ co-twin's mother
DZ co-twin 1
DZ co-twin 2
DZ co-twin's mother
culture collection strains
METSMIALI
METSMIF1
MsmPS
DZ co-twin 1
DZ co-twin 2
MZ co-twins' Mother
C
PC1 (23%)
PC2 (11%)
METSMITS95C
METSMITS95A
METSMITS95B
METSMITS96B
METSMITS95D
METSMITS96A
METSMITS94A
METSMITS94B
METSMITS94C
METSMITS147B
METSMITS147C
METSMITS147A
METSMITS146E
METSMITS146C
METSMIALI
METSMIF1
MsmPS
METSMITS146A
METSMITS146B
METSMITS146D
METSMITS145B
METSMITS145A
METSMITS96C
D
percent shared OGUs
within individual
MZ cotwins
DZ cotwins
mother-twin
unrelated
p<0.01
p<0.001 Tukey's post-hoc, after one-way ANOVA

Figure S4.

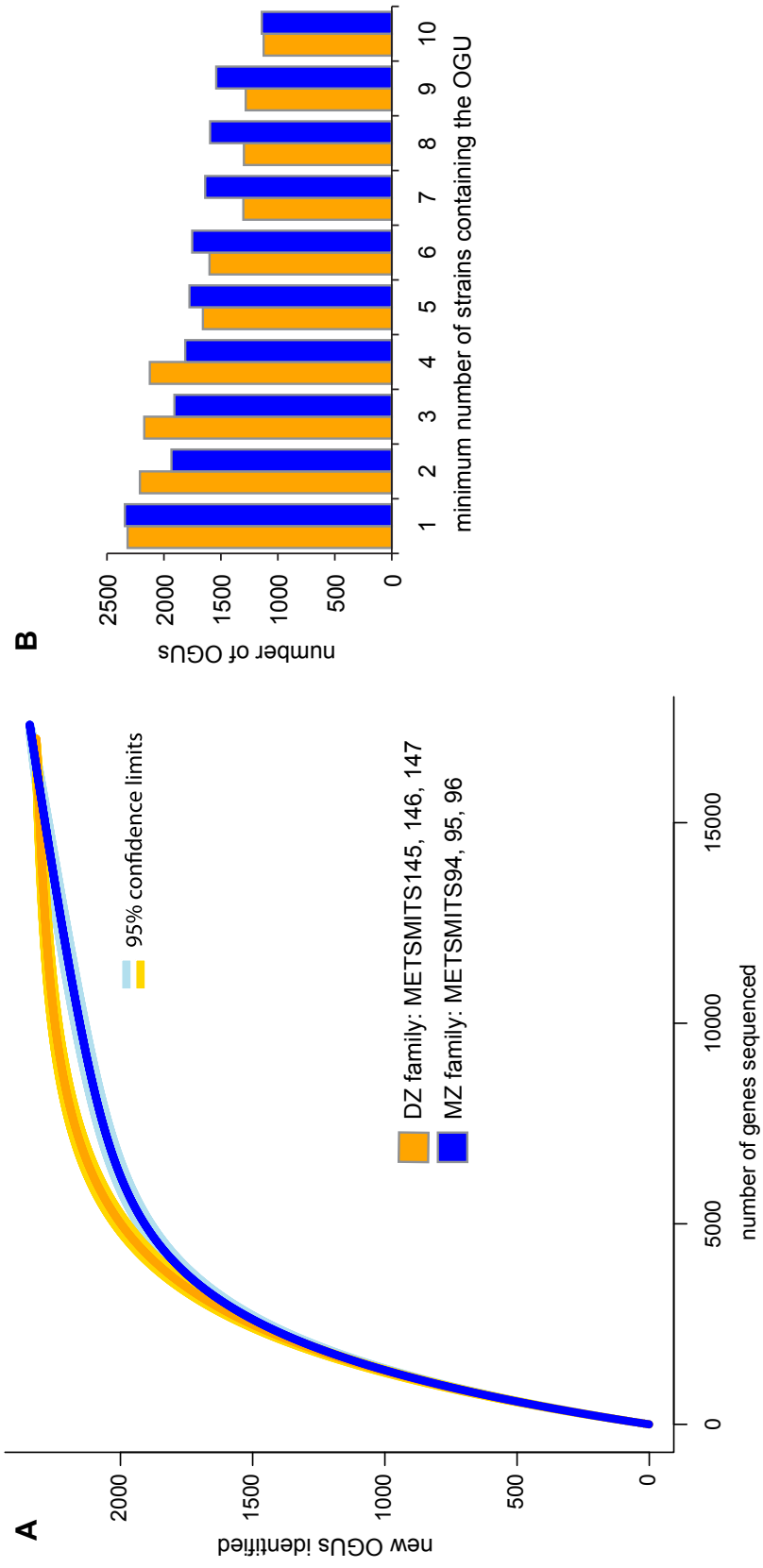

A
new OGUs identified
0
500
1000
1500
2000
0
5000
10000
15000
number of genes sequenced
95% confidence limits
DZ family: METSMITS145, 146, 147
MZ family: METSMITS94, 95, 96
B
number of OGUs
0
500
1000
1500
2000
2500
1
2
3
4
5
6
7
8
9
10
minimum number of strains containing the OGU

Figure S5.

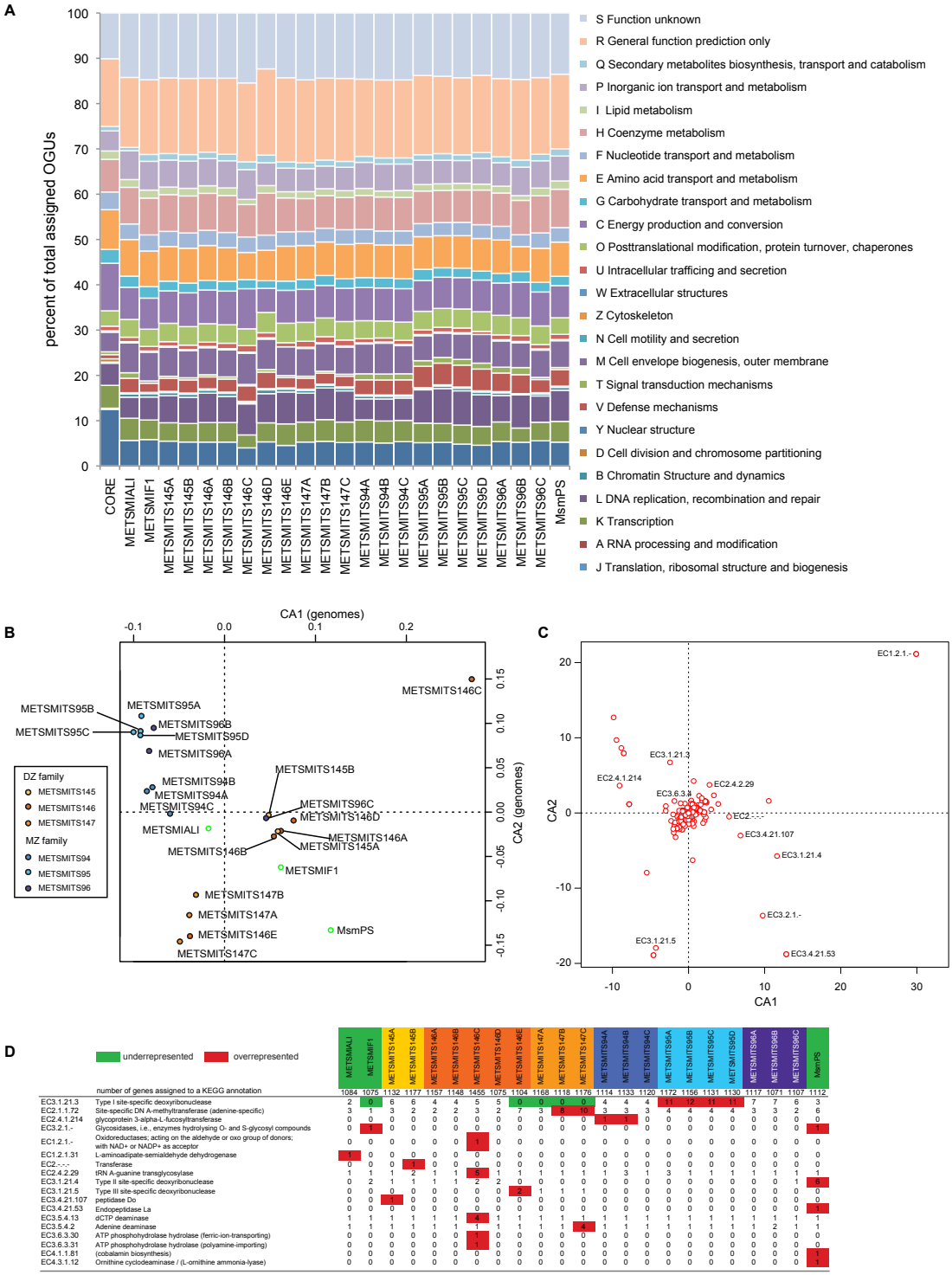

A
percent of total assigned OGUs
S Function unknown
R General function prediction only
Q Secondary metabolites biosynthesis, transport and catabolism
P Inorganic ion transport and metabolism
I Lipid metabolism
H Coenzyme metabolism
F Nucleotide transport and metabolism
E Amino acid transport and metabolism
G Carbohydrate transport and metabolism
C Energy production and conversion
O Posttranslational modification, protein turnover, chaperones
U Intracellular trafficing and secretion
W Extracellular structures
Z Cytoskeleton
N Cell motility and secretion
M Cell envelope biogenesis, outer membrane
T Signal transduction mechanisms
V Defense mechanisms
Y Nuclear structure
D Cell division and chromosome partitioning
B Chromatin Structure and dynamics
L DNA replication, recombination and repair
K Transcription
A RNA processing and modification
J Translation, ribosomal structure and biogenesis
CORE
METSMIALI
METSMIF1
METSMITS145A
METSMITS145B
METSMITS146A
METSMITS146B
METSMITS146C
METSMITS146D
METSMITS146E
METSMITS147A
METSMITS147B
METSMITS147C
METSMITS94A
METSMITS94B
METSMITS94C
METSMITS95A
METSMITS95B
METSMITS95C
METSMITS95D
METSMITS96A
METSMITS96B
METSMITS96C
MsmPS
B
CA1 (genomes)
CA2 (genomes)
DZ family
METSMITS145
METSMITS146
METSMITS147
MZ family
METSMITS94
METSMITS95
METSMITS96
C
CA1
CA2
EC1.2.1.-
EC3.1.21.3
EC2.4.1.214
EC2.4.2.29
EC3.6.3.8
EC2.-.-.-
EC3.4.21.107
EC3.1.21.4
EC3.2.1.-
EC3.1.21.5
EC3.4.21.53
D
underrepresented
overrepresented
number of genes assigned to a KEGG annotation
EC3.1.21.3 Type I site-specific deoxyribonuclease
EC2.1.1.72 Site-specific DN A-methyltransferase (adenine-specific)
EC2.4.1.214 glycoprotein 3-alpha-L-fucosyltransferase
EC3.2.1.- Glycosidases, i.e., enzymes hydrolysing O- and S-glycosyl compounds
EC1.2.1.- Oxidoreductases; acting on the aldehyde or oxo group of donors; with NAD+ or NADP+ as acceptor
EC1.2.1.31 L-aminoadipate-semialdehyde dehydrogenase
EC2.-.-.- Transferase
EC2.4.2.29 tRN A-guanine transglycosylase
EC3.1.21.4 Type II site-specific deoxyribonuclease
EC3.1.21.5 Type III site-specific deoxyribonuclease
EC3.4.21.107 peptidase Do
EC3.4.21.53 Endopeptidase La
EC3.5.4.13 dCTP deaminase
EC3.5.4.2 Adenine deaminase
EC3.6.3.30 ATP phosphohydrolase hydrolase (ferric-ion-transporting)
EC3.6.3.31 ATP phosphohydrolase hydrolase (polyamine-importing)
EC4.1.1.81 (cobalamin biosynthesis)
EC4.3.1.12 Ornithine cyclodeaminase / (L-ornithine ammonia-lyase)

Figure S6.

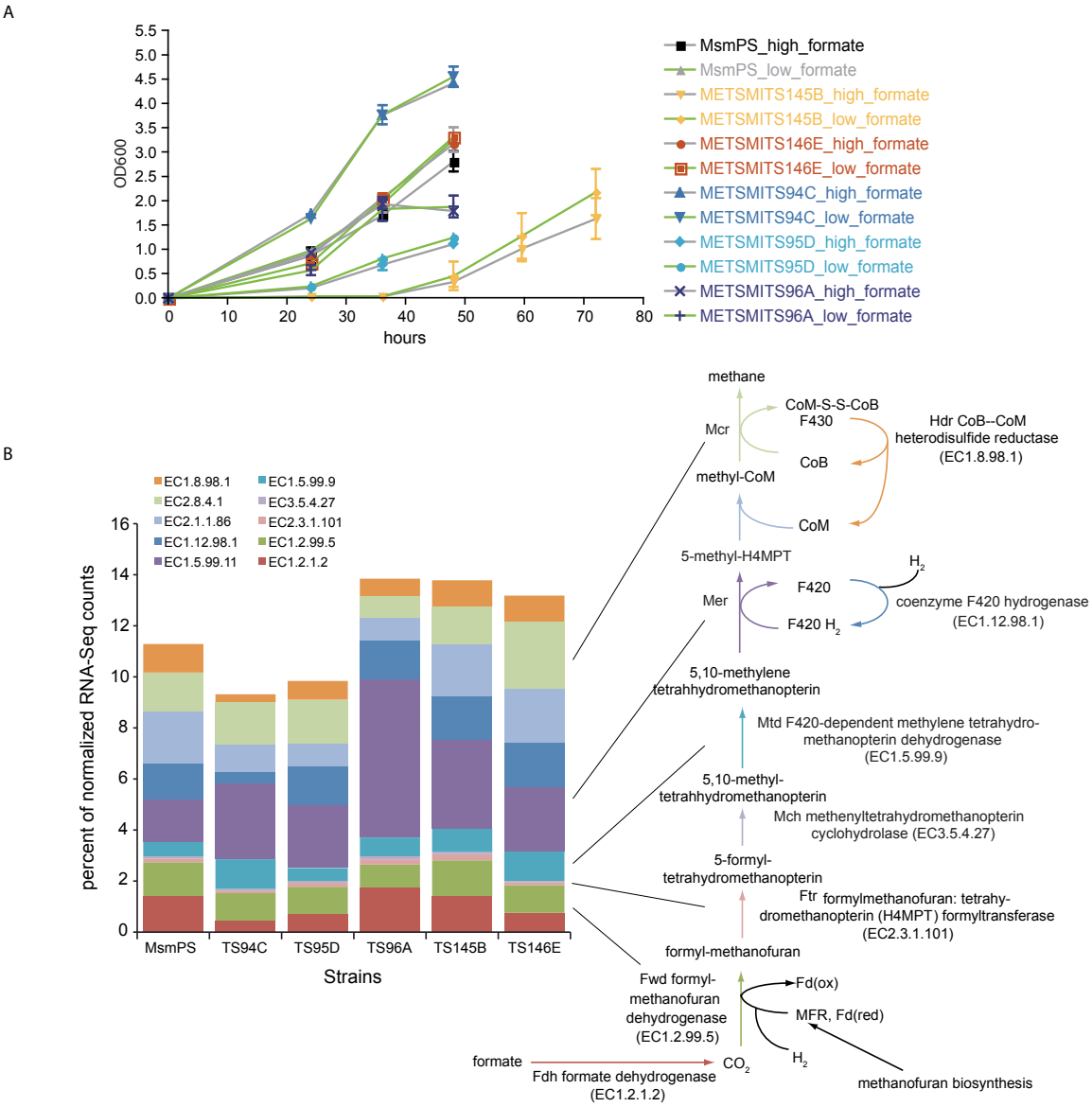

A
OD600
hours
MsmPS_high_formate
MsmPS_low_formate
METSMITS145B_high_formate
METSMITS145B_low_formate
METSMITS146E_high_formate
METSMITS146E_low_formate
METSMITS94C_high_formate
METSMITS94C_low_formate
METSMITS95D_high_formate
METSMITS95D_low_formate
METSMITS96A_high_formate
METSMITS96A_low_formate
B
percent of normalized RNA-Seq counts
Strains
MsmPS
TS94C
TS95D
TS96A
TS145B
TS146E
EC1.8.98.1
EC1.5.99.9
EC2.8.4.1
EC3.5.4.27
EC2.1.1.86
EC2.3.1.101
EC1.12.98.1
EC1.2.99.5
EC1.5.99.11
EC1.2.1.2
methane
Mcr
CoM-S-S-CoB
F430
CoB
Hdr CoB--CoM heterodisulfide reductase (EC1.8.98.1)
methyl-CoM
CoM
5-methyl-H4MPT
Mer
F420
F420 H2
H2
coenzyme F420 hydrogenase (EC1.12.98.1)
5,10-methylene tetrahhydromethanopterin
Mtd F420-dependent methylene tetrahydro-methanopterin dehydrogenase (EC1.5.99.9)
5,10-methyl-tetrahhydromethanopterin
Mch methenyltetrahydromethanopterin cyclohydrolase (EC3.5.4.27)
5-formyl-tetrahydromethanopterin
Ftr formylmethanofuran: tetrahydromethanopterin (H4MPT) formyltransferase (EC2.3.1.101)
formyl-methanofuran
Fwd formyl-methanofuran dehydrogenase (EC1.2.99.5)
Fd(ox)
MFR, Fd(red)
H2
formate
CO2
Fdh formate dehydrogenase (EC1.2.1.2)
methanofuran biosynthesis

Figure S7.

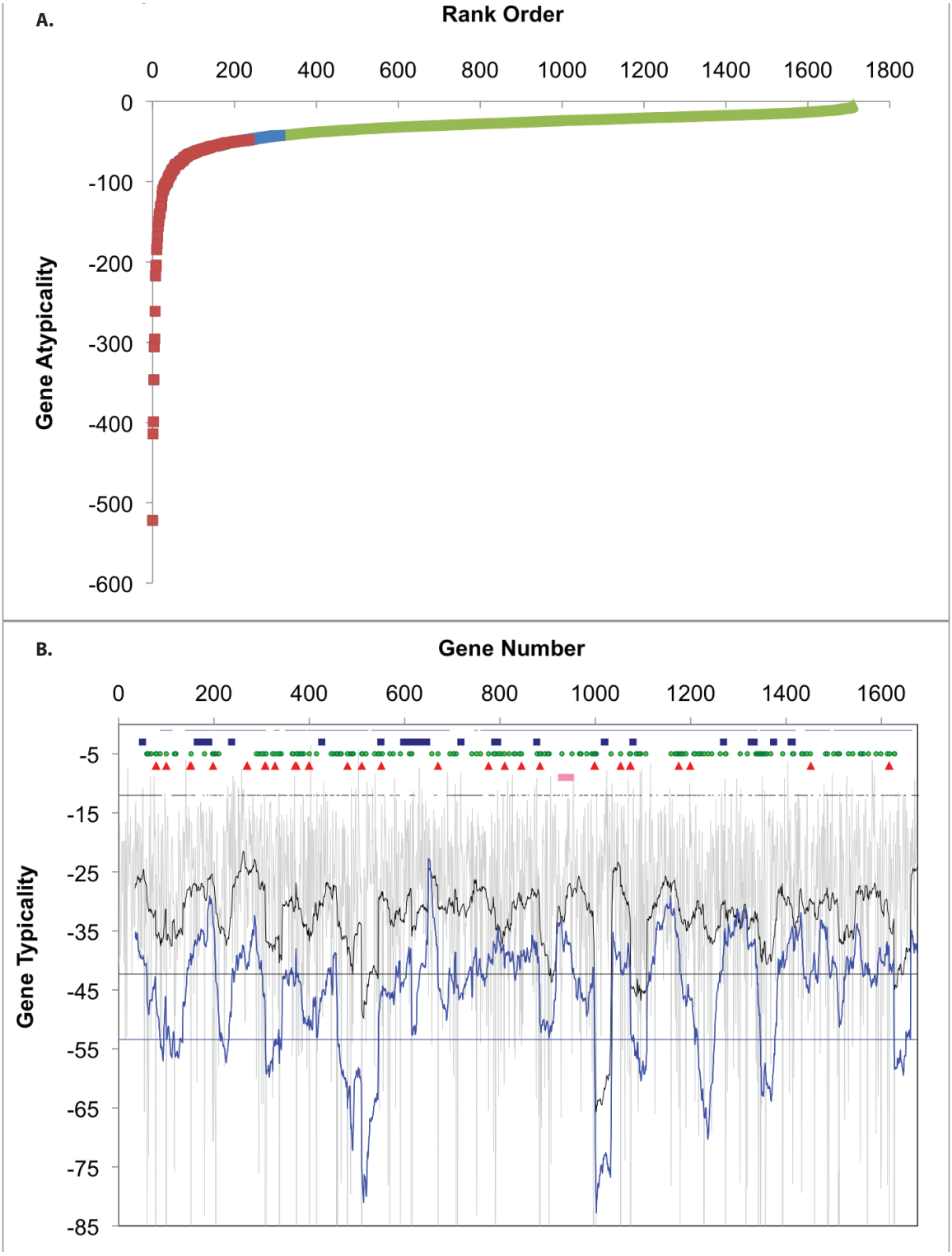

A.
Rank Order
0
200
400
600
800
1000
1200
1400
1600
1800
0
-100
-200
-300
-400
-500
-600
Gene Atypicality
B.
Gene Number
0
200
400
600
800
1000
1200
1400
1600
-5
-15
-25
-35
-45
-55
-65
-75
-85
Gene Typicality

Figure S8.

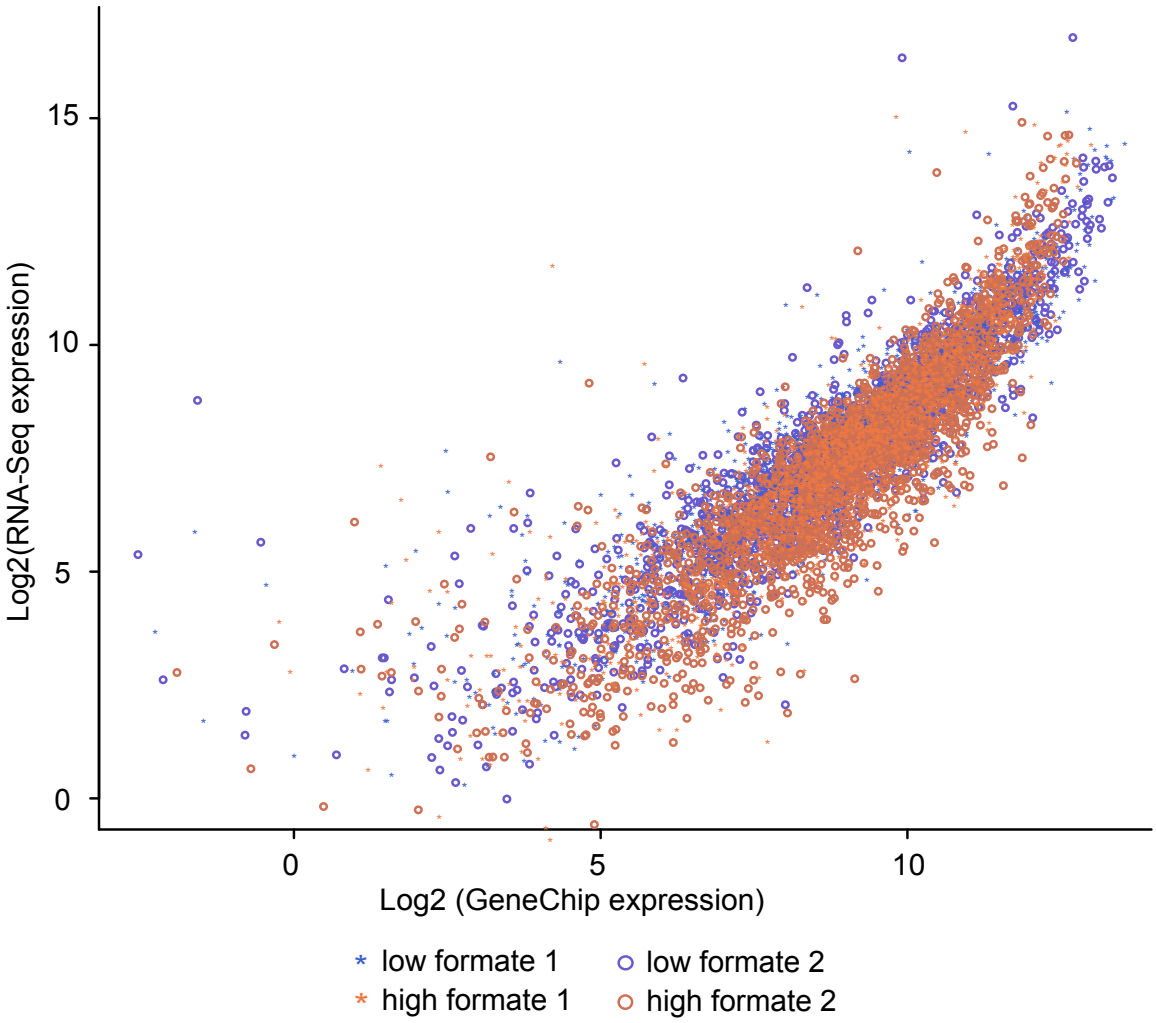
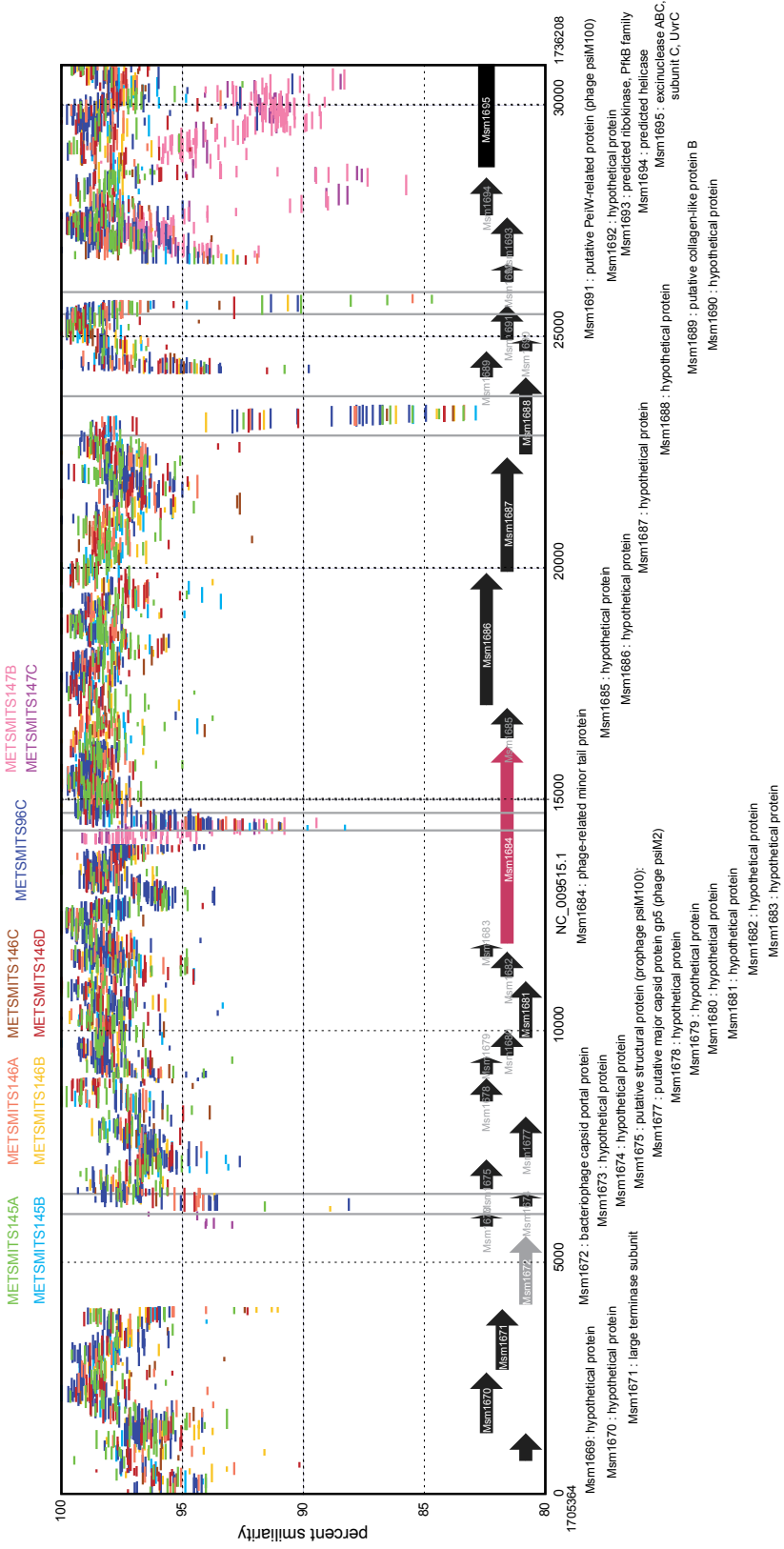

15
10
5
0
Log2(RNA-Seq expression)
0
5
10
Log2 (GeneChip expression)
low formate 1
low formate 2
high formate 1
high formate 2

Figure S9.


METSMITS145A
METSMITS145B
METSMITS146A
METSMITS146B
METSMITS146C
METSMITS146D
METSMITS96C
METSMITS147B
METSMITS147C
percent similarity
100
95
90
85
80
0
1705364
5000
10000
15000
20000
25000
30000
1736208
NC_009515.1
Msm1669: hypothetical protein
Msm1670 : hypothetical protein
Msm1671 : large terminase subunit
Msm1672 : bacteriophage capsid portal protein
Msm1673 : hypothetical protein
Msm1674 : hypothetical protein
Msm1675 : putative structural protein (prophage psiM100):
Msm1677 : putative major capsid protein gp5 (phage psiM2)
Msm1678 : hypothetical protein
Msm1679 : hypothetical protein
Msm1680 : hypothetical protein
Msm1681 : hypothetical protein
Msm1682 : hypothetical protein
Msm1683 : hypothetical protein
Msm1684 : phage-related minor tail protein
Msm1685 : hypothetical protein
Msm1686 : hypothetical protein
Msm1687 : hypothetical protein
Msm1688 : hypothetical protein
Msm1689 : putative collagen-like protein B
Msm1690 : hypothetical protein
Msm1691 : putative PeiW-related protein (phage psiM100)
Msm1692 : hypothetical protein
Msm1693 : predicted ribokinase, PfkB family
Msm1694 : predicted helicase
Msm1695 : excinuclease ABC, subunit C, UvrC

SI Tables

Table S1. Summary of qPCR results for *mcrA* (methanogens) and *aps* (SRB) in fecal samples from MZ and DZ twins, plus the relative abundance of Desulfovibrio taxa (as defined by sequencing the V2 regions of their 16S rRNA genes). qPCR results are shown as $\log_{10}$(genome equivalents per nanogram of DNA). For *mcrA*, results in bold are above our threshold for calling a sample "positive" for methanogens.

Table S2. Bacterial taxa that co-occur with methanogens. OTUs found to be significantly co-occurring with methanogens are shown, together with information about their phylogeny, the percent identity of the V2 regions of their 16S rRNA gene sequence with previously described related bacterial taxa, a p-value for co-occurrence as defined by ANOVA and corrected for multiple hypothesis testing (false discovery rate correction). Significant p-values are noted in red while insignificant values are shown in black or denoted with "NS." The rank is for the ANOVA p-values. G-test p-values are only given for the ones that were significant after applying the FDR correction. Related isolates are followed by their percent nucleotide sequence identity (%ID) to the listed organism over the V2 region of their 16S rRNA genes (after the Lane mask for hypervariable positions was applied).

Table S3. Summary of genome sequencing effort, assembly statistics and annotation results obtained for the 20 *M. smithii* strains isolated in the present study and three previously identified isolates. Genomic DNA was sequenced using an Illumina GA-IIx instrument and a 454 pyrosequencer (Titanium chemistry). Deep draft genome assemblies were obtained using Velvet and Newbler, and hybrid assemblies were produced using AMOS minimus2. Genes were called using Glimmer (v3.02). Prophages were identified using PhageFinder and by Blast against the PS type strain's prophage. Total IS elements identified are shown (using IS Finder, see ***SI methods***) as well as the number of those reaching a minimum length threshold of 58nt.

Table S4. KEGG categories of genes with evidence of horizontal gene transfer (HGT). **(A)** Genes with evidence of HGT by KEGG category compared to the total number of

genes in the *M. smithii* pan-genome assigned to each KEGG category. (B) Genes involved in methane metabolism and folate biosynthesis that show evidence of HGT. (C) Distribution of HGT genes in the core, variable and pan-genome by detection method.

Table S5. Overview of RNA-Seq dataset. (**A**) Average number of reads assigned to protein coding regions (CDS), the total number of mapped reads, and the total number of reads for each strain, averaged across all samples for that strain. (**B**) Sequences of depletion oligos designed to remove *M. smithii* 16S and 23S rRNAs. (C) OGUs present in the *M. smithii* core genome but not in other sequenced methanogens.

Table S6. Compositional evidence for HGT in (A) adhesin-like proteins and (B) the *M. smithii* genome. Fold enrichment is relative to the overall levels of HGT predicted by a given method

Table S7. Genes regulated by formate concentration by strain. Genes significantly regulated by formate were identified for each strain by analyzing normalized reads by CyberT, which calculates a posterior probability of differential expression (PPDE) statistic to determine significance (PPDE ≥ 0.97 and at least a two-fold difference between conditions).

Table S1.

Table S1 - Summary of qPCR results for mcrA (methanogens) and aps (SRB) in fecal samples from MZ and DZ twins, plus the relative abundance of Desulfovibrio taxa (as defined by sequencing the V2 regions of their 16S rRNA genes).

			Quantification of methanogens log (*mcrA*)				SRB	OTUs in lineages related to SRB (relative abundance)			
TS#		zygosity	timepoint 1	timepoint 2	timepoint 3	lineage	log *aps*	taxon 7973	taxon 12216	taxon 12050	taxon 1908
1	Co-twin 1	MZ	**3.163**	**2.542**	**3.053**	*M. smithii*	3.509	0.000503694	0	0	0
2	Co-twin 2	MZ	**3.293**	**3.408**	**3.901**	*M. smithii*		0	0	0	0
4	Co-twin 1	MZ	0.000	0.000	0.000		0.000	0	0	0	0
5	Co-twin 2	MZ	0.000	0.000	0.000		0.000	0	0	0	0
7	Co-twin 1	MZ	0.000	0.000	0.000		0.000	0	0.000179469	8.97344E-05	0
8	Co-twin 2	MZ	0.000	0.000	0.000		3.741	0	0.001734713	0.001053219	0
10	Co-twin 1	MZ	0.000	0.000			0.000	0	0	0	0
11	Co-twin 2	MZ	0.000					0	0	0	0
13	Co-twin 1	MZ	0.000					0	0.001230769	0.001107692	0
14	Co-twin 2	MZ	0.000					0	0.000129266	0	0
16	Co-twin 1	MZ	0.000				3.402	0	0	0.002105263	0
17	Co-twin 2	MZ	**3.243**			*M. smithii* and *M. stadtmanae*		0	0	0	0
19	Co-twin 1	MZ	0.000					0	0	0	0
20	Co-twin 2	MZ	0.000					0	0	0	0
22	Co-twin 1	MZ	0.000				1.744	0	0	0	0
23	Co-twin 2	MZ	0.000				0.000	0	0	0	0
25	Co-twin 1	MZ	0.000	**3.053**				0	0	0	0
26	Co-twin 2	MZ	**2.751**	**2.781**				6.27983E-05	0	0	0
28	Co-twin 1	MZ	**3.790**					0	0	0	0
29	Co-twin 2	MZ	**3.344**					0	0	0	0
31	Co-twin 1	MZ	0.000					0	0	0	0
32	Co-twin 2	MZ	0.000					0	0.001546278	0.001546278	0
34	Co-twin 1	MZ	**3.012**			*M. smithii*	3.073	0	0	0.003594536	0
35	Co-twin 2	MZ	**3.132**			*M. smithii*	3.356	0	0	0.004326123	0
37	Co-twin 1	MZ	0.000				0.000	0	0	0	0
38	Co-twin 2	MZ	0.000					0	0	0	0
40	Co-twin 1	MZ	0.000				2.490				
41	Co-twin 2	MZ	0.000								
43	Co-twin 1	MZ	0.000				1.958	0	0	0	0
44	Co-twin 2	MZ	**3.065**				0.000	0	0	0	0
46	Co-twin 1	MZ	0.000								
47	Co-twin 2	MZ	1.126				0.000				
49	Co-twin 1	MZ	0.000	0.000				0	0	0	0

50	Co-twin 2	MZ	0.000	0.000			0	0	0	0
52	Co-twin 1	MZ	0.000							
53	Co-twin 2	MZ	**2.830**			2.615				
55	Co-twin 1	DZ	0.000				0	0	0	0
56	Co-twin 2	DZ	0.000				0	0	0	0
58	Co-twin 1	MZ	0.000							
59	Co-twin 2	MZ	1.582							
61	Co-twin 1	DZ	0.000	0.000		0.000	0	0	0	0
62	Co-twin 2	DZ	0.052	0.000		0.000	0	0	0	0
64	Co-twin 1	MZ	**3.002**				0	0	0	0
65	Co-twin 2	MZ	0.000			0.000	0	0	0	0
67	Co-twin 1	DZ	0.000	0.000		0.000	0	0	0	0
68	Co-twin 2	DZ	0.769	**2.815**		3.086	0	0	0	0.001277139
70	Co-twin 1	DZ	**3.270**	**3.083**			0	0	0	0
71	Co-twin 2	DZ	0.000	0.858		0.000	0	0	0	0
73	Co-twin 1	DZ	0.000	**2.119**		2.458	0	0	0	0
74	Co-twin 2	DZ	**3.109**	**3.076**		0.000	0	0	0	0
76	Co-twin 1	MZ	0.484	**2.120**		3.293	0	0	0	0.000676361
77	Co-twin 2	MZ	0.037	**1.894**		0.000	0	0	0	0
79	Co-twin 1	MZ	0.000							
80	Co-twin 2	MZ	0.000			0.000				
82	Co-twin 1	MZ	0.000	0.000		2.536	0	0	0	0
83	Co-twin 2	MZ	0.039	0.000		2.613	0	0	0.000645161	0
85	Co-twin 1	DZ	0.000			0.000	0	0	0	0
86	Co-twin 2	DZ	**2.995**			0.000	0	0	0	0
88	Co-twin 1	DZ	0.056	0.103		0.000	0	0	0	0
89	Co-twin 2	DZ	0.000	0.000		2.700	0	0	0	0.000959233
91	Co-twin 1	MZ	0.000			2.084	0	0	0	0
92	Co-twin 2	MZ	0.000			0.000	0	0	0	0
94	Co-twin 1	MZ	**3.212**	**3.159**	*M. smithii*		0	0	0	0.008077544
95	Co-twin 2	MZ	**2.793**	**2.442**	*M. smithii*	2.462	0	0	0	0.011243851
97	Co-twin 1	DZ	0.038	0.000		0.000	0	0	0	0
98	Co-twin 2	DZ	0.010	0.000		2.044	0	0	0	0
100	Co-twin 1	MZ	**1.930**	**1.622**		2.302	0	0	0	0
101	Co-twin 2	MZ	**3.215**	0.685						
103	Co-twin 1	MZ	0.080	0.000		0.000	0	0	0	0
104	Co-twin 2	MZ	0.036	0.000		0.000	0	0	0	0
106	Co-twin 1	MZ	0.000			0.000	0	0	0	0
107	Co-twin 2	MZ	0.000				0	0	0	0

109	Co-twin 1	DZ	0.078				2.006	0	0	0.002912621	0
110	Co-twin 2	DZ	0.249					0	0	0	0
112	Co-twin 1	MZ	0.000				0.000				
113	Co-twin 2	MZ	0.000								
115	Co-twin 1	MZ	**2.381**	**2.860**				0	0	0	0.002368733
116	Co-twin 2	MZ	**2.893**	**3.150**				0	0	0.003847563	0.001832173
118	Co-twin 1	DZ	0.000				0.000	0	0	0	0
119	Co-twin 2	DZ	0.909					0	0	0	0
121	Co-twin 1	MZ	0.000				2.665				
122	Co-twin 2	MZ	0.000								
124	Co-twin 1	DZ	0.000				3.002	0	0	0	0
125	Co-twin 2	DZ	0.000				3.133	0	0	0.00084317	0
127	Co-twin 1	DZ	**5.718**			*M. smithii*		0.000312305	0	0	0
128	Co-twin 2	DZ	0.000				0.000	0	0	0	0
130	Co-twin 1	MZ	0.000				0.000	0	0	0	0
131	Co-twin 2	MZ	0.000					0	0	0	0
133	Co-twin 1	MZ	0.000					0	0	0	0
134	Co-twin 2	MZ	0.000				0.000	0	0	0	0
136	Co-twin 1	DZ	**4.761**					0	0	0	0.003103448
137	Co-twin 2	DZ	0.000				2.833	0	0	0.001086957	0
139	Co-twin 1	DZ	**1.890**					0	0	0.000363504	0
140	Co-twin 2	DZ	**2.044**			*M. smithii*	2.957	0	0	0.002235469	0.002980626
142	Co-twin 1	DZ	0.000				3.857	0	0	0	0
143	Co-twin 2	DZ	0.221				0.000	0	0	0	0
145	Co-twin 1	DZ	1.502			*M. smithii*	4.191	0	0	0	0
146	Co-twin 2	DZ	**2.655**			*M. smithii*	0.000	0	0	0	0
148	Co-twin 1	MZ	0.000					0	0.001706193	0.001023716	0
149	Co-twin 2	MZ	0.000					0	0	0	0
151	Co-twin 1	DZ	0.000				0.000	0	0	0	0
152	Co-twin 2	DZ	**3.004**				2.942	0	0	0	0
154	Co-twin 1	MZ	**3.388**			*M. smithii*	0.000				
155	Co-twin 2	MZ	**3.107**			*M. smithii*	2.221	0	0	0	0
157	Co-twin 1	DZ	1.467				0.000				
158	Co-twin 2	DZ	0.000								
160	Co-twin 1	DZ	0.610				0.000	0	0	0.001730104	0
161	Co-twin 2	DZ	0.000				0.000	0	0	0	0
163	Co-twin 1	MZ	0.000				0.000	0	0	0	0
164	Co-twin 2	MZ	0.000				4.550	0	0	0	0
166	Co-twin 1	DZ	1.378				0.000	0	0	0	0

167	Co-twin 2	DZ	0.000	0.000	0	0	0	0
169	Co-twin 1	DZ	**2.955**	3.880	0	0	0	0.000481696
170	Co-twin 2	DZ	0.000	0.000	0	0	0	0
172	Co-twin 1	MZ	0.000	3.416				
173	Co-twin 2	MZ	0.000	0.000				
175	Co-twin 1	MZ	0.000					
176	Co-twin 2	MZ	0.000	0.000				
178	Co-twin 1	DZ	0.613		0	0	0	0
179	Co-twin 2	DZ	**2.282**	1.651	0	0	0	0
181	Co-twin 1	DZ	0.000	2.505	0	0	0	0
182	Co-twin 2	DZ	**2.430**	4.587	0	0	0	0.012687428
184	Co-twin 1	MZ	**1.996**	0.000	0	0	0	0
185	Co-twin 2	MZ	0.000	0.000	0	0	0	0
187	Co-twin 1	DZ	0.000					
188	Co-twin 2	DZ	0.000					
190	Co-twin 1	MZ	0.000	3.375	0	0	0	0.000644122
191	Co-twin 2	MZ	0.000	0.000	0	0	0	0
193	Co-twin 1	DZ	0.000	3.233	0	0	0.002008032	0
194	Co-twin 2	DZ	0.000	0.000	0	0	0	0
196	Co-twin 1	DZ	0.000	2.820				
197	Co-twin 2	DZ	0.000	0.000				
199	Co-twin 1	DZ	0.000	2.989				
200	Co-twin 2	DZ	0.000	0.000				
202	Co-twin 1	DZ	0.000					
203	Co-twin 2	DZ	0.000					
205	Co-twin 1	DZ	**2.727**					

Table S2.

Table S2 - Bacterial taxa that co-occur with methanogens.

OTU #	Related bacteria (% identity)	ANOVA p-value			G-test p-value			rank
		Raw	Bonferroni	fdr	Raw	Bonferroni	fdr	
Delta Proteobacteria; Desulfovibrio;								
1908	*D. piger* (87.4) *D. desulfuricans* (90)	4.07E-04	2.47E-01	2.24E-02		NS	NS	11
Bacteroidetes;Alistipes;								
4544	*Alistipes putridinis* (91.6)	7.10E-04	4.31E-01	2.87E-02		NS	NS	15
Firmicutes; Clostridiales; Cluster IV; Sporobacter/Oscillospira;								
994	*Oscillospira guilliermondi* (94) *Sporobacter termitidis* (89)	3.07E-06	1.86E-03	1.86E-03	7.48E-05	4.54E-02	2.27E-02	1
7178	*Oscillospira guilliermondi* (95.6) *Sporobacter termitidis* (89.5)	1.80E-05	1.09E-02	5.45E-03		NS	NS	2
11076	*Oscillospira guilliermondi* (96) *Sporobacter termitidis* (89.7)	4.12E-05	2.50E-02	8.33E-03		NS	NS	3
12187	*Oscillospira guilliermondi* (93) *Sporobacter termitidis* (93)	5.46E-05	3.32E-02	8.29E-03	7.55E-05	4.58E-02	1.53E-02	4
10817	*Oscillospira guilliermondi* (89) *Sporobacter termitidis* (88.5)	9.70E-04	5.89E-01	3.10E-02		NS	NS	19
10188	*Oscillospira guilliermondi* (92.6) *Sporobacter termitidis* (88)	1.06E-03	6.43E-01	3.22E-02	2.44E-04	1.48E-01	2.96E-02	20
Firmicutes; Clostridiales; Cluster IV; Rennanqily;								
10297	Rennanqilyf3_AY363375 (91.9)	2.82E-04	1.71E-01	2.14E-02		NS	NS	8
10741	Rennanqilyf3_AY363375 (87)	3.03E-04	1.84E-01	2.05E-02	2.28E-04	1.39E-01	3.46E-02	9
Firmicutes; Clostridiales; Cluster IV; Anaerotruncus;								
10014	*Clostridium methylpentosum* (92) *Anaerotruncus colihominis* (91)	2.07E-04	1.26E-01	1.79E-02		NS	NS	7
8310	*Clostridium methylpentosum* (92.6) *Anaerotruncus colihominis* (92)	6.13E-04	3.72E-01	2.86E-02		NS	NS	13
Firmicutes; Clostridiales; Catabacter;								
3231	Catabacter sp. YIT12065 AB490809 (85)	1.48E-04	8.98E-02	1.80E-02		NS	NS	5
6560	Catabacter sp. YIT12065 AB490809 (92)	1.56E-04	9.46E-02	0.016		NS	NS	6
4838	Catabacter sp. YIT12065 AB490809 (81.9)	9.07E-03	5.50E+00	1.34E-01	6.71E-05	4.07E-02	4.07E-02	41
Firmicutes; Clostridiales; Cluster I; Gut Clone Group;								
3247	*Clostridium cellulovorans* (83.3) Kangaroo forestomach isolate YE57 AY442821 (86.5)	3.13E-04	1.90E-01	1.90E-02		NS	NS	10
7622	*Clostridium cellulovorans* (85.7) Kangaroo forestomach isolate YE57 AY442821 (83.5)	7.32E-04	4.44E-01	2.78E-02		NS	NS	16
9347	*Clostridium cellulovorans* (81.7) Kangaroo forestomach isolate YE57 AY442821 (83)	9.21E-04	5.59E-01	3.11E-02		NS	NS	18
8770	*Clostridium cellulovorans* (78.4) Kangaroo forestomach isolate YE57 AY442821 (94.9)	1.41E-03	8.59E-01	4.10E-02		NS	NS	21
Firmicutes; Clostridiales; Cluster XIVa;								
2502	*Blautia hydrogenotrophica* (92.5)	6.37E-04	3.87E-01	2.76E-02		NS	NS	14
4531	*Blautia hydrogenotrophica* (89)	1.75E-03	1.06E+00	4.82E-02		NS	NS	22
4683	*Coprococcus eutactus* (98.9)	7.72E-04	4.69E-01	2.76E-02		NS	NS	17

Table S3 - Summary of genome sequencing effort, assembly statistics and annotation results obtained for the 20 strains isolated in the present study and 3 previously identified isolates

	strain name	number of 36 nt Illumina reads	number of 454 Titanium reads	number of contigs	N50 contig size	total assembly size	coverage by Illumina	coverage by Titanium	total fold-coverage	number of CDS	number of IS elements identified total (number >58nt)	presence of prophage
MZ twin 1	METSMITS94A	5,049,552	449,545	47	120,002	1,889,378	96	83	179	1808	12(7)	
	METSMITS94B	4,785,200	76,513	58	90,573	1,886,020	91	14	106	1856	12(9)	
	METSMITS94C	20,939,658	433,652	50	108,845	1,910,054	395	79	474	1812	13(9)	
MZ twin 2	METSMITS95A	6,264,402	73,255	56	77,936	1,992,157	113	13	126	1961	17(11)	
	METSMITS95B	3,557,512	85,737	44	133,694	1,972,498	65	15	80	1895	16(8)	
	METSMITS95C	4,559,830	96,757	37	96,923	1,978,848	83	17	100	1874	17(9)	
	METSMITS95D	22,316,058	415,598	58	94,662	2,011,683	399	72	472	1860	20(10)	
Mother of MZ twins	METSMITS96A	29,499,134	260,162	47	98,370	1,975,004	538	46	584	1852	19(11)	
	METSMITS96B	28,356,554	274,657	45	94,662	1,869,210	546	51	598	1742	21(11)	
	METSMITS96C	25,292,727	190,329	108	43,698	1,818,239	501	37	537	1764	3(1)	present
DZ twin 1	METSMITS145A	6,536,457	83,667	44	103,481	1,782,572	132	16	148	1786	2(1)	present
	METSMITS145B	8,277,390	45,203	54	80,226	1,797,373	166	9	175	1880	2(1)	present
DZ twin 2	METSMITS146A	27,011,849	49,854	66	73,601	1,791,997	543	10	552	1823	2(1)	present
	METSMITS146B	26,899,427	58,633	43	147,680	1,794,702	540	11	551	1814	2(1)	present
	METSMITS146C	8,007,300	27,844	102	43,081	1,947,483	148	5	153	2355	3(1)	
	METSMITS146D	9,210,075	73,182	33	139,646	1,713,264	194	15	208	1693	2(1)	present
	METSMITS146E	9,763,978	107,106	64	81,915	1,952,171	180	19	199	1887	11(4)	present
Mother of DZ twins	METSMITS147A	10,284,342	375,219	61	87,700	2,008,979	184	65	250	1969	9(3)	
	METSMITS147B	8,551,491	230,907	40	99,611	1,965,064	157	41	198	1911	11(5)	
	METSMITS147C	9,321,088	68,487	40	256,349	1,973,030	170	12	182	2014	10(3)	
Culture Collection (previously sequenced)	MsmPS (NC_009515)			1		1,853,160				1793	71(51)	present
	METSMIALI (DSM2375)			24	226,159	1,704,865				1679	14(9)	
	METSMIF1 (DSM2374)			25	1,043,555	1,727,775				1688	2(1)	

Table S4A.

Table S4A - KEGG categories of genes with evidence of horizontal gene transfer

KEGG Pathway	Compositionally Atypical Genes in pathway*	Percent	All genes in pan-genome in pathway	Percent	Fold Enrichment
Unclassified; Poorly Characterized	215	12.1	3067	13.0	0.93
Metabolism; Metabolism of Cofactors and Vitamins	201	11.3	2395	10.1	1.12
Unclassified; Cellular Processes and Signaling	197	11.1	1031	4.4	2.54
Genetic Information Processing; Replication and Repair	187	10.5	1259	5.3	1.97
Unclassified; Genetic Information Processing	143	8.0	1918	8.1	0.99
Environmental Information Processing; Membrane Transport	133	7.5	1268	5.4	1.39
Unclassified; Metabolism	125	7.0	1881	8.0	0.88
Metabolism; Carbohydrate Metabolism	75	4.2	1371	5.8	0.73
Metabolism; Nucleotide Metabolism	70	3.9	1237	5.2	0.75
Metabolism; Glycan Biosynthesis and Metabolism	62	3.5	298	1.3	2.74
Metabolism; Enzyme Families	60	3.4	402	1.7	1.97
Metabolism; Amino Acid Metabolism	58	3.3	1981	8.4	0.39
Metabolism; Energy Metabolism	57	3.2	963	4.1	0.78
Environmental Information Processing; Signaling Molecules and Interaction	52	2.9	78	0.3	8.89
Genetic Information Processing; Folding, Sorting and Degradation	24	1.4	384	1.6	0.84
Metabolism; Xenobiotics Biodegradation and Metabolism	21	1.2	516	2.2	0.53
Metabolism; Metabolism of Other Amino Acids	17	1.0	269	1.1	0.85
Cellular Processes; Cell Motility	15	0.8	57	0.2	3.50
Human Diseases; Infectious Diseases	11	0.6	69	0.3	2.09
Environmental Information Processing; Signal Transduction	10	0.6	119	0.5	1.12
Genetic Information Processing; Translation	9	0.5	2010	8.5	0.06
Cellular Processes; Transport and Catabolism	8	0.4	34	0.1	3.09
Genetic Information Processing; Transcription	8	0.4	382	1.6	0.28
Organismal Systems; Immune System	7	0.4	23	0.1	4.38
Human Diseases; Neurodegenerative Diseases	7	0.4	53	0.2	1.69
Organismal Systems; Excretory System	3	0.2	21	0.1	1.77
Metabolism; Biosynthesis of Polyketides and Terpenoids	2	0.1	292	1.2	0.08
Organismal Systems; Environmental Adaptation	1	0.1	6	0.0	2.80
Organismal Systems; Circulatory System	1	0.1	3	0.0	4.80
Metabolism; Lipid Metabolism	1	0.1	246	1.0	0.05
Metabolism; Biosynthesis of Other Secondary Metabolites	0	0.0	135	0.6	0.02

*Genes shown are atypical in 3-1 dinucleotide usage

Table S4B.

Table S4B-Genes involved in methane metabolism and folate biosynthesis that show evidence of HGT

Analyses used	GENE_ID	KO_ID	CAT1	CAT2	CAT3	DESCRIPTION
PhyloNet and codon_usage_G_score_rank_order_th	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMIF1_(	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMIF1_	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K08264	Metabolism	Metabolism c	Folate biosyn	heterodisulfide reductase subunit D [EC:1.8.98.1]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K08264	Metabolism	Metabolism c	Folate biosyn	heterodisulfide reductase subunit D [EC:1.8.98.1]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K08264	Metabolism	Metabolism c	Folate biosyr	heterodisulfide reductase subunit D [EC:1.8.98.1]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and codon_usage_G_score_rank_order_th	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMIF1_(	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMIF1_(	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00320	Metabolism	Metabolism c	Folate biosyr	coenzyme F420-dependent N5,N10-methenyltetrahydromethanopterin reductase [EC:1.5.99.11]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and dinuc_3_1_G_score_rank_order_thres	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIALI_	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIF1_(	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]

PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIF1_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMIF1_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_2_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIALI_	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIF1_	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00122	Metabolism	Energy Metal	Methane met	formate dehydrogenase [EC:1.2.1.2]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIF1_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMIF1_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00579	Metabolism	Metabolism c	Folate biosyr	tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_3_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIALI_	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIF1_(	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIF1_(	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMIF1_'	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyr	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_4_G_score_rank_order_	METSMITS9	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_5_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_5_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]
PhyloNet and kwords_order_5_G_score_rank_order_	METSMITS1	K00205	Metabolism	Metabolism c	Folate biosyn	formylmethanofuran dehydrogenase subunit F [EC:1.2.99.5]

Table S4C.

Table 4C- Distribution of HGT genes in the core, variable and pan-genome by detection method.

Category*	Variable Genome		Core Genome	
	Genes	%	Genes	%
Codons	2695	67.8%	1278	32.2%
Codons (with KO mappings)	816	46.6%	935	53.4%
Dinuc 3-1	2858	68.0%	1342	32.0%
Dinuc 3-1 (with KO mappings)	756	42.5%	1023	57.5%
K-words order 5	1386	59.3%	950	40.7%
K-words order 5 (with KO mappings)	418	32.2%	879	67.8%
PhyloNet	1333	26.0%	3790	73.4%
PhyloNet and codons	174	54.5%	145	45.5%
PhyloNet and dinuc 3-1	146	45.9%	172	54.1%
PhyloNet and kwords order 5	114	40.7%	166	59.3%
Phage	17	10.9%	139	89.1%

*Categories listed as 'with KO mappings' represent the subset of the pan-genome that could be mapped to KEGG orthology groups.

Table S5A.

Table S5A - Overview of RNA-Seq dataset

strain	fraction_CDS	number_CDS	total_mapped	total_reads
METSMITS94C	0.0294	93481	3302490	3429629
METSMITS95D	0.04526	138514	3170100	3278630
METSMITS96A	0.06311	234981	3994000	4095260
METSMITS145B	0.05809	153439	2873157	3116025
METSMITS146E	0.08068	190337	2756408	2895607
MsmPS	0.1027	219511	2621639	2713609
overall	0.06321	171710	3119632	3254793

Table S5B.

Table S5B - Depletion Oligos

16S_depl_61	CTACGACTAAGTTTAGAGGATTACCTCCGC
16S_depl_346	TTGTCTCAGGTTCCATCTCCGGGCTCTTGC
16S_depl_595	CTAAGGGTAGGTTATCCACGTGTTACTGAG
16S_depl_746	AGGACTACCCGGGTATCTAATCCGGTTCGC
16S_depl_1092	GCGTGGGTCTCGCTCGTTGCCTGACTTAAC
23S_depl_269	AAAAGGGATTCAGTTTGTTCTAAGTCGATT
23S_depl_733	TTCCCTACGACTACAAGGATAAAAACCTTT
23S_depl_1146	AGTCTGAGTTGTTTCTCTTTCGGGACACA
23S_depl_1401	CTGCTACTACTACCAGGATCCACATACCTG
23S_depl_2644	CAGGATGGAAAGAACCGACATCGAAGTAGC
23S_depl_2704	CCAGCTCACGTTCCCCTTTAATGGGCGAAC

Table S5C.

Table S5C-OGUs present in the *M. smithii* core genome but not in other sequenced methanogens*

Cluster	Annotation (M.smithii type strain)	mRNA detecetd in vitro?
Cluster 1042	hypothetical protein Msm_0799 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1066	hypothetical protein Msm_0212 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1086	hypothetical protein Msm_0258 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1102	O-linked GlcNAc transferase [Methanobrevibacter smithii ATCC 35061]	
Cluster 1114	hypothetical protein Msm_0067 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1145	hypothetical protein Msm_1152 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1260	acetylesterase [Methanobrevibacter smithii ATCC 35061]	
Cluster 1348	hypothetical protein Msm_1729 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1388	putative SAM-dependent methyltransferase [Methanobrevibacter smithii ATCC 35061]	
Cluster 1414	cobalt ABC transporter, permease component, CbiQ [Methanobrevibacter smithii ATCC 35061]	
Cluster 1463	hypothetical protein Msm_0499 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1483	hypothetical protein Msm_0529 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1503	hypothetical protein Msm_1205 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1510	putative calcium-binding protein [Methanobrevibacter smithii ATCC 35061]	
Cluster 1641	hypothetical protein Msm_1696 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1665	hypothetical protein Msm_1458 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1672	major facilitator superfamily permease [Methanobrevibacter smithii ATCC 35061]	
Cluster 1826	hypothetical protein Msm_0259 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1876	hypothetical protein Msm_1490 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1883	hypothetical protein Msm_1571 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1888	hypothetical protein Msm_0546 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1933	hypothetical protein Msm_1199 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 1943	hypothetical protein Msm_1470 [Methanobrevibacter smithii ATCC 35061]	Marginal/no expression
Cluster 2011	hypothetical protein Msm_0003 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2016	hypothetical protein Msm_0698 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2030	hypothetical protein Msm_0180 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2035	hypothetical protein Msm_1255 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2052	hypothetical protein Msm_0712 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2069	hypothetical protein Msm_1509 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2089	hypothetical protein Msm_0454 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2134	hypothetical protein Msm_0139 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2169	hypothetical protein Msm_0098 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2174	hypothetical protein Msm_0005 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2181	hypothetical protein Msm_0442 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2206	hypothetical protein Msm_0211 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2269	hypothetical protein Msm_0667 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2299	hypothetical protein Msm_1697 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2338	hypothetical protein Msm_0685 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2390	hypothetical protein Msm_1563 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2402	putative monovalent cation/H+ antiporter subunit F [Methanobrevibacter smithii ATCC 35061]	
Cluster 2427	hypothetical protein Msm_0366 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2491	hypothetical protein Msm_0478 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2521	hypothetical protein Msm_0587 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2545	hypothetical protein Msm_1605 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2579	hypothetical protein Msm_0658 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2590	hypothetical protein Msm_0278 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2595	ferredoxin [Methanobrevibacter smithii ATCC 35061]	
Cluster 2597	preprotein translocase subunit SecE [Methanobrevibacter smithii ATCC 35061]	
Cluster 2606	hypothetical protein Msm_1163 [Methanobrevibacter smithii ATCC 35061]	Marginal
Cluster 2617	hypothetical protein Msm_0782 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 2807	rubredoxin [Methanobrevibacter smithii ATCC 35061]	
Cluster 551	glycerol-3-phosphate cytidyltransferase, TagD [Methanobrevibacter smithii ATCC 35061]	
Cluster 573	hypothetical protein Msm_1543 [Methanobrevibacter smithii ATCC 35061]	yes
Cluster 591	ATPase [Methanobrevibacter smithii ATCC 35061]	
Cluster 810	integrase-recombinase protein [Methanobrevibacter smithii ATCC 35061]	

*Blastp threshold E<10-10; methanogenic species used for the analysis:Methanobrevibacter_ruminantium_M1, Methanocaldococcus_FS406_22, Methanocaldococcus_fervens_AG86, Methanocaldococcus_infernus_ME, Methanocaldococcus_jannaschii_DSM_2661, Methanocaldococcus_vulcanius_M7, Methanocella_paludicola_SANAE, Methanococcoides_burtonii_DSM_6242, Methanococcus_aeolicus_Nankai_3, Methanococcus_maripaludis_C5, Methanococcus_maripaludis_C6, Methanococcus_maripaludis_C7, Methanococcus_maripaludis_S2, Methanococcus_vannielii_SB, Methanococcus_voltae_A3, Methanocorpusculum_labreanum_Z, Methanoculleus_marisnigri_JR1, Methanohalobium_evestigatum_Z_7303, Methanohalophilus_mahii_DSM_5219, Methanoplanus_petrolearius_DSM_11571, Methanopyrus_kandleri_AV19, Methanosaeta_thermophila_PT, Methanosarcina_acetivorans_C2A, Methanosarcina_barkeri_Fusaro, Methanosarcina_mazei_Go1, Methanosphaera_stadtmanae_DSM_3091, Methanosphaerula_palustris_E1_9c, Methanospirillum_hungatei_JF_1, Methanothermobacter_marburgensis_Marburg,Methanothermobacter_thermautotrophicus_Delta_H, Methanothermus_fervidus_DSM_2088. Genome sequences were downloaded from NCBI website.

Table S6A.

Table S6A - Compositional evidence for HGT in adhesin-like proteins

Method	Significance Measure	Atypical/total	Percent	Fold enrichment
3-1 Dinucleotide	Rank order threshold; G-score	558/ 853	65%	6.4
Codon Usage	Rank order threshold; G-score	538/853	63%	6.6
K-words (length 4)	Rank order threshold; G-score	525/853	62%	8.1
K-words (length 6)	Rank order threshold; G-score	445/853	52%	9.3

Table S6B.

Table S6B - Compositional evidence for HGT in the M. smithii genome

Method	Significance Measure	Atypical/Total	Percent
3-1 Dinucleotide	Rank order threshold; G-score	4200/41694	10.1%
3-1 Dinucleotide	Rank order threshold; Pearson correlation	1410/41694	3.3%
Codon Usage	Rank order threshold; G-score	3973/41694	9.5%
Codon Usage	Rank order threshold; Pearson correlation	1675/41694	4.0%
K-words (length 4)	Rank order threshold; G-score	3230/41694	7.7%
K-words (length 4)	Rank order threshold; Pearson correlation	2223/41694	5.3%
K-words (length 6)	Rank order threshold; G-score	2336/41694	5.6%
K-words(length 6)	Rank order threshold; Pearson correlation	3300/41694	7.9%

Table S7.

Table S7 - Genes regulated by formate concentration by strain

Gene	annotation	Normalized RNA-Seq counts High formate	Low formate	fold change	PPDE(p)
Msm1453	hypothetical protein	90.1	253.2	2.81	0.9929
Msm1119	hypothetical protein	1965.4	3984.0	2.03	0.9918
Msm1488	cobalt ABC transporter, permease component, CbiM	869.1	1763.4	2.03	0.9902
Msm1649	hypothetical protein	126.8	39.6	-3.20	0.9858
Msm0585	cobalt ABC transporter, permease component, CbiQ	131.5	320.6	2.44	0.9853
Msm1306	adhesin-like protein (Cluster 86)	93.3	208.0	2.23	0.9841
Msm0957	adhesin-like protein (Cluster 287)	731.2	1805.1	2.47	0.9759
Msm0051	adhesin-like protein (Cluster 133)	2412.7	6249.6	2.59	0.9755
Msm1747	type II restriction enzyme, methylase subunit	52.6	110.5	2.10	0.9722
METSMITS146E_0738	hypothetical protein	3740.8	1409.7	-2.65	0.9999
METSMITS146E_0960	GTPase of unknown function	430.6	185.3	-2.32	0.9994
METSMITS146E_1448	4Fe-4S binding domain	6952.5	13982.6	2.01	0.9973
METSMITS146E_0599	ABC transporter	88.7	30.9	-2.87	0.9967
METSMITS146E_0243	hypothetical protein	59.3	233.1	3.93	0.9953
METSMITS146E_0461	hypothetical protein	267.2	73.7	-3.63	0.9948
METSMITS146E_1097	Tetrahydromethanopterin S-methyltransferase	3051.2	7097.8	2.33	0.9942
METSMITS146E_0307	Carboxymuconolactone decarboxylase family	3318.9	1513.6	-2.19	0.9934
METSMITS146E_1103	Tetrahydromethanopterin S-methyltransferase,	2099.9	5557.2	2.65	0.9883
METSMITS146E_0686	eRF1 domain 1	335.3	143.8	-2.33	0.9875
METSMITS146E_0385	Alcohol dehydrogenase GroES-like domain	357.5	172.6	-2.07	0.9872
METSMITS146E_0783	Formate/nitrite transporter	4542.5	11725.6	2.58	0.9862
METSMITS146E_1104	Tetrahydromethanopterin S-methyltransferase,	2314.3	5551.7	2.40	0.9861
METSMITS146E_1121	N2,N2-dimethylguanosine tRNA methyltransfera	290.6	137.6	-2.11	0.9840
METSMITS146E_0493	hypothetical protein	65.0	12.0	-5.43	0.9817
METSMITS146E_1244	hypothetical protein	208.9	90.1	-2.32	0.9778
METSMITS146E_1854	YLP motif	34.0	9.7	-3.49	0.9753
METSMITS146E_1202	Bacterial regulatory protein, arsR family	1344.2	326.5	-4.12	0.9744
METSMITS146E_1583	MarR family	2807.7	1332.7	-2.11	0.9733
METSMITS146E_0848	Fibronectin-binding protein A N-terminus (Fb	142.8	69.0	-2.07	0.9732
METSMITS146E_0278	Pyridoxal-phosphate dependent enzyme	80.5	33.0	-2.44	0.9726
METSMITS146E_1163	NADH-ubiquinone/plastoquinone oxidoreduct	527.8	1067.8	2.02	0.9714
METSMITS146E_1164	hypothetical protein	262.4	529.8	2.02	0.9707
METSMITS145B_0176	Lyase	352.2	113.7	-3.10	0.9999999
METSMITS145B_1436	Chlamydia polymorphic membrane protein (Chl	298.6	30.9	-9.68	0.9999996
METSMITS145B_0056	tRNA synthetases class I (M)	1177.8	308.9	-3.81	0.9999988
METSMITS145B_1144	Peptidase family M50	289.6	95.7	-3.03	0.9999983
METSMITS145B_0784	hypothetical protein	130.9	454.8	3.47	0.9999942
METSMITS145B_1676	Thiolase, C-terminal domain	2052.7	1019.5	-2.01	0.9999928
METSMITS145B_1188	hypothetical protein	4813.1	911.2	-5.28	0.9999909
METSMITS145B_0880	Ribosomal protein S5, N-terminal domai	563.5	145.1	-3.88	0.9999902
METSMITS145B_1454	Protein of unknown function DUF75	362.4	112.3	-3.23	0.9999869
METSMITS145B_1212	KH domain	2266.0	536.7	-4.22	0.9999837
METSMITS145B_0374	hypothetical protein	817.0	295.3	-2.77	0.9999826
METSMITS145B_1216	RNA polymerase Rpb2, domain 6	457.5	92.3	-4.96	0.9999824
METSMITS145B_0870	TruB family pseudouridylate synthase (N term	261.4	45.0	-5.81	0.9999807
METSMITS145B_0187	Nucleoside diphosphate kinase	1403.1	224.9	-6.24	0.9999804
METSMITS145B_0847	GHMP kinases N terminal domain	279.7	83.4	-3.35	0.9999792
METSMITS145B_0067	Thiamine pyrophosphate enzyme, C-termina	1382.7	387.2	-3.57	0.9999792
METSMITS145B_0414	Permease family	515.1	186.2	-2.77	0.9999771
METSMITS145B_0185	Ribosomal protein S6e	677.9	155.1	-4.37	0.9999756
METSMITS145B_1306	Ribosomal protein L16p/L10e	1838.1	517.9	-3.55	0.9999755
METSMITS145B_0901	Ribosomal protein L3	894.0	183.9	-4.86	0.9999753
METSMITS145B_0387	Glutamine amidotransferases class-II	726.5	116.9	-6.21	0.9999752
METSMITS145B_0644	Ribosomal protein L10	5205.4	905.5	-5.75	0.9999750
METSMITS145B_0184	Elongation factor Tu GTP binding domain	851.1	202.3	-4.21	0.9999750
METSMITS145B_1737	Cobalt transport protein component CbiN	6656.3	2065.1	-3.22	0.9999708
METSMITS145B_0799	Ribosomal protein S8e	4787.9	895.1	-5.35	0.9999677
METSMITS145B_1215	RNA polymerase Rpb1, domain 2	508.8	133.9	-3.80	0.9999634
METSMITS145B_1438	CobN/Magnesium Chelatase	489.4	194.0	-2.52	0.9999603
METSMITS145B_1077	Hsp20/alpha crystallin family	2923.3	5909.2	2.02	0.9999597
METSMITS145B_0242	MarR family	2774.0	7441.9	2.68	0.9999571
METSMITS145B_1847	hypothetical protein	164.0	17.6	-9.29	0.9999571
METSMITS145B_0895	KH domain	543.1	145.8	-3.72	0.9999558
METSMITS145B_0385	Conserved region in glutamate synthase	1020.6	257.5	-3.96	0.9999548

METSMITS145B_0920	Fibronectin-binding protein A N-terminus (Fb	143.5	59.5	-2.41	0.9999506
METSMITS145B_1828	Ferritin-like domain	4084.7	13266.8	3.25	0.9999480
METSMITS145B_0055	Protein of unknown function (DUF530)	590.1	158.8	-3.72	0.9999469
METSMITS145B_1456	Eukaryotic translation initiation factor	652.6	214.8	-3.04	0.9999450
METSMITS145B_1202	Elongation factor Tu GTP binding domain	1306.1	319.2	-4.09	0.9999449
METSMITS145B_0645	Ribosomal protein L1p/L10e family	6968.3	1572.0	-4.43	0.9999448
METSMITS145B_1217	RNA polymerase beta subunit	553.0	113.1	-4.89	0.9999402
METSMITS145B_0860	Ribosomal protein S13/S18	1428.8	335.2	-4.26	0.9999378
METSMITS145B_0585	Binding-protein-dependent transport syst	1242.9	168.7	-7.37	0.9999323
METSMITS145B_0125	M42 glutamyl aminopeptidase	1233.1	419.3	-2.94	0.9999312
METSMITS145B_1525	Protein of unknown function (DUF521)	785.5	268.9	-2.92	0.9999263
METSMITS145B_1214	RNA polymerase Rpb1, domain 5	2850.7	471.7	-6.04	0.9999254
METSMITS145B_0060	Eukaryotic and archaeal DNA primase sma	589.9	282.4	-2.09	0.9999223
METSMITS145B_0655	Tetrahydromethanopterin S-methyltransferase	3532.9	573.1	-6.16	0.9999221
METSMITS145B_1569	Carbonic anhydrase	5284.9	11711.4	2.22	0.9999208
METSMITS145B_1433	DnaJ domain	351.4	114.3	-3.07	0.9999192
METSMITS145B_0851	Enolase, C-terminal TIM barrel domain	243.2	50.1	-4.86	0.9999189
METSMITS145B_1203	Ribosomal protein S7p/S5e	863.9	222.6	-3.88	0.9999147
METSMITS145B_0646	Ribosomal protein L11, N-terminal dom	2943.1	909.8	-3.24	0.9999070
METSMITS145B_0065	Radical SAM superfamily	1138.1	394.6	-2.88	0.9999007
METSMITS145B_1200	Ribosomal protein S10p/S20e	1683.7	511.8	-3.29	0.9998999
METSMITS145B_0845	FMN-dependent dehydrogenase	184.0	47.6	-3.87	0.9998983
METSMITS145B_0317	Ribosomal L15	2318.5	804.7	-2.88	0.9998920
METSMITS145B_0584	hypothetical protein	603.8	65.7	-9.19	0.9998727
METSMITS145B_0780	MotA/TolQ/ExbB proton channel family	203.1	547.6	2.70	0.9998725
METSMITS145B_0053	hypothetical protein	415.8	1428.8	3.44	0.9998667
METSMITS145B_0126	Coenzyme F420 hydrogenase/dehydrogenase,	1760.1	505.3	-3.48	0.9998661
METSMITS145B_0582	ABC transporter	981.0	134.7	-7.28	0.9998638
METSMITS145B_1526	DHH family	434.8	102.7	-4.23	0.9998629
METSMITS145B_0973	TCP-1/cpn60 chaperonin family	2008.9	543.5	-3.70	0.9998577
METSMITS145B_1168	hypothetical protein	152.5	43.1	-3.54	0.9998449
METSMITS145B_0857	RNA polymerase Rpb3/RpoA insert domain	917.6	186.2	-4.93	0.9998397
METSMITS145B_0249	DHH family	387.9	186.1	-2.08	0.9998334
METSMITS145B_0763	Glutamine synthetase, catalytic domain	3520.8	1064.5	-3.31	0.9998141
METSMITS145B_1495	hypothetical protein	22.7	55.2	2.44	0.9998125
METSMITS145B_1572	Aspartate/ornithine carbamoyltransferase, As	475.6	149.2	-3.19	0.9998077
METSMITS145B_0709	Aminotransferase class-V	3124.1	1317.0	-2.37	0.9998010
METSMITS145B_0504	CBS domain pair	844.3	420.0	-2.01	0.9997991
METSMITS145B_0186	Elongation factor Tu GTP binding domain	1355.8	217.8	-6.22	0.9997922
METSMITS145B_1383	hypothetical protein	257.4	72.7	-3.54	0.9997910
METSMITS145B_0739	Ribosomal protein S19e	1414.1	295.0	-4.79	0.9997828
METSMITS145B_0876	hypothetical protein	564.2	184.2	-3.06	0.9997722
METSMITS145B_1314	adhesin-like protein (Cluster 199)	181.4	41.6	-4.36	0.9997633
METSMITS145B_1189	Glutamate/Leucine/Phenylalanine/Valin	3763.6	842.9	-4.46	0.9997607
METSMITS145B_0737	hypothetical protein	1476.0	249.9	-5.91	0.9997553
METSMITS145B_0783	Cna protein B-type domain	406.0	2803.3	6.90	0.9997532
METSMITS145B_0855	Ribosomal protein L13	817.1	154.2	-5.30	0.9997492
METSMITS145B_1213	Ribosomal protein L7Ae/L30e/S12e/Gadd4	2448.7	494.7	-4.95	0.9997411
METSMITS145B_0750	haloacid dehalogenase-like hydrolase	1233.6	352.2	-3.50	0.9997362
METSMITS145B_0388	SNO glutamine amidotransferase family	987.2	172.5	-5.72	0.9997182
METSMITS145B_0486	Mov34/MPN/PAD-1 family	164.5	61.4	-2.68	0.9997154
METSMITS145B_0586	hypothetical protein	5082.5	618.0	-8.22	0.9997120
METSMITS145B_0814	CoA binding domain	1264.6	377.6	-3.35	0.9997018
METSMITS145B_0188	Ribosomal protein L24e	1296.2	242.6	-5.34	0.9996973
METSMITS145B_1747	IMP dehydrogenase / GMP reductase domain	484.8	161.0	-3.01	0.9996970
METSMITS145B_0066	hypothetical protein	2898.0	780.7	-3.71	0.9996914
METSMITS145B_1665	hypothetical protein	154.4	380.2	2.46	0.9996806
METSMITS145B_0858	Ribosomal protein S11	909.4	201.7	-4.51	0.9996784
METSMITS145B_0902	Uncharacterized ACR, COG2106	1763.8	633.2	-2.79	0.9996432
METSMITS145B_0449	BioY family	1347.2	570.2	-2.36	0.9996428
METSMITS145B_1776	hypothetical protein	270.2	775.6	2.87	0.9996396
METSMITS145B_0995	Aconitase C-terminal domain	435.5	153.2	-2.84	0.9996190
METSMITS145B_0115	hypothetical protein	1003.7	292.8	-3.43	0.9996127
METSMITS145B_0190	Ribosomal protein L7Ae/L30e/S12e/Gadd4	2003.0	486.0	-4.12	0.9996116
METSMITS145B_1613	CDC6, C terminal	282.9	105.3	-2.69	0.9996114
METSMITS145B_0477	hypothetical protein	251.0	99.3	-2.53	0.9995923
METSMITS145B_0734	eIF-6 family	1958.1	526.3	-3.72	0.9995904
METSMITS145B_1291	hypothetical protein	1348.3	4333.5	3.21	0.9995674
METSMITS145B_1267	Topoisomerase VI B subunit, transducer	312.3	139.2	-2.24	0.9995495
METSMITS145B_0854	Ribosomal protein S9/S16	758.5	180.4	-4.20	0.9995292

METSMITS145B_0581	PhoU domain	637.5	75.6	-8.43	0.9995225
METSMITS145B_1458	Ribosomal protein L44	1860.0	750.1	-2.48	0.9995215
METSMITS145B_0647	KOW motif	3352.1	996.4	-3.36	0.9995023
METSMITS145B_0656	Domain of unknown function (DUF1867)	666.4	1352.6	2.03	0.9994987
METSMITS145B_1359	Proteasome A-type and B-type	1175.0	339.2	-3.46	0.9994870
METSMITS145B_0859	S4 domain	913.7	172.1	-5.31	0.9994849
METSMITS145B_0692	Ribosomal S3Ae family	3724.4	1025.3	-3.63	0.9994778
METSMITS145B_1434	DnaJ C terminal region	606.2	119.8	-5.06	0.9994702
METSMITS145B_1685	hypothetical protein	461.3	210.2	-2.19	0.9994676
METSMITS145B_1661	Periplasmic binding protein	154.4	366.7	2.38	0.9994386
METSMITS145B_1631	hypothetical protein	1793.0	754.3	-2.38	0.9994340
METSMITS145B_1120	hypothetical protein	123.6	398.8	3.23	0.9994298
METSMITS145B_1455	Nucleolar RNA-binding protein, Nop10p family	366.9	65.6	-5.59	0.9994282
METSMITS145B_1835	Uncharacterized conserved protein (DUF2149)	846.0	238.5	-3.55	0.9994112
METSMITS145B_0843	Polyprenyl synthetase	303.9	124.2	-2.45	0.9994048
METSMITS145B_0711	hypothetical protein	86.5	391.4	4.52	0.9994019
METSMITS145B_0275	adhesin-like protein (Cluster 317)	252.8	976.3	3.86	0.9993871
METSMITS145B_0900	Ribosomal protein L4/L1 family	559.6	135.5	-4.13	0.9993812
METSMITS145B_0817	Adenylosuccinate synthetase	758.6	197.3	-3.84	0.9993764
METSMITS145B_0054	hypothetical protein	411.8	157.2	-2.62	0.9993431
METSMITS145B_1531	Glycoprotease family	208.9	431.2	2.06	0.9993307
METSMITS145B_0415	Phosphoribosyl transferase domain	763.7	174.7	-4.37	0.9993204
METSMITS145B_0228	3' exoribonuclease family, domain 1	618.2	209.4	-2.95	0.9992944
METSMITS145B_1749	Ribosomal L37ae protein family	2003.5	944.9	-2.12	0.9992646
METSMITS145B_0896	Ribosomal protein L22p/L17e	1031.7	317.7	-3.25	0.9992573
METSMITS145B_1585	tRNA synthetases class II (D, K and N)	489.8	178.9	-2.74	0.9992398
METSMITS145B_1060	Staphylococcal nuclease homologue	773.1	1570.0	2.03	0.9992376
METSMITS145B_0898	Ribosomal Proteins L2, C-terminal doma	953.7	279.8	-3.41	0.9992289
METSMITS145B_1584	HI0933-like protein	73.6	23.7	-3.11	0.9992283
METSMITS145B_1307	ABC transporter	65.7	15.2	-4.31	0.9992127
METSMITS145B_0829	Aminotransferase class I and II	575.1	266.6	-2.16	0.9992114
METSMITS145B_0844	Metallo-beta-lactamase superfamily	268.7	66.2	-4.06	0.9991340
METSMITS145B_0189	Ribosomal protein S28e	1819.2	324.7	-5.60	0.9991304
METSMITS145B_0178	Ribosomal protein S24e	1310.5	545.1	-2.40	0.9991251
METSMITS145B_0199	tRNA synthetases class I (W and Y)	340.6	113.2	-3.01	0.9991108
METSMITS145B_0204	TCP-1/cpn60 chaperonin family	1542.1	483.0	-3.19	0.9990865
METSMITS145B_0732	Prefoldin subunit	1034.8	453.6	-2.28	0.9990781
METSMITS145B_1352	hypothetical protein	1820.2	541.5	-3.36	0.9990488
METSMITS145B_0505	Universal stress protein family	3104.8	7598.8	2.45	0.9990446
METSMITS145B_1238	FKBP-type peptidyl-prolyl cis-trans isomeras	902.4	213.9	-4.22	0.9990183
METSMITS145B_0014	hypothetical protein	286.4	808.0	2.82	0.9990161
METSMITS145B_0356	PET112 family, N terminal region	291.8	106.4	-2.74	0.9990137
METSMITS145B_1113	hypothetical protein	712.6	291.6	-2.44	0.9989838
METSMITS145B_1143	MoeA N-terminal region (domain I and II	493.6	191.3	-2.58	0.9989256
METSMITS145B_0386	GXGXG motif	914.6	223.2	-4.10	0.9989225
METSMITS145B_1748	IMP dehydrogenase / GMP reductase domain	769.7	339.6	-2.27	0.9989211
METSMITS145B_1736	Cobalt uptake substrate-specific transmembra	3330.3	912.2	-3.65	0.9989004
METSMITS145B_0888	Ribosomal family S4e	313.8	77.7	-4.04	0.9988336
METSMITS145B_1818	FAD binding domain	1018.4	482.3	-2.11	0.9988300
METSMITS145B_0506	Amidohydrolase family	193.8	399.2	2.06	0.9988200
METSMITS145B_0968	PRC-barrel domain	1888.1	4744.9	2.51	0.9988084
METSMITS145B_1141	tRNA synthetases class I (I, L, M and V)	304.8	146.7	-2.08	0.9987920
METSMITS145B_0933	CBS domain pair	1036.5	2349.7	2.27	0.9987890
METSMITS145B_1093	adhesin-like protein (Cluster 222)	757.7	312.1	-2.43	0.9987764
METSMITS145B_1360	Metallo-beta-lactamase superfamily	142.6	42.4	-3.36	0.9987727
METSMITS145B_1174	RNA polymerase Rpb4	1196.9	381.6	-3.14	0.9987565
METSMITS145B_0144	2,3-bisphosphoglycerate-independent pho	707.8	319.3	-2.22	0.9987488
METSMITS145B_1660	FdhD/NarQ family	72.9	187.5	2.57	0.9987478
METSMITS145B_0290	CAAX amino terminal protease family	935.2	448.3	-2.09	0.9987370
METSMITS145B_0106	hypothetical protein	940.3	466.0	-2.02	0.9987099
METSMITS145B_1602	Amidase	339.2	162.6	-2.09	0.9986677
METSMITS145B_0764	Domain of unknown function DUF128	545.8	161.1	-3.39	0.9986243
METSMITS145B_0534	NMD3 family	326.3	129.9	-2.51	0.9985967
METSMITS145B_1656	ThiC family	5845.1	2037.2	-2.87	0.9985545
METSMITS145B_1262	MoeA N-terminal region (domain I and II	191.0	84.3	-2.27	0.9985525
METSMITS145B_0217	hypothetical protein	190.8	1685.3	8.83	0.9985134
METSMITS145B_1204	Ribosomal protein S12	930.4	310.1	-3.00	0.9984510
METSMITS145B_1738	Cobalt transport protein	276.5	79.6	-3.48	0.9984390
METSMITS145B_0544	Peptidase family U32	172.8	84.6	-2.04	0.9984250
METSMITS145B_0832	hypothetical protein	261.7	64.1	-4.08	0.9984085

METSMITS145B_0894	Ribosomal L29 protein	433.4	92.5	-4.68	0.9984004
METSMITS145B_0887	ribosomal L5P family C-terminus	436.3	99.8	-4.37	0.9983973
METSMITS145B_1249	Conserved carboxylase domain	1286.5	569.9	-2.26	0.9983772
METSMITS145B_1123	ABC-2 type transporter	253.3	118.2	-2.14	0.9983630
METSMITS145B_0070	Cysteine-rich domain	211.0	457.7	2.17	0.9983307
METSMITS145B_0738	Double-stranded DNA-binding domain	2332.3	1120.4	-2.08	0.9983227
METSMITS145B_0166	LSM domain	2243.6	712.9	-3.15	0.9983088
METSMITS145B_0177	Ribosomal protein S27a	732.1	281.9	-2.60	0.9983070
METSMITS145B_0183	hypothetical protein	193.9	34.7	-5.59	0.9982994
METSMITS145B_0892	Domain of unknown function UPF0086	444.7	65.1	-6.83	0.9982397
METSMITS145B_0741	RNAse P Rpr2/Rpp21/SNM1 subunit domain	2696.9	777.6	-3.47	0.9982195
METSMITS145B_0154	adhesin-like protein (Cluster 92)	55.7	120.1	2.16	0.9982156
METSMITS145B_0595	NIF3 (NGG1p interacting factor 3)	153.5	47.2	-3.25	0.9982128
METSMITS145B_0747	DNA topoisomerase	235.6	107.2	-2.20	0.9981900
METSMITS145B_0406	ACT domain	363.3	771.6	2.12	0.9981515
METSMITS145B_0503	hypothetical protein	347.7	142.0	-2.45	0.9981511
METSMITS145B_0875	Integral membrane protein DUF106	726.1	150.6	-4.82	0.9981432
METSMITS145B_0740	CRS1 / YhbY (CRM) domain	2241.9	303.2	-7.40	0.9980914
METSMITS145B_0035	2',5' RNA ligase family	229.0	87.6	-2.61	0.9979889
METSMITS145B_0268	hypothetical protein	36.2	103.1	2.85	0.9979688
METSMITS145B_0179	Protein of unknown function (DUF359)	952.4	388.4	-2.45	0.9979441
METSMITS145B_0883	Ribosomal protein L32	477.7	120.5	-3.97	0.9979379
METSMITS145B_0884	Ribosomal protein L6	448.3	125.7	-3.57	0.9979229
METSMITS145B_1457	Ribosomal protein S27	1013.8	378.9	-2.68	0.9977081
METSMITS145B_0980	hypothetical protein	2545.5	1171.2	-2.17	0.9976843
METSMITS145B_1092	hypothetical protein	658.7	278.5	-2.37	0.9976820
METSMITS145B_1163	8-oxoguanine DNA glycosylase, N-terminal dom	80.8	31.3	-2.58	0.9976816
METSMITS145B_1518	Nitrogen regulatory protein P-II	1068.9	139.1	-7.69	0.9976814
METSMITS145B_0220	Nitrogen regulatory protein P-II	1068.9	139.1	-7.69	0.9976814
METSMITS145B_1850	Rubrerythrin	4998.3	14204.1	2.84	0.9975017
METSMITS145B_1313	hypothetical protein	70.1	19.9	-3.52	0.9974803
METSMITS145B_1430	GrpE	141.5	33.7	-4.19	0.9974753
METSMITS145B_0579	4Fe-4S binding domain	194.2	59.6	-3.26	0.9974621
METSMITS145B_0624	hypothetical protein	276.1	96.5	-2.86	0.9973842
METSMITS145B_0380	hypothetical protein	50.7	23.7	-2.14	0.9973443
METSMITS145B_0181	RNA polymerase Rpb7-like, N-terminal d	1063.1	433.0	-2.46	0.9973353
METSMITS145B_0889	KOW motif	942.2	281.6	-3.35	0.9971608
METSMITS145B_0423	hypothetical protein	681.3	320.1	-2.13	0.9970890
METSMITS145B_1565	hypothetical protein	423.6	865.4	2.04	0.9970848
METSMITS145B_0288	ABC transporter	268.3	126.6	-2.12	0.9970833
METSMITS145B_0877	eubacterial secY protein	1062.7	445.0	-2.39	0.9970693
METSMITS145B_1431	Hsp70 protein	642.9	256.0	-2.51	0.9969834
METSMITS145B_1605	3,4-dihydroxy-2-butanone 4-phosphate sy	554.2	202.0	-2.74	0.9968535
METSMITS145B_1410	Sir2 family	284.3	121.8	-2.33	0.9968306
METSMITS145B_0760	S-adenosyl-L-homocysteine hydrolase, NA	854.7	368.6	-2.32	0.9967602
METSMITS145B_0059	hypothetical protein	386.7	167.5	-2.31	0.9967564
METSMITS145B_1444	Hydrogenase maturation protease	3568.0	7435.4	2.08	0.9966401
METSMITS145B_1096	Chlamydia polymorphic membrane protein (Chl	98.6	31.9	-3.09	0.9966163
METSMITS145B_0856	Ribosomal protein L15	858.0	178.6	-4.80	0.9965206
METSMITS145B_0848	Memo-like protein	132.2	45.9	-2.88	0.9964448
METSMITS145B_0872	hypothetical protein	2094.6	460.3	-4.55	0.9963996
METSMITS145B_0949	ABC transporter	415.5	156.6	-2.65	0.9963777
METSMITS145B_0427	hypothetical protein	520.9	161.3	-3.23	0.9963302
METSMITS145B_1302	Pyridoxal-dependent decarboxylase conse	136.5	37.9	-3.60	0.9962536
METSMITS145B_1538	methylene-5,6,7,8-tetrahydromethanopterin de	8996.2	24598.8	2.73	0.9962188
METSMITS145B_0576	Pyruvate flavodoxin/ferredoxin oxidor	851.8	401.3	-2.12	0.9958999
METSMITS145B_1496	hypothetical protein	438.9	1271.5	2.90	0.9958583
METSMITS145B_1222	Thiamine monophosphate synthase/TENI	152.4	50.0	-3.05	0.9957322
METSMITS145B_0052	hypothetical protein	184.2	546.8	2.97	0.9957191
METSMITS145B_0948	Initiation factor 2 subunit family	476.7	221.5	-2.15	0.9956836
METSMITS145B_0964	Domain of unknown function (DUF1724)	37.1	148.0	3.99	0.9954538
METSMITS145B_1739	ABC transporter	181.6	52.0	-3.49	0.9954323
METSMITS145B_1136	Serine hydroxymethyltransferase	720.1	334.8	-2.15	0.9953723
METSMITS145B_0497	hypothetical protein	185.6	88.8	-2.09	0.9953688
METSMITS145B_1308	Substrate binding domain of ABC-type gly	52.1	13.4	-3.90	0.9950982
METSMITS145B_0623	hypothetical protein	5695.6	1596.0	-3.57	0.9950323
METSMITS145B_1481	tRNA pseudouridine synthase D (TruD)	203.1	100.0	-2.03	0.9949935
METSMITS145B_1750	Brix domain	264.3	78.2	-3.38	0.9949691
METSMITS145B_0062	Thymidylate kinase	328.9	136.9	-2.40	0.9949192
METSMITS145B_0266	Anticodon-binding domain	554.3	247.9	-2.24	0.9948838

METSMITS145B_0028	NADP oxidoreductase coenzyme F420-depe	732.8	1959.3	2.67	0.9948097
METSMITS145B_1182	hypothetical protein	465.3	141.4	-3.29	0.9947834
METSMITS145B_0535	tRNA synthetases class I (W and Y)	385.6	142.9	-2.70	0.9946715
METSMITS145B_0164	hypothetical protein	528.8	1108.3	2.10	0.9946358
METSMITS145B_0853	RNA polymerases N / 8 kDa subunit	239.1	80.0	-2.99	0.9945515
METSMITS145B_1321	hypothetical protein	370.5	174.2	-2.13	0.9944052
METSMITS145B_1177	Zinc-binding dehydrogenase	1040.7	6448.2	6.20	0.9943331
METSMITS145B_0622	EF-1 guanine nucleotide exchange domain	641.0	252.5	-2.54	0.9942381
METSMITS145B_0660	PUA domain	447.6	203.6	-2.20	0.9940320
METSMITS145B_0349	hypothetical protein	131.9	44.8	-2.94	0.9936418
METSMITS145B_0879	Ribosomal protein L30p/L7e	559.9	206.9	-2.71	0.9935049
METSMITS145B_0063	hypothetical protein	359.4	133.5	-2.69	0.9934970
METSMITS145B_0583	hypothetical protein	1081.9	205.1	-5.27	0.9933050
METSMITS145B_1269	Type IIB DNA topoisomerase	559.3	215.7	-2.59	0.9932221
METSMITS145B_0609	Ferrous iron transport protein B	133.3	315.1	2.36	0.9928589
METSMITS145B_0897	Ribosomal protein S19	815.1	325.5	-2.50	0.9926403
METSMITS145B_1219	Tetratricopeptide repeat	12.9	43.1	3.33	0.9924422
METSMITS145B_1394	NADH-Ubiquinone/plastoquinone (complex I)	273.2	123.0	-2.22	0.9923872
METSMITS145B_0022	Aminotransferase class I and II	71.4	24.3	-2.94	0.9922219
METSMITS145B_0831	hypothetical protein	203.4	95.8	-2.12	0.9921651
METSMITS145B_0578	4Fe-4S binding domain	869.8	397.6	-2.19	0.9921494
METSMITS145B_0735	Ribosomal protein L31e	1124.2	429.2	-2.62	0.9918193
METSMITS145B_0893	Translation initiation factor SUI1	280.1	74.6	-3.76	0.9917441
METSMITS145B_0034	adhesin-like protein (Cluster 18)	245.6	96.6	-2.54	0.9917377
METSMITS145B_1114	hypothetical protein	315.2	104.0	-3.03	0.9916374
METSMITS145B_1432	hypothetical protein	455.9	128.2	-3.56	0.9916028
METSMITS145B_0322	ABC transporter	328.2	138.0	-2.38	0.9913972
METSMITS145B_0447	Toprim domain	498.8	245.5	-2.03	0.9909964
METSMITS145B_1825	hypothetical protein	71.8	27.3	-2.62	0.9909923
METSMITS145B_0749	hypothetical protein	475.5	158.5	-3.00	0.9907966
METSMITS145B_1356	YLP motif	662.5	255.6	-2.59	0.9907192
METSMITS145B_1002	hypothetical protein	2145.8	5907.5	2.75	0.9903513
METSMITS145B_0524	hypothetical protein	92.0	200.6	2.18	0.9901892
METSMITS145B_0246	hypothetical protein	68.8	27.3	-2.52	0.9899688
METSMITS145B_0891	Ribosomal protein S17	499.7	207.6	-2.41	0.9898280
METSMITS145B_0885	Ribosomal protein S8	550.5	208.5	-2.64	0.9898115
METSMITS145B_0881	Ribosomal L18p/L5e family	841.1	349.9	-2.40	0.9896522
METSMITS145B_1752	Prefoldin subunit	604.8	280.1	-2.16	0.9895336
METSMITS145B_0850	4Fe-4S binding domain	252.8	46.8	-5.40	0.9895237
METSMITS145B_1678	hypothetical protein	82.5	40.2	-2.06	0.9891309
METSMITS145B_1218	RNA polymerase Rpb5, C-terminal domain	463.5	212.8	-2.18	0.9886263
METSMITS145B_0537	hypothetical protein	504.4	1372.1	2.72	0.9884802
METSMITS145B_0230	KH domain	244.8	95.3	-2.57	0.9884130
METSMITS145B_0564	Cytidylyltransferase	48.4	21.3	-2.27	0.9882508
METSMITS145B_1517	Ammonium Transporter Family	1021.8	93.0	-10.98	0.9879685
METSMITS145B_0221	Ammonium Transporter Family	1021.8	93.0	-10.98	0.9879685
METSMITS145B_1029	Sodium:neurotransmitter symporter family	94.0	31.1	-3.03	0.9878348
METSMITS145B_0804	DNA polymerase family B	158.2	352.3	2.23	0.9876538
METSMITS145B_0599	adhesin-like protein (Cluster 1267)	191.4	90.4	-2.12	0.9874649
METSMITS145B_0890	Ribosomal protein L14p/L23e	536.9	185.0	-2.90	0.9874616
METSMITS145B_1686	hypothetical protein	342.5	88.7	-3.86	0.9873751
METSMITS145B_1039	Methyltransferase domain	288.9	130.6	-2.21	0.9873505
METSMITS145B_0143	MatE	71.6	23.9	-3.00	0.9869973
METSMITS145B_0874	Integral membrane protein DUF106	330.0	113.0	-2.92	0.9869525
METSMITS145B_1547	Peptidase family M48	75.1	162.2	2.16	0.9866127
METSMITS145B_0036	3-dehydroquinate synthase (EC 4.6.1.3)	400.8	199.6	-2.01	0.9865625
METSMITS145B_0367	Protein of unknown function (DUF509)	221.5	100.5	-2.20	0.9865053
METSMITS145B_0878	Ribosomal protein L15	516.5	237.4	-2.18	0.9864541
METSMITS145B_0600	Protein of unknown function DUF70	109.3	31.0	-3.53	0.9864365
METSMITS145B_1537	Peptidase family M48	65.8	140.4	2.13	0.9862686
METSMITS145B_0536	hypothetical protein	691.9	316.8	-2.18	0.9862085
METSMITS145B_0196	Histone-like transcription factor (CBF/	116600.0	233350.9	2.00	0.9860558
METSMITS145B_1804	hypothetical protein	6.5	14.7	2.26	0.9858840
METSMITS145B_0838	Cupin domain	989.7	2032.4	2.05	0.9858283
METSMITS145B_0707	hypothetical protein	54.4	19.0	-2.86	0.9848908
METSMITS145B_1834	hypothetical protein	378.5	124.5	-3.04	0.9847138
METSMITS145B_1666	hypothetical protein	69.3	27.5	-2.52	0.9838153
METSMITS145B_0827	hypothetical protein	265.7	111.7	-2.38	0.9837558
METSMITS145B_0994	hypothetical protein	296.2	121.7	-2.43	0.9836185
METSMITS145B_0826	hypothetical protein	76.9	23.6	-3.26	0.9830315

METSMITS145B_0538	B12 binding domain	362.5	1019.1	2.81	0.9828607
METSMITS145B_0357	hypothetical protein	111.7	245.7	2.20	0.9821259
METSMITS145B_1424	Phosphoribosyl-ATP pyrophosphohydrolase	447.5	208.0	-2.15	0.9820992
METSMITS145B_1281	Shikimate / quinate 5-dehydrogenase	126.8	58.6	-2.16	0.9820247
METSMITS145B_0632	yrdC domain	106.0	44.9	-2.36	0.9811305
METSMITS145B_0899	Ribosomal protein L23	524.8	190.1	-2.76	0.9809105
METSMITS145B_0396	hypothetical protein	4585.6	332.4	-13.80	0.9804666
METSMITS145B_0105	hypothetical protein	155.8	29.4	-5.30	0.9802722
METSMITS145B_1663	ABC transporter	75.6	163.2	2.16	0.9788921
METSMITS145B_0963	hypothetical protein	68.3	230.8	3.38	0.9781729
METSMITS145B_1346	Sodium/calcium exchanger protein	148.2	66.5	-2.23	0.9779362
METSMITS145B_0300	hypothetical protein	90.9	30.7	-2.96	0.9775189
METSMITS145B_1030	Sodium:neurotransmitter symporter family	156.3	68.8	-2.27	0.9774789
METSMITS145B_1824	4Fe-4S iron sulfur cluster binding proteins	141.7	55.0	-2.58	0.9772790
METSMITS145B_0464	hypothetical protein	255.8	565.0	2.21	0.9770922
METSMITS145B_1670	hypothetical protein	26.1	11.1	-2.34	0.9768899
METSMITS145B_0434	hypothetical protein	24.5	62.1	2.54	0.9766617
METSMITS145B_0276	hypothetical protein	158.0	672.7	4.26	0.9765445
METSMITS145B_0381	NikR C terminal nickel binding domain	2712.3	1008.3	-2.69	0.9760396
METSMITS145B_0292	hypothetical protein	147.9	465.8	3.15	0.9759447
METSMITS145B_0064	hypothetical protein	976.2	486.1	-2.01	0.9759221
METSMITS145B_1577	MatE	53.0	26.2	-2.02	0.9738624
METSMITS145B_0248	hypothetical protein	126.4	41.7	-3.03	0.9730076
METSMITS145B_0450	hypothetical protein	137.1	51.2	-2.68	0.9726703
METSMITS145B_0700	Protein of unknown function DUF101	56.9	161.7	2.84	0.9724525
METSMITS145B_0305	hypothetical protein	52.7	26.2	-2.01	0.9721916
METSMITS145B_0765	Uncharacterized protein conserved in archaea	303.7	640.0	2.11	0.9714954
METSMITS145B_0662	ACT domain	150.3	45.5	-3.30	0.9704155
METSMITS96A_1127	Uncharacterized protein conserved in archaea	459.6	176.9	-2.60	1.0000
METSMITS96A_0937	Elongation factor Tu GTP binding domain	439.8	988.1	2.25	1.0000
METSMITS96A_1571	CoA binding domain	167.2	844.3	5.05	0.9999
METSMITS96A_0605	4Fe-4S binding domain	50.1	287.7	5.75	0.9999
METSMITS96A_0778	Ribosomal protein L3	331.4	736.7	2.22	0.9998
METSMITS96A_0026	Acetyltransferase (GNAT) family	2152.7	954.9	-2.25	0.9998
METSMITS96A_0075	adhesin-like protein (Cluster 18)	150.1	72.3	-2.08	0.9998
METSMITS96A_1071	AsnC family	1146.2	443.9	-2.58	0.9998
METSMITS96A_0603	Pyruvate flavodoxin/ferredoxin oxidor	45.1	251.4	5.57	0.9997
METSMITS96A_1455	adhesin-like protein (Cluster 37)	84.5	30.0	-2.82	0.9996
METSMITS96A_0777	Ribosomal protein L4/L1 family	155.9	464.0	2.98	0.9996
METSMITS96A_0593	hypothetical protein	974.4	479.5	-2.03	0.9993
METSMITS96A_1126	Major intrinsic protein	281.8	107.5	-2.62	0.9993
METSMITS96A_0604	Thiamine pyrophosphate enzyme, C-termina	53.6	335.9	6.26	0.9993
METSMITS96A_0948	RNA polymerase Rpb1, domain 2	128.9	258.2	2.00	0.9993
METSMITS96A_0403	Helix-turn-helix	1264.1	597.2	-2.12	0.9989
METSMITS96A_0626	Peptide methionine sulfoxide reductase	523.4	230.3	-2.27	0.9987
METSMITS96A_1014	hypothetical protein	4182.6	1728.2	-2.42	0.9986
METSMITS96A_1260	hypothetical protein	125.6	46.8	-2.68	0.9985
METSMITS96A_0601	Pyruvate ferredoxin/flavodoxin oxidoreductas	336.6	1075.0	3.19	0.9984
METSMITS96A_1456	Chlamydia polymorphic membrane protein (Chl	243.5	118.7	-2.05	0.9984
METSMITS96A_0239	TCP-1/cpn60 chaperonin family	207.4	422.6	2.04	0.9983
METSMITS96A_0947	RNA polymerase Rpb1, domain 5	656.1	1373.2	2.09	0.9982
METSMITS96A_0913	Glutamate/Leucine/Phenylalanine/Valin	347.6	782.8	2.25	0.9980
METSMITS96A_0926	Glutamate/Leucine/Phenylalanine/Valin	347.6	782.8	2.25	0.9980
METSMITS96A_1542	Sugar-specific transcriptional regulator Trm	228.7	53.0	-4.31	0.9977
METSMITS96A_0732	hypothetical protein	1541.1	634.2	-2.43	0.9975
METSMITS96A_1524	S4 domain	326.0	659.6	2.02	0.9974
METSMITS96A_1119	adhesin-like protein (Cluster 226)	673.8	179.8	-3.75	0.9973
METSMITS96A_0349	Ribosomal L15	1139.3	2798.6	2.46	0.9969
METSMITS96A_1374	Chlamydia polymorphic membrane protein (Chl	149.9	73.9	-2.03	0.9961
METSMITS96A_0373	Predicted membrane protein (DUF2107)	734.6	341.7	-2.15	0.9957
METSMITS96A_1733	Uncharacterized conserved protein (DUF2304)	2328.9	742.3	-3.14	0.9955
METSMITS96A_1793	hypothetical protein	1206.1	594.3	-2.03	0.9954
METSMITS96A_1758	hypothetical protein	1312.0	595.6	-2.20	0.9954
METSMITS96A_1532	Enolase, C-terminal TIM barrel domain	47.7	95.7	2.01	0.9952
METSMITS96A_1849	hypothetical protein	269.7	126.4	-2.13	0.9951
METSMITS96A_1403	4Fe-4S binding domain	2597.7	5467.9	2.10	0.9950
METSMITS96A_0945	KH domain	790.6	1791.8	2.27	0.9950
METSMITS96A_0935	Ribosomal protein S10p/S20e	302.4	868.1	2.87	0.9937
METSMITS96A_1519	Transposase DDE domain	49.5	21.1	-2.34	0.9934

METSMITS96A_0833	hypothetical protein	394.4	133.6	-2.95	0.9929
METSMITS96A_0720	hypothetical protein	434.5	161.3	-2.69	0.9926
METSMITS96A_0087	hypothetical protein	43.7	19.0	-2.31	0.9917
METSMITS96A_0304	Uncharacterised protein family UPF0047	388.8	95.0	-4.09	0.9911
METSMITS96A_0859	Chlamydia polymorphic membrane protein (Chl	209.8	98.3	-2.13	0.9910
METSMITS96A_0973	hypothetical protein	91.0	36.1	-2.52	0.9900
METSMITS96A_0974	hypothetical protein	661.1	284.5	-2.32	0.9895
METSMITS96A_0347	Archaeal ATPase	141.9	68.1	-2.08	0.9893
METSMITS96A_0272	hypothetical protein	439.7	216.2	-2.03	0.9893
METSMITS96A_0005	hypothetical protein	171.4	76.9	-2.23	0.9884
METSMITS96A_0664	Ribosomal protein L11, N-terminal dom	1444.2	3049.9	2.11	0.9877
METSMITS96A_1347	hypothetical protein	496.1	171.7	-2.89	0.9873
METSMITS96A_0501	Helix-turn-helix	683.8	270.7	-2.53	0.9867
METSMITS96A_0919	E1-E2 ATPase	171.4	82.3	-2.08	0.9862
METSMITS96A_1650	Glycosyl transferase family 2	105.8	46.5	-2.28	0.9861
METSMITS96A_0602	4Fe-4S binding domain	57.2	250.6	4.38	0.9859
METSMITS96A_1529	Ribosomal protein S9/S16	316.2	679.4	2.15	0.9852
METSMITS96A_0050	hypothetical protein	576.9	280.1	-2.06	0.9846
METSMITS96A_1591	hypothetical protein	59.5	169.9	2.86	0.9832
METSMITS96A_0093	hypothetical protein	349.3	168.8	-2.07	0.9826
METSMITS96A_0019	Exonuclease VII small subunit	68.7	24.2	-2.84	0.9824
METSMITS96A_1783	Transcription factor S-II (TFIIS)	633.2	302.7	-2.09	0.9820
METSMITS96A_0189	hypothetical protein	149.7	73.4	-2.04	0.9816
METSMITS96A_1107	Domain related to MnhB subunit of Na+/H+ ant	21.7	51.5	2.37	0.9810
METSMITS96A_0885	HxlR-like helix-turn-helix	332.2	126.0	-2.64	0.9809
METSMITS96A_1237	6-O-methylguanine DNA methyltransferase	221.2	99.0	-2.23	0.9805
METSMITS96A_0253	hypothetical protein	412.8	126.2	-3.27	0.9799
METSMITS96A_1566	Histidine kinase-, DNA gyrase B-, and HSP90	75.0	34.0	-2.21	0.9781
METSMITS96A_0852	GHMP kinases N terminal domain	45.2	91.8	2.03	0.9775
METSMITS96A_0746	RNAse P Rpr2/Rpp21/SNM1 subunit domain	707.8	1480.0	2.09	0.9769
METSMITS96A_1611	hypothetical protein	57.6	25.1	-2.30	0.9765
METSMITS96A_1628	hypothetical protein	44.4	21.8	-2.04	0.9764
METSMITS96A_1064	Domain of unknown function (DUF1922)	4893.8	2191.7	-2.23	0.9764
METSMITS96A_0765	Ribosomal family S4e	101.0	214.3	2.12	0.9750
METSMITS96A_0116	NADP oxidoreductase coenzyme F420-depe	38.1	17.3	-2.20	0.9744
METSMITS96A_1102	hypothetical protein	25.6	57.4	2.25	0.9735
METSMITS96A_1559	Coenzyme F420 hydrogenase/dehydrogenase,	37.4	16.6	-2.26	0.9733
METSMITS96A_1822	YLP motif	34.3	69.0	2.01	0.9715
METSMITS96A_0301	hypothetical protein	844.2	361.9	-2.33	0.9714
METSMITS96A_0061	hypothetical protein	2255.9	803.8	-2.81	0.9709

Chapter 4

Future Directions

Discussion and Future Directions

This thesis addresses a number of basic questions: what forces determine if *Methanobrevibacter smithii* is part of a human gut microbiota; what other species in the gut microbial community co-occur with it; how does its genome evolve within an individual and between individuals within and between families.

Exploring the possible genetic determinants of methanogen carriage

Our studies in twins have revealed evidence for a host genetic determinant for carriage and levels of methanogens. We found that MZ co-twins had a statistically significant concordance in their colonization with methanogens and that those who were colonized had similar levels of methanogens. However, in DZ twins there was no significant concordance. The host factor(s) that contribute to colonization with methanogens have not been identified. Numerous possible mechanisms/factors could mediate host selection of the intestinal microbiota. These include direct selection for methanogens (e.g., by producing a receptor or receptors that recognize an expressed repertoire of *M. smithii* adhesin-like proteins (such a receptor has been observed in Crohn's disease with *E. coli*; Barnich *et al*., 2007), indirection selection (e.g., by supporting the growth of a bacterial syntrophic partner, or by selecting against a competitor, or by having an increased gut transit time phenotype to allow the slower-growing methanogens to more readily persist), or a combination of both direct and indirection selection mechanisms (**Fig. 1**). Exclusion may be achieved by direct inhibition of methanogen growth by factors produced by the host's innate or adaptive immune system, or indirect inhibition (e.g., exclusion of a required bacterial symbiotic partner from the microbiota).

Previous work in non-human primates found uniform colonization of certain phylogenic groups and uniformly absent colonization in other groups: the pattern did not follow any identifiable morphological or physiological features of their digestive tracts nor any other identifiable features of their host biology. Importantly, it did not track with behavioral

characteristics. This observation provides additional support for a host genetic determinant of the presence or absence of gut methanogens (Hackstein *et al*., 1995). Another study examining the distribution of methanogens within the guts of 253 vertebrate species found "methanogenic branches" of the host phylogenetic tree (i.e., branches containing ruminants – bovidae, cervidae, giraffidae) and non-methane producing branches (those containing felidae, canidae and ursidae (**Fig. 2**). The methane-producing groups could not be distinguished from non-methane producers based on their diets, or features of their gut structure/ physiology. The authors concluded that presence of methanogens is under 'phylogenetic' control, rather than dietary regulation (Hackstein *et al*., 1996).Members of the lab have validated procedures for transplanting entire human fecal microbial communities into germ-free mouse recipients who are then fed diets that do or do not resemble those of the human donors (Turnbaugh *et al*., 2009b). We have found that a remarkable proportion of human fecal microbial diversity can be transferred in this fashion even if the donor specimen had been frozen at -80^{o}C for 1–2 years (all bacterial phyla, up to 90% of class-level and genus-level taxa, and 60-90% of species level-phylotypes in donor samples are identifiable in recipient mice using 16S rRNA-based pyrosequencing). Once engrafted, the transplanted human microbial communities are remarkably stable: they can be reliably transmitted across generations of mice, and exhibit reproducible bio-geographical features along the length of the recipient mouse gut (Turnbaugh *et al*., 2009b). Efficient intergenerational transfer of transplanted human fecal microbiota allows the microbiota and the host's innate/adaptive immune system to co-evolve beginning at birth in 'second generation' mice. Humanized gnotobiotic mice can be used for proof-of-mechanism studies that cannot be readily conducted in humans where potentially confounding variables, including host genotype and diet, are very difficult to control. Transplantation of fecal microbiota from human donors who do or do not harbor *M. smithii* into mice with engineered mutations, such as those involving the innate or adaptive immune system, represents one way of testing the role of specified host factors in determining colonization with this (or other methanogens). An al-

ternative strategy would be to colonize germ-free mice with defined collections of cultured members of the human gut microbiota, with and without various combinations of sulfate reducers and acetogens and/or with other taxa shown in Chapter 3 to co-occur with *M. smithii*: colonization of these mice with defined communities, with or without *M. smithii*, and monitoring the transcriptional responses of the gut could identify *M. smithii*-associated responses. These functional genomics datasets could be used to nominate responsive host genes as potential mediators/determinants of *M. smithii* colonization: the contributions of these genes could then be tested further in genetically manipulated gnotobiotic mice and/or through genome wide association studies (GWAS).Both *M. smithii* and *M. stadtmanae* are extremely oxygen sensitive (as are most methanogens), and do not persist in the environment (except in sewage digester plants, and other anaerobic sites contaminated with fecal waste). We do not know how methanogens colonize a new host: opportunities for exposure may be affected by route of delivery including the history of exposure to other children or animals (e.g., pets, livestock; Fujimura *et al*., 2010), the nutrient content of breast milk, the pattern of assembly of the microbiota during postnatal life, or the pattern of introduction of foods and the nature of these foods. One study examined different types of food for the presence of methanogens, and while they found methanogens in vegetables, meat, fish and cheese, they did not find *M. smithii* or *M. stadtmanae*, and instead identified methanogens that normally live in soil (Brusa *et al*., 1998). Others have argued that exposure is probably not the limiting factor in determining whether a host harbors methanogens, because they believe that most humans (and other animals) are colonized with at least a very low level of methanogens (Miller and Wolin, 1982; Dridi *et al*., 2009).

Studies of early colonization of humans are confounded by methodologic challenges: methanogens are difficult to culture; methane breath tests have limited applicability and reliability in this age range; there is a need to develop PCR primers that are specific for archaeal 16S rRNA genes (i.e., that do not primarily hybridize to bacterial small subunit rRNA gene sequences); there have been very few shotgun sequencing datasets of the fecal

microbiomes generated in studies of the community assembly during the first two years of life. Methanogens have been identified in 16S rRNA datasets generated from children under the age of one (Palmer *et al*., 2007, reviewed in Vael *et al*., 2009). One study followed a single infant from birth to 2.5 years, and found Archaeal sequences in each of their 12 datasets obtained from shotgun sequencing of fecal DNA, including a sample of meconium (Koenig *et al*., 2010). One report suggested that a smaller percentage of children were positive for methanogens when compared to an adult population (Stewart *et al*., 2006).

Members of our lab are interested in the role of the microbiome in risk for childhood malnutrition, and the effect of diet on the development of the microbiota. These metagenomic analyses of fecal samples obtained from large numbers of twins enrolled in two birth cohort studies (one in Malawi and the other in Bangladesh) have the advantage of serial collection of fecal samples from each individual, plus detailed information about their host phenotypes, exposures (i.e., breast feeding history, foods, antibiotics, vaccinations, animals, etc). Their fecal DNA is being subjected to shotgun pyrsosequencing, which should provide an unbiased view of the representation of Archaeal, as well as bacterial and eukaryotic genomes.Experiments involving systematic variation in diet components (macronutrients, micronutrients and food additives) are ongoing in the lab to determine the effects of these components on microbial community composition, patterns of microbiome gene expression, and community metabolism: these experiments involve gnotobiotic mice colonized with an intact human fecal microbiota as well as mice colonized with defined communities composed of sequenced members of the human gut microbiota, with or without hydrogen-consuming acetogens, SRB, methanogens. These experiments could examine the impact of diet on community assembly, and on the representation of different classes of hydrogen consumers. The relationships between diet and community composition and expressed functions can be documented using a variety of methods, including those centered around the use of massively parallel DNA sequencers [e.g., shotgun sequencing of fecal DNA to quantify community membership (Faith *et al*., 2010; Goodman

et al., 2009) or microbial RNA-Seq (Rey *et al*., 2010)], and RNA-Seq informed targeted metabolomic studies.

H_2-dependent sulfate reduction and methanogenesis are thermodynamically more favorable than acetogenesis. In most anaerobic environments, given a high concentration of sulfate, sulfate-reducers dominate. In the gut, sulfated glycans are the main source of sulfate for SRB. This sulfate is accessible to SRB when saccharolytic bacteria degrade host mucins. In the presence of a normal or high protein diet, we predict increased fermentation of protein and lower consumption of host glycans from mucin will result in lower levels of sulfate accessible to SRB. In this scenario, it is likely that *M. smithii* will dominate in a host. *Blautia hydrogenotrophicus* may also thrive with high protein, mainly because of its capability to ferment amino acids. However, as most aceogens, *B. hydrogenotrophicus* is a relatively poor H_2 consumer compared to a sulfate reducing bacterium (Cord-Ruwisch, *et al*., 1988, Leclerc, *et al*., 1997). Importantly, these gnotobiotic mouse experiments involving defined communities of human gut symbionts will create a data- and knowledge-base that will impact other experiments: adding *M. smithii* to a defined bacterial community, we can examine the inter-relationships between methanogens, acetogens and sulfate reducers, and between hydrogen-consumers and other constituents in the model human gut microbiota, as a function of all of diet perturbations.

Methanogens may be more likely to be present in people with slower gut transit time (Attaluri *et al*., 2009), including those with constipation-predominant irritable bowel syndrome (IBS) (Pimentel *et al*., 2003, Chatterjee *et al*., 2007). Methanogens are also more likely to be present in patients with diverticulosis (Weaver *et al*., 1986). Methane-production has been observed more often in women than men (Pitt *et al*., 1980), which may be related to the differences in transit time between men and women (Stephen *et al*., 1986; McKay *et al*., 1985). In addition, methane in the gut may be a cause of slowed transit time (Pimentel *et al*., 2006). Slower transit time may allow slower-growing methanogens to avoid washout from the gut.Gut motility is regulated by various host signals, including

ghrelin and motilin (which augment motility): levels of these enteroendocrine cell products are modulated by a variety of factors, including signals that may be derived from microbes (Hellström *et al*., 2009). One study noted that circulating levels of ghrelin are significantly higher in lean compared to obese individuals: a study that also found differences in their gut microbial community composition (Tiihonen *et al*., 2010). Genetic factors influencing levels of ghrelin and motilin could be responsible for changes in gut motility that in turn allow methanogens to persist (lower ghrelin, slower transit time).

Expanding our understanding of co-occurrence

Many of the OTUs identified that positively co-occur with methanogens are part of undefined lineages with few or no sequenced representatives. In an effort to understand the interactions (if any) between these co-occurring organisms and methanogens, representative cultured isolates need to be identified and their genomes sequenced. For example, several of the significantly co-occurring OTUs are part of a gut-clone group within the Firmicutes that has no sequenced representatives. Ongoing anaerobic culturing initiatives in the lab have been able to capture >50% of the species level bacterial diversity present in a single donor's fecal microbiota (Goodman *et al.*, in preparation): the cultured taxa can be arrayed in multiwell plates so that specific members can be retrieved for sequencing or for introduction into gnotobiotic mice. When isolating the *M. smithii* strains characterized in Chapter 3, I used antibiotics to inhibit the growth of contaminating bacteria. However, adding *M. smithii* to fecal microbiota prior to culturing, or systematically characterizing the interactions of (i) cultured bacteria taxa (identified from metagenomic studies as co-occurring with *M. smithii* in human gut microbiota) and (ii) *M. smithii* using for example multi-well plates or microfluidic devices should be very informative.

Once the genomes of the co-occurring organisms are defined, they can be used in metabolic reconstructions, and models created to identify potential metabolites or substrates that are being shared between symbionts (**Fig. 3**). Our collaborator, Catherine Lozupone has used an approach first described by Borenstein (2008) to identify potential points

of metabolic interaction between microbes. Based on a metabolic reconstruction analysis, networks made up of reactions and compounds can be predicted for each organism (reactions are predicted based on enzymes encoded in the genome, compounds are either substrates or end-products of these reactions). The network is built so that nodes (symbols) are compounds, and edges (arrows) are reactions. Based on this information, strongly connected components (SCCs) are defined as either "seeds" which have no incoming edge, or "products" which have no outgoing edge. Seeds may be metabolites that are obtained from outside the cell, either from a symbiotic microbe or from the host or diet. Products may be exported from to a nearby microbe, the host or the lumen of the gut. Products from one microbe may then become seeds for another. Competitive indices are then defined as the number of shared seeds divided by the number of total seeds in both microbes' networks. If two organisms' networks contain many of the same seeds, they are more likely to compete with each other (both need the same substrates), and thus would have a higher competitive index. If a pair of microbes have no overlapping seeds, they would not be likely to compete. If one has a set of compounds as seeds, and the other as products, they would have a lower competitive index, and a higher cooperative index. Ongoing work has shown that competitive indices for gut microbes are related to phylogeny. Existing genome sequences from 101 human gut microbiota members were used to create networks and calculate competitive indices. These pairwise competitive indices were then compared to distances based on 16S rRNA based taxonomy using the Ribosomal Database Project (Cole *et al*., 2009). Organisms that are more related phylogenetically tend to have higher competitive indices, which suggests that organisms share both an evolutionary lineage and a similar metabolic network. Additionally, microbes with larger genome sizes tended to have higher cooperative indices, indicating that they provide more metabolites to organisms with smaller genomes.

Together these types of analyses, plus the approaches described in Chapters 2 and 3 of this thesis should enhance our understanding of what niche is occupied by *M smithii* in the human gut microbiota and how its representation may be manipulated to enhance digestive and host health.

References

Attaluri A, *et al*., (2010) Methanogenic flora is associated with altered colonic transit but not stool characteristics in constipation without IBS. *Am J Gastroenterol* 105: 1407-1411.

Barnich N, *et al*., (2007) CEACAM6 acts as a receptor for adherent-invasive E. coli, supporting ileal mucosa colonization in Crohn disease. *J Clin Invest* 117(6): 1566–1574.Borenstein E, *et al*., (2008) Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc Natl Acad Sci U S A* 105: 14482-14487.

Brusa T, *et al*., (1998) Methanogenic bacteria: presence in foodstuffs. *Journal of Basic Microbiology* 38: 79-84

Chatterjee S, *et al*., (2007) The degree of breath methane production in IBS correlates with the severity of constipation. *Am J Gastroenterol* 102: 837-841.

Cole JR, *et al*., (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37: D141-145.

Cord-Ruwisch R, *et al*., (1988) The capacity of hydrogenotrophic anaerobic bacteria to compete for traces of hydrogen depends on the redox potential of the terminal electron acceptor. Arch Microbiol 149, 350-357

Dridi B, *et al*., (2009) High prevalence of Methanobrevibacter smithii and Methanosphaera stadtmanae detected in the human gut using an improved DNA detection protocol. *PLoS One* 4: e7063.

Faith JJ, *et al*., (2010) Creating and characterizing communities of human gut microbes in gnotobiotic mice. *ISME J* 4: 1094-1098.

Fujimura KE, *et al*., (2010) Role of the gut microbiota in defining human health. *Expert Rev Anti Infect Ther* 8: 435-454.

Goodman AL, Gordon JI (2010) Our unindicted coconspirators: human metabolism from a microbial perspective. *Cell Metab* 12: 111-116.

Hackstein, JHP, *et al*., (1995) Intestinal Methanogenesis in Primates - a Genetic and Evolutionary Approach. *Dtsch. tierärztl. Wschr*. 102(4), 143-178.

Koenig JE, *et al*., (2010) Microbes and Health Sackler Colloquium: Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A*. doi 10.1073/pnas.100008110

Hackstein, JHP and TA van Alen (1996) Fecal Methanogens and Vertebrate Evolution. *Evolution* Vol 50(2) 559-572.Hellstrom PM (2009) Faces of ghrelin--research for the 21st century. *Neurogastroenterol Motil* 21: 2-5.

Leclerc M, *et al*., (1997) H2/CO2 Metabolism in Acetogenic Bacteria Isolated From the Human Colon. Anaerobe 3, 307-315

McKay LF, *et al*., (1985) Methane excretion in man--a study of breath, flatus, and faeces. *Gut* 26: 69-74.

Miller, TL and Wolin MJ (1982) Enumeration of Methanobrevibacter smithii in human feces. *Arch Micrbiol*. 131(1):14-8.

Palmer C, *et al*., (2007) Development of the human infant intestinal microbiota. *PLoS Biol* 5: e177.

Pimentel M, *et al*., (2003) Methane production during lactulose breath test is associated with gastrointestinal disease presentation. *Dig Dis Sci* 48: 86-92.

Pimentel M, *et al*., (2006) Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. *Am J Physiol Gastrointest Liver Physiol* 290: G1089-1095.

Pitt P, *et al*., (1980) Studies on breath methane: the effect of ethnic origins and lactulose. *Gut* 21: 951-954.

Rey FE, *et al*., (2010) Dissecting the in vivo metabolic potential of two human gut acetogens. *J Biol Chem* 285: 22082-22090.

Stephen AM, *et al*., (1986) The effect of age, sex and level of intake of dietary fibre from wheat on large-bowel function in thirty healthy subjects. *Br J Nutr* 56: 349-361.

Stewart, J.A., *et al*., (2005) Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J Med Microbiol* 54: 1239–1242.

Tiihonen K, *et al*., (2010) Effect of overweight on gastrointestinal microbiology and immunology: correlation with blood biomarkers. *Br J Nutr* 103: 1070-1078.

Turnbaugh PJ, *et al*., (2009a) A core gut microbiome in obese and lean twins. *Nature* 457: 480-484.

Turnbaugh PJ, *et al*., (2009b) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* 1: 6-14.

Vael C, Desager K (2009) The importance of the development of the intestinal microbiota in infancy. *Curr Opin Pediatr* 21: 794-800.

Weaver GA, *et al*., (1986) Incidence of methanogenic bacteria in a sigmoidoscopy population: an association of methanogenic bacteria and diverticulosis. *Gut* 27: 698-704.

Figure Legends

Fig. 1. Schematic of the interplay between human gut methanogens, the microbiota and host. See text for discussion

Fig. 2. Phylogenetic tree of vertebrates with a methane-producing branch highlighted in purple, and a non-methane-producing branch in green. Figure adapted from JHP Hackstein and TA van Alen, (1996) Fecal Methanogens and Vertebrate Evolution. *Evolution* Vol 50(2) 559-572.

Fig. 3. Metabolic network analysis. In the example shown, there are seeds (metabolites with no incoming edge or reaction, which would need to be supplied exogenously) and products (metabolites with no outgoing edge or reaction, which could potentially be supplied to another organism). If seeds are shared between micro-organisms, they are likely have overlapping niche space. On the other hand, if two organisms produce each other's seeds, they are likely to be co-operating (adapted from C. Lozupone).

Figures

Figure 1.

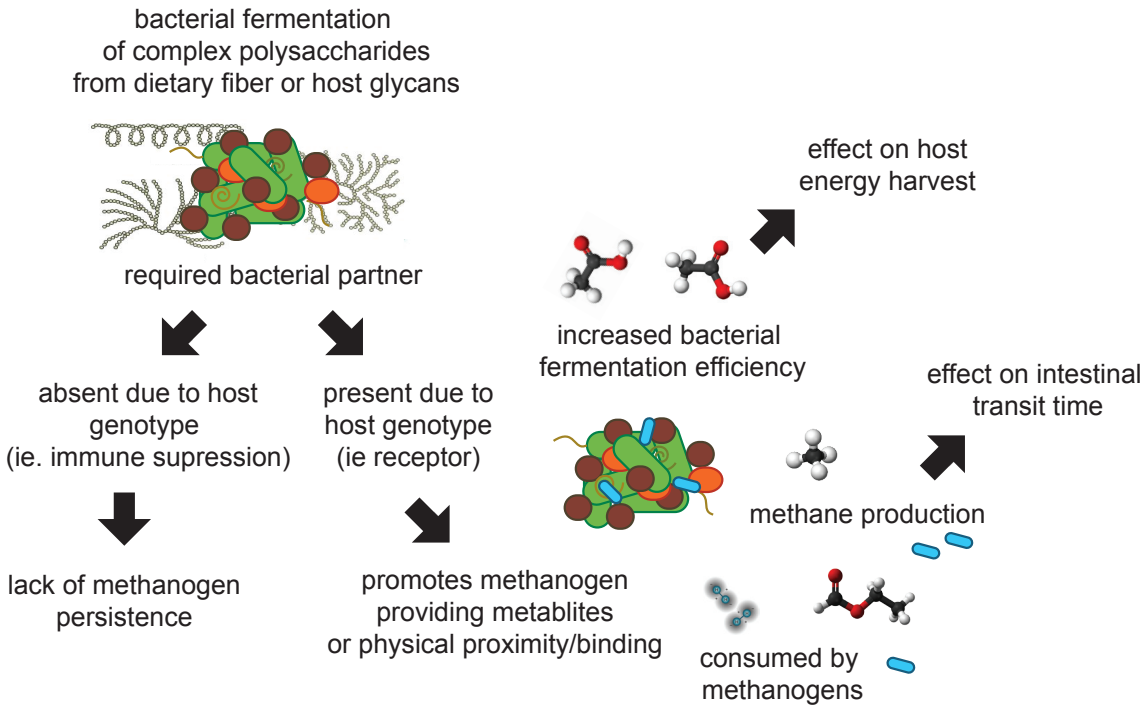
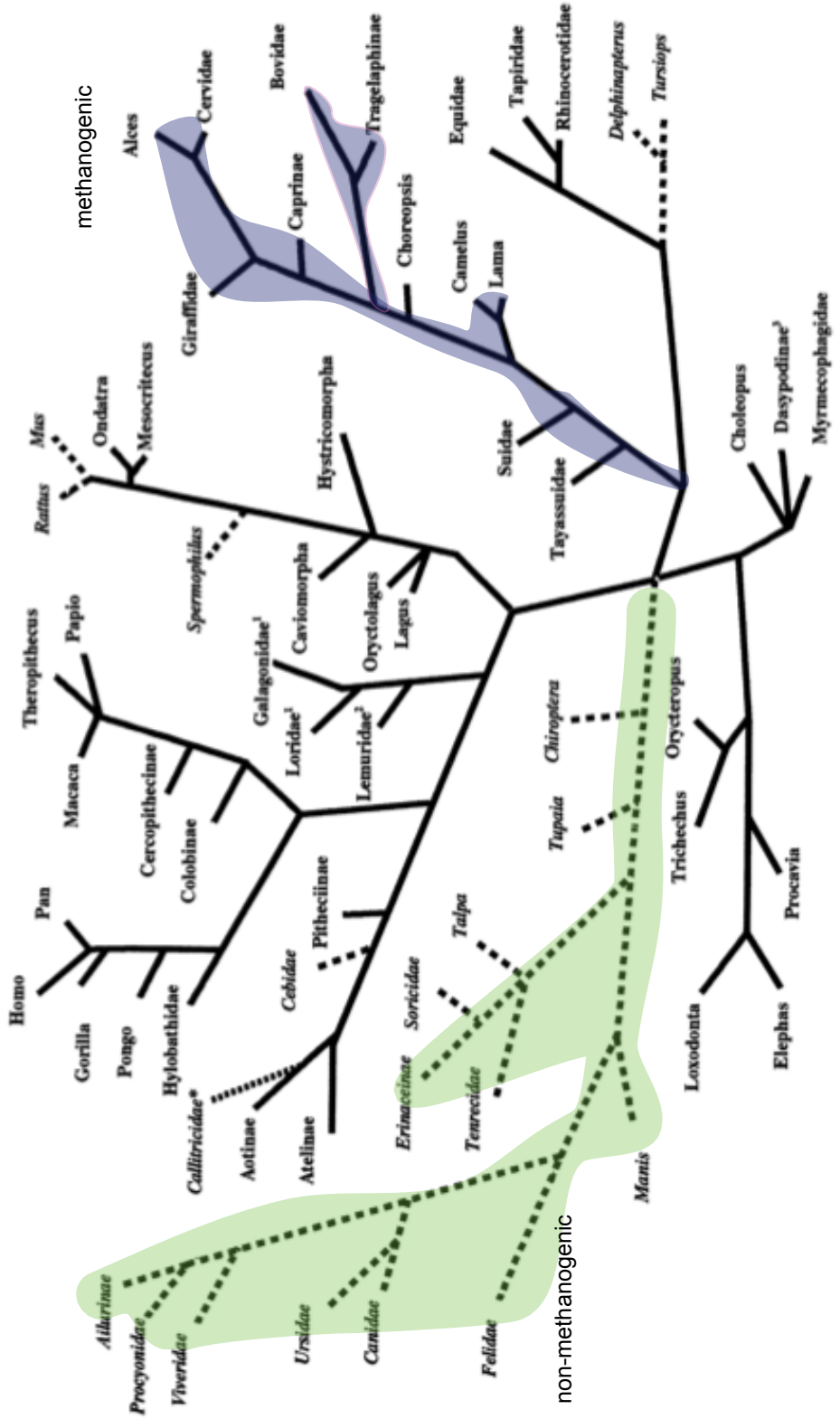


Figure 2.


methanogenic
non-methanogenic
Alces
Cervidae
Bovidae
Caprinae
Tragelaphinae
Giraffidae
Choreopsis
Camelus
Lama
Equidae
Tapiridae
Rhinocerotidae
Delphinapterus
Tursiops
Mus
Rattus
Ondatra
Mesocricetus
Hystricomorpha
Suidae
Tayassuidae
Choleopus
Dasypodinae
Myrmecophagidae
Spermophilus
Caviomorpha
Oryctolagus
Lagus
Galagonidae
Lorida
Lemuridae
Chiroptera
Oryceropus
Theropithecus
Papio
Macaca
Cercopithecinae
Colobinae
Tupaia
Trichechus
Homo
Pan
Gorilla
Pongo
Hylobathidae
Pitheciinae
Cebidae
Talpa
Soricidae
Erinaceinae
Tenrecidae
Callitrichidae
Aotinae
Atelinae
Manis
Loxodonta
Elephas
Procavia
Ailurinae
Procyonidae
Viveridae
Ursidae
Canidae
Felidae

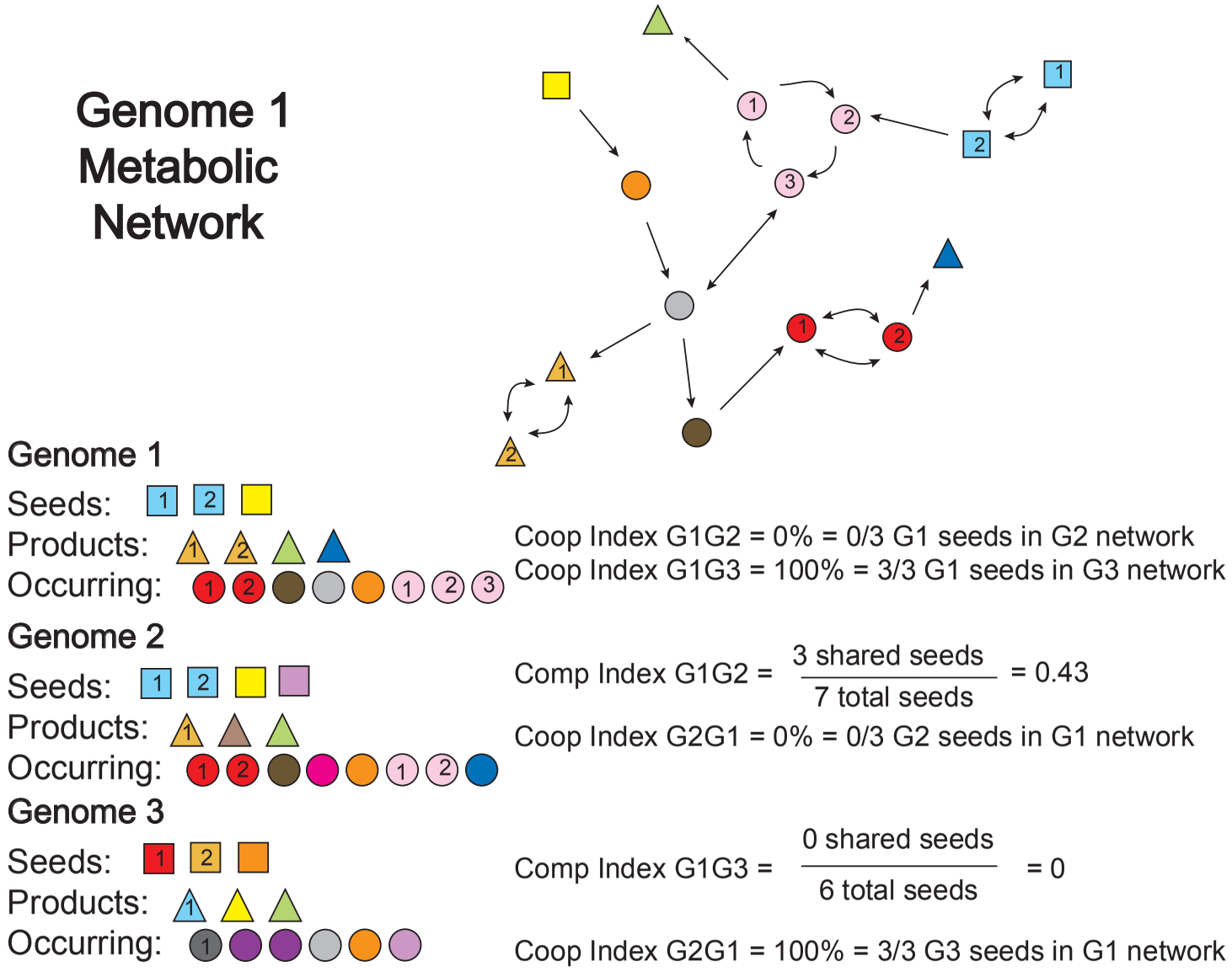

Genome 1
Metabolic
Network
Genome 1
Seeds:
Products:
Occurring:
Genome 2
Seeds:
Products:
Occurring:
Genome 3
Seeds:
Products:
Occurring:
Coop Index G1G2 = 0% = 0/3 G1 seeds in G2 network
Coop Index G1G3 = 100% = 3/3 G1 seeds in G3 network
Comp Index G1G2 = 3 shared seeds / 7 total seeds = 0.43
Coop Index G2G1 = 0% = 0/3 G2 seeds in G1 network
Comp Index G1G3 = 0 shared seeds / 6 total seeds = 0
Coop Index G2G1 = 100% = 3/3 G3 seeds in G1 network

Figure 3.

Appendix

Amanda L. Lewis, Nolan Desa, Elizabeth E. Hansen, Yuriy A. Knirel, Jeffrey I. Gordon, Pascal Gagneux, Victor Nizet, and Ajit Varki

Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure

Proc Natl Acad Sci USA. 2009 August 11; 106(32):13552–13557.

Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure

Amanda L. Lewis[a,b,1,2], Nolan Desa[a], Elizabeth E. Hansen[c], Yuriy A. Knirel[d], Jeffrey I. Gordon[c], Pascal Gagneux[a,e], Victor Nizet[a,b,f], and Ajit Varki[a,e,g,1]

[a]Glycobiology Research and Training Center, Departments of [b]Pediatrics, [g]Medicine, and [e]Cellular and Molecular Medicine, School of Medicine, and [f]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093; [d]N.D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninsky Prospekt 47, 11991 Moscow, Russia; and [c]Center for Genome Sciences, Washington University, St. Louis, MO 63108

Edited by Sen-itiroh Hakomori, Pacific Northwest Diabetes Research Institute, Seattle, WA, and approved June 19, 2009 (received for review March 9, 2009)

Sialic acids (Sias) are nonulosonic acid (NulO) sugars prominently displayed on vertebrate cells and occasionally mimicked by bacterial pathogens using homologous biosynthetic pathways. It has been suggested that Sias were an animal innovation and later emerged in pathogens by convergent evolution or horizontal gene transfer. To better illuminate the evolutionary processes underlying the phenomenon of Sia molecular mimicry, we performed phylogenomic analyses of biosynthetic pathways for Sias and related higher sugars derived from 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids. Examination of ≈1,000 sequenced microbial genomes indicated that such biosynthetic pathways are far more widely distributed than previously realized. Phylogenetic analysis, validated by targeted biochemistry, was used to predict NulO types (i.e., neuraminic, legionaminic, or pseudaminic acids) expressed by various organisms. This approach uncovered previously unreported occurrences of Sia pathways in pathogenic and symbiotic bacteria and identified at least one instance in which a human archaeal symbiont tentatively reported to express Sias in fact expressed the related pseudaminic acid structure. Evaluation of targeted phylogenies and protein domain organization revealed that the ''unique'' Sia biosynthetic pathway of animals was instead a much more ancient innovation. Pathway phylogenies suggest that bacterial pathogens may have acquired Sia expression via adaptation of pathways for legionaminic acid biosynthesis, one of at least 3 evolutionary paths for de novo Sia synthesis. Together, these data indicate that some of the long-standing paradigms in Sia biology should be reconsidered in a wider evolutionary context of the extended family of NulO sugars.

legionaminic acid | phylogeny | pseudaminic acid | neuraminic acid | biosynthetic pathway

Sialic acids (Sias) are displayed in prominent positions on vertebrate cells and are critical for such physiological processes as cellular repulsion, renal filtration, and neuronal plasticity (1, 2). Many other Sia-dependent functions occur in conjunction with Sia-binding lectins, including down-modulation of complement activity and the regulation of leukocyte activation, migration, and apoptosis (1, 3). The divergence of the superphyla protostomes and deuterostomes[†] created a dichotomy in animal Sia expression, and heralded the emergence of widespread Sia-dependent biological functions in deuterostomes. Nearly 5 decades of research have confirmed that with few exceptions, these unique 9-carbon backbone sugars are conspicuously absent from many eukaryotic lineages, including most protostomes, plants, fungi, and protists (1, 4). Sia decoration by de novo biosynthesis or via metabolic scavenging pathways has been reported in more than a dozen pathogenic bacterial species (5, 6) and also was recently described in a human gut–associated methanogenic archaeon (7). During infection, microbes displaying Sia mimicry can exploit host factor H and/or Siglec-9 to down-regulate alternative complement deposition and neutrophil bactericidal activities (8–11).

Sias are 9-carbon backbone derivatives of neuraminic (Neu) and ketodeoxynonulosonic (Kdn) acids. They are actually part of a larger family of carbohydrate structures collectively called nonulosonic acids (NulOs)[‡]. A number of NulO sugars other than Sias have been found in microbes, all of which are derivatives of 4 isomeric 5,7-diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids (12). At least 2 of these, the D-*glycero*-d-*galacto* isomer [legionaminic acid (Leg)] (13, 14) and L-*glycero*-l-*manno* isomer [pseudaminic acid (Pse)] (15, 16), have striking structural and biosynthetic similarities to Sias (Fig. 1). These commonalities among NulO pathways reflect the structural similarity of all of the NulO sugars, as well as their uniqueness compared with other monosaccharides. Similar steps in each NulO biosynthetic (NAB)[§] pathway are catalyzed by homologous enzymes, including the condensation of a 6-carbon sugar intermediate with 3-carbon phosphoenolpyruvate (3C) to generate the 9-carbon backbone NulO sugar, followed by the activation of free NulO residues using cytidine triphosphate to form cytidine monophosphate (CMP)-NulO intermediates (Fig. 1). In *Campylobacter* species and other ε-proteobacteria, Pse modifications play critical roles in flagellar assembly and, consequently, motility (16, 17), an important physiological function in aquatic environments and for association with animals. Leg modifications also have been identified on flagellar subunits, but have less well-defined functions (14, 18, 19). Both Leg and Pse also have

Author contributions: A.L.L., V.N., and A.V. designed research; A.L.L., N.D., and E.E.H. performed research; Y.A.K. and J.I.G. contributed new reagents/analytic tools; A.L.L., E.E.H., P.G., V.N., and A.V. analyzed data; and A.L.L., Y.A.K., J.I.G., P.G., V.N., and A.V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: allewis@wustl.edu or alvarki@ucsd.edu.

[2]Current address: Washington University School of Medicine, St. Louis, MO 63110.

[†]Protostomes and deuterostomes are bilaterian animals with distinct developmental programs for gut tube formation in which the first opening of the embryo, the blastopore, becomes either the mouth (protostomes, ''mouth first'') or the anus (deuterostomes, ''mouth second'').

[‡]There currently is no established nomenclature defining the 9 carbon α-keto acids as a group. Following discussions with Hans Kamerling, Roland Schauer, and Nathan Sharon, here we use the abbreviation ''NulO'' for non-2-ulosonic acids, which assumes Nul for non-2-uloses and maintains the discrimination between aldonic acids and uronic acids, such as glucuronic acid. We suggest that the use of the term ''sialic acid'' (Sia) continue to be limited to its original use in describing Neu, Kdn, and their derivatives in deuterostomes and their pathogens, and that NulO be used to encompass the entire group of 9 carbon α-keto acids.

[§]NAB pathways have been defined in various bacterial pathogens. Enzymes in each of these pathways have been given different designations, beginning with ''Neu'' for neuraminic, ''Pse'' for pseudaminic, and ''Ptm'' for posttranslational modification (see Fig. 1 for published designations). To avoid confusion, here we use ''NAB'' to refer to homologous steps in these related biosynthetic pathways, numbering the enzymatic steps as shown in Fig. 1.

This article contains supporting information online at www.pnas.org/cgi/content/full/0902431106/DCSupplemental.

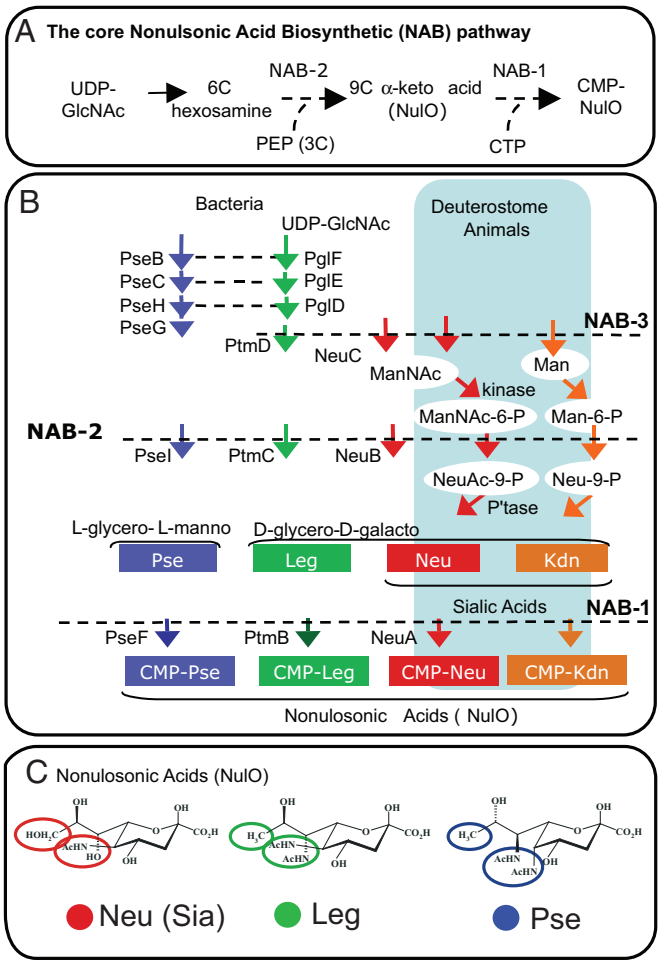


Fig. 1. Related NAB pathways synthesize chemically related sugars. NulOs include all 9-carbon backbone α-keto acid sugars. NulO sugars described to date conform to one of several core backbones that can be further modified by epimerization or modification (1, 12). Major core backbones include the Neu, Kdn, Leg, and Pse acids. (*A*), The unique and shared ''core'' features of all NAB pathways include UDP-*N*-acetylglucosamine as a starting point, condensation of a 6-carbon intermediate with phosphoenolpyruvate to yield a 9-carbon α-keto acid (NAB-2), and formation of a CMP-activated NulO intermediate (NAB-1). (*B*), Architecture and nomenclature of NulO pathways. Horizontal dashed lines denote NAB enzymes in different pathways that share a common ancestor as deduced by amino acid sequence similarity. (*C*), Chemical structures of *N*-acetyl derivatives of Neu, Leg, and Pse.

been identified as part of lipopolysaccharide (LPS) O antigens in some Gram-negative bacteria (12), where they conceivably could contribute to biofilm formation, resistance to phage predation, or animal associations. Despite the similarities of Leg and Pse to Sias, the potential roles of these sugars in host–pathogen interactions remain poorly defined, and their distribution among microbes has not yet been systematically investigated.

In the present work, we probed the existing paradigms of Sia evolution using genomic, phylogenetic, and biochemical approaches to ask whether Sias were a unique innovation of the deuterostome lineage, whether bacterial mimicry of host Sias was the result of lateral gene transfer from an animal host or convergent evolution from microbial Sia-like biosynthetic pathways, and whether the chemical structure of Sias and related sugars can be predicted from genomic sequence information.

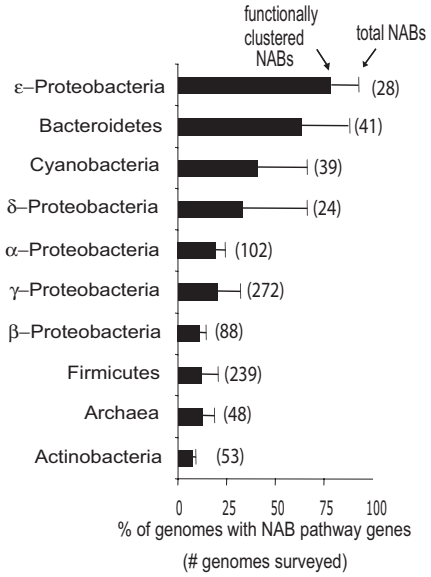


Fig. 2. Distribution of predicted NAB pathways among bacteria and archaea. NAB enzymes NAB-1 and NAB-2 were identified by BLASTp in the genomes of various bacterial phyla and in sequenced archaeons. The number of genomes in each group is given in parentheses. Bars reflect the percentage of genomes in each group with one or more physically clustered NAB gene pairs, and thin line extensions reflect total NAB homolog pairs irrespective of functional clustering.

Results and Discussion

''Functional Clustering'' Predicts a Remarkably Wide Distribution of NulO Sugar Expression Among Bacteria and Archaea. To define the distribution of biosynthetic pathways for NulO sugars in members of bacteria and archaea, nearly 1,000 sequenced microbial genomes[¶] were examined by BLAST for evidence of ''functional clusters'' (20) of NAB pathway genes. Unexpectedly, about 20% of all microbial phylogenetic types (phylotypes) sequenced to date were found to encode NAB pathway cassettes (Fig. 2).

Many species/subsets, as well as entire phyla in which NulO sugars have never been documented, were found to have NAB enzymes in their genomes, including remarkably large proportions of available Bacteroidetes (36/41), Cyanobacteria (26/39), and δ-Proteobacteria (16/24), certain pathogenic members of the order *Spirochaetales*, and nearly 19% of sequenced Archaea (9/48) (Fig. 2). NAB pathways were identified in a larger fraction of the available genomes in bacterial phyla (divisions) previously known to include Sia-decorated pathogens (i.e., γ-Proteobacteria, β-Proteobacteria, and Firmicutes). Our analysis reveals a far wider distribution of NulOs and a deeper evolutionary history of this class of sugars than originally assumed.

Phylogenetic Prediction of NulO Types Reveals an Evolutionary Context for Sias and Sia-Like Sugars. To better illuminate the evolutionary history of these 9-carbon backbone NulO sugars and predict their distribution and structure, we performed phylogenetic analysis of the most highly conserved enzyme in the pathway (NAB-2) and overlaid this tree with published biochemical data [Fig. 3*A*; summarized with strain and sequence identifiers in supporting information (SI) Table S1]. NAB-2 condenses a 6-carbon intermediate with the 3-carbon molecular phosphoenolpyruvate to generate NulOs of different types (Fig.

[¶]These sequences had been previously deposited in the Genbank database. For a list of accession numbers, see Table S1. Annotations have been updated in The SEED, an annotation/analysis tool provided by the Fellowship for Interpretation of Genomes.

MICROBIOLOGY

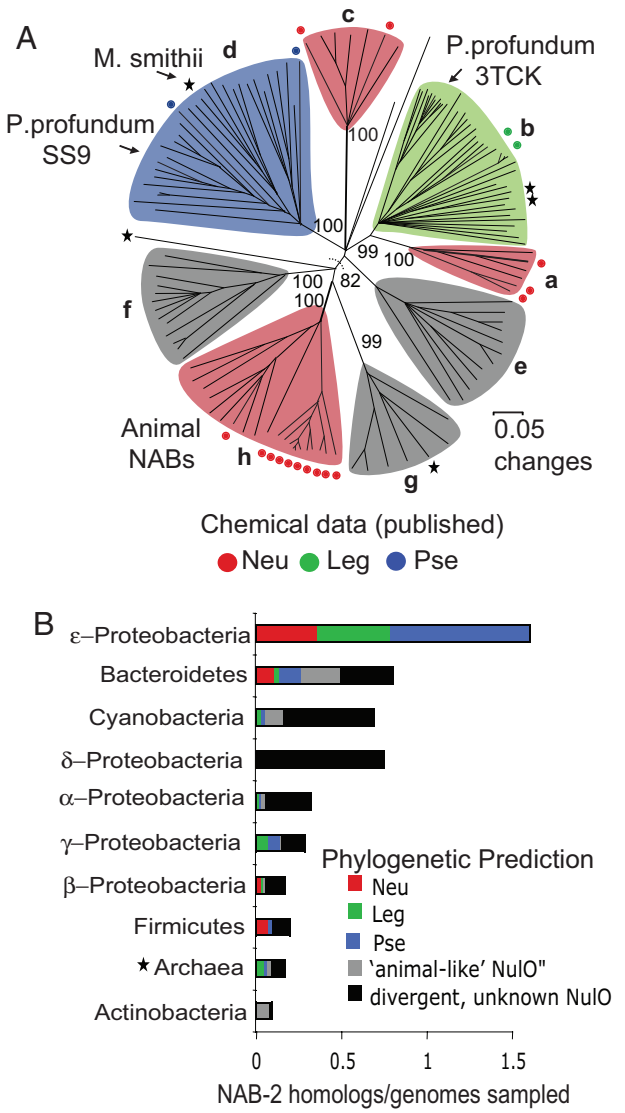


Fig. 3. NAB-2 phylogeny for predicting NulO structure. (*A*), A distance-based neighbor-joining tree was constructed as described in *Materials and Methods*, and published biochemical data were overlaid onto the tree (colored circles). A ''cohesion group'' approach (21) was used to infer enzyme specificities by extrapolation of the published biochemical data (13–15, 22–24) to phylogenetic clades supported by high bootstrap values (shown at relevant nodes). Color shading reflects the phylogenetic predictions of chemical structure for each clade. Clades are designated ''a''–''h'' for reference. Note that clade ''h'' is composed entirely of NAB-2 enzymes from animals. Organisms for which biochemical data are presented in later figures are indicated by name. (*B*), Percentages of NAB-2 clade affiliations were calculated and expressed as a function of the number of genomes surveyed in each NCBI-classified taxonomic group (see *Materials and Methods*). Note that ''pruned'' tree branches representing diverged NAB-2 enzymes were included in the tabulation of NAB-positive genomes as ''unknown'' NulO type (shown in black) if a nearby NAB-1 homolog was identified. Some individual genomes encode multiple NAB pathways leading to NAB-2 homolog/genome sampled values of >1. These data, with strain and sequence identifiers, are summarized in Table S1.

1). Microbial NAB-2 homologs with published roles in Neu synthesis were identified in multiple clades that are phylogenetically distinct from the clade with animal NAB-2 homologs (Fig. 3*A*). Specifically, the analysis revealed 2 groups of bacterial Neu synthases, represented by clades ''a'' and ''c'' (Fig. 3*A*), which include sequences from well-known Sia-decorated pathogens (Table 1). These data point to the existence of at least 3 evolutionarily discrete types of Neu synthases that ultimately share a common ancestor (i.e., semiconvergence). Examination of other NAB pathway enzymes revealed similar phylogenetic signatures, indicating that most microbial NAB pathway cassettes persist as evolutionarily conserved units (Fig. S1). These data predict Neu expression in several pathogenic and commensal bacterial species in which NAB biosynthesis has not been characterized previously (Table 1). Moreover, the data strongly suggest that Neu expression is *not* limited to microbes that associate with animals, as has been commonly assumed (see the composition of clade ''c'' in Table 1).

One evolutionary explanation for the emergence of Sia mimicry is represented by highly supported nodes in the NAB-2 and NAB-3 trees showing that clade ''a'' (Neu-specific enzymes encoded exclusively in animal-associated bacteria) shares a common ancestor with clade ''b'' (Leg-specific enzymes encoded mostly by aquatic organisms) (Fig. 3*A* and Fig. S1). Interestingly, the phylogenetic relationship between NAB pathways represented by clades ''a'' and ''b'' do not reflect known evolutionary relationships between organisms represented in these clades. Consistent with the shared ancestry of clades ''a'' and ''b,'' organisms represented in these clades (but not those in clade ''c'') encode acetyltransferases as part of their NAB gene cassettes. Previous studies have indicated that such acetyltransferases are required for overall NulO expression (16, 25, 26). For example, *Campylobacter jejuni* (clade ''b'') requires PtmH for *N*-acetylation at the 7-carbon position of Pse residues (16), whereas *Streptococcus agalactiae* and *Escherichia coli* (clade ''a'') use the homologous NeuD enzyme for *O*-acetylation at the same carbon position of Neu (26). These data indicate that Neu biosynthetic pathways in clade ''a'' were not acquired by lateral gene transfer from an animal host, but rather the Neu mimicry by organisms represented in clade ''a'' arose by recruitment and modification of an ancestral NulO pathway that requires an acetylation reaction at C7 (Fig. 1).

NulO Biosynthetic Pathways Originated and Diversified Early in the History of Cellular Life. Based on the distribution of Sias among animals, it has been suggested that the biosynthetic pathway for this NulO sugar may have been an innovation of the deuterostome lineage (1). But the NAB-2-based phylogeny (Fig. 3) revealed novel phylogenetic clusters of microbial NAB homologs highly similar to components of animal Sia pathways (clades ''f,'' ''g,'' and possibly ''e''), suggesting that Sia synthesis pathways of animals may have much deeper evolutionary roots. Notably, the branching pattern of the limited number of taxa within these clades is consistent with known evolutionary relationships among these organisms. Targeted phylogenetic trees and protein domain analyses of NAB-1 and NAB-2 pairs encoded in the animal and ''animal-like'' clades further support the deep but firm evolutionary relationship between these biosynthetic pathways (Fig. S2). A few of the organisms represented in these clades associate in various ways with animals, including the spirochetal zoonotic pathogen *Leptospira interrogans* and the Actinobacteria *Brevibacterium linens* and *Thermobifida fusca*, which that cause body odor and farmers' lung, respectively. All other microbes represented in these clades are environmentally associated, however.

A model of early cellular diversification of NAB pathways (Fig. S3), including those for Sias, is supported by multiple lines of evidence, including phylogenetic and protein domain comparisons (Fig. 3*A* and Fig. S2), the wide distribution of NAB pathways within members of Bacteria and the considerable diversity of their predicted sugar structures (Figs. 2 and 3*B*), and the diverse composition of taxa in the animal-like clades (Table S1). The presence of archaeal NAB sequences in several distinct clades of the phylogenetic tree (Fig. 3*A*) also supports the conclusion that paralogous gene duplications with divergence of

Table 1. Phylogenetic prediction of sialic acid synthesis in bacteria

	Organism	NCBI class*	Host/environment	Disease	NulO known
a	*S.agalactiae*	4	Human	MG, SP, PN	Neu
	S. suis	4	Swine, human	MG, SP, endocarditis	Neu
	E. coli K1	3	Human, avian	MG, SP, GE, cystitis	Neu
	R. gnavus	4	Human	NK, gut symbiont	NK
	F. nucleatum	5	Human, animal	Periodontal diseases	NK
c	*C. jejuni*	1	Chicken, human	GE, autoimmunity	Neu
	N. meningitidis	2	Human	MG, SP	Neu
	B. cereus G9241	4	Soil, animal	Anthrax-like PN	NK
	H. acinonychis	1	Feline	Gastritis in cheetah	NK
	F. johnsoniae	3	Soil, aquatic	NK	NK
	F. psychrophilum	3	Aquatic, fish	Fry syndrome, fish	NK
	Algoriphagus sp. *PR1*	3	Aquatic	NK	NK
	Flavobacterium sp. *MED217*	3	Seawater	NK	NK
	I. ioihiensis	4	NK, hydrothermal vent	NK	NK

MG; meningitis. SP, septicemia. PN, pneumonia. GE, gastroenteritis. NK; none known. Here "a" and "c" refer to phylogenetic clades shown in Fig. 3*A*.
*NCBI classifications are as follows: 1, ε-Proteobacteria; 2, β-Proteobacteria; 3, Bacteroidetes; 4, Firmicutes; 5, Fusobacteria.

enzymatic function occurred very early in cellular evolution. We conclude that NAB enzymes, which at first glance appear to be animal-like, should be more accurately considered present-day components of an ancient pathway that was universally adopted by deuterostome animals. The distant sequence relationship between animal and animal-like clades suggests that Sia biosynthesis did not arise by lateral transfer into the animal lineage, but rather that Sia synthesis was likely inherited in the traditional sense, accompanied by multiple gene losses along other eukaryotic lineages (27, 28).

Chemical Validation of the Phylogenomic/Phyloglycomic Approach. To provide functional validation of our phylogenomic findings, we present 2 striking examples that illustrate the utility of a phylogenetic approach for predicting chemical structure and inferring the evolutionary history of this class of monosaccharides.

***Photobacterium profundum* Strains Encode Phylogenetically Distinct NAB Enzymes and Biochemically Distinct NulO Sugars.** The phylogenetic data predict that 2 otherwise closely related strains of *Photobacterium profundum* (3TCK and SS9) encode biosynthetic pathways for distinct NulO sugar structures (Fig. 3). We tested this prediction using a well-validated approach for Sia structure identification. Leg and Pse acid standards were isolated from purified LPS preparations (29, 30) and, after fluorescent derivatization with 1,2-diamino-4,5-methylene dioxybenzene (DMB) (26), they eluted at distinct HPLC retention times (Fig. 4*A* and *B*). Tandem mass spectrometry confirmed the expected masses of these eluted Leg and Pse standards, showing that the DMB-HPLC approach can be effectively applied to the broader class of NulO sugars. In another experiment, *P. profundum* SS9 and 3TCK strains were grown under optimized conditions, and mild acid hydrolysis was used to release NulO sugars. NulO sugars isolated from *P. profundum* genome strains 3TCK and SS9 exhibited the expected retention times and masses of Leg and Pse standards based on the phylogenetic prediction (Fig. 4*C* and *D*).

***Methanobrevibacter smithii*: A Case of Mistaken Identity.** The published genome sequence of the principal human gut–associated methanogen, *Methanobrevibacter smithii*, contains a gene cluster originally annotated to encode a Sia (Neu) biosynthesis pathway. This archaeon has a prominent capsule when grown in vitro, and previous biochemical assays plus lectin-binding immunohistochemical studies suggested that NulO sugars, presumably Neu5Ac, are expressed by the cultured-type strain (7). Further analyses using custom *M. smithii* GeneChips (see *Materials and Methods*) revealed that this gene cluster was present in 7/7 *M. smithii* isolates examined, and that expression of these genes is differentially regulated in vitro (Fig. S4).

Here we show that the amino acid sequence of *M. smithii* NAB-2 clusters is within clade "d," and thus is predicted to synthesize Pse rather than Neu acids (Fig. 3*A*). To test this hypothesis, we first compared the retention times and masses of DMB-derivatized α-keto acid standards Neu and Pse using

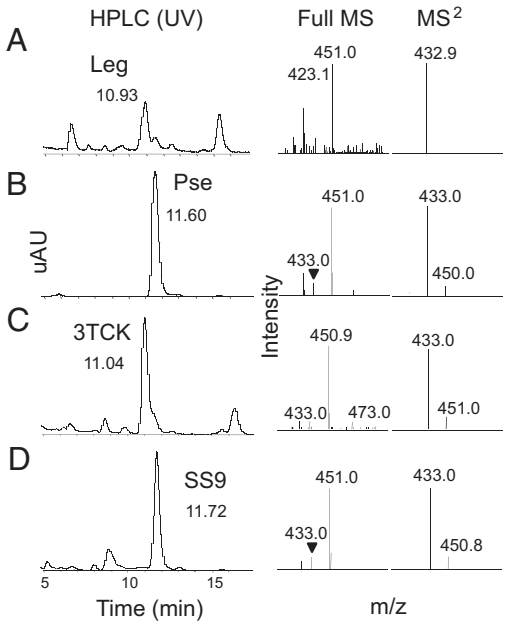


Fig. 4. Targeted chemical validation of *Photobacterium profundum* NAB pathways. LCMS analysis of NulO sugars isolated from purified LPS of Leg (*A*) or Pse (*B*) or from cultures of *P. profundum* genome strain (*C*) 3TCK or SS9 (*D*). Different HPLC retention times enable differentiation between Leg and Pse and reliable identification of NulO in 3TCK and SS9. MS and MS^2 analyses provide additional confirmation of the expected mass for Leg and Pse for both standards and unknowns. In all cases, MS data are shown from 400–500 *m/z*.

MICROBIOLOGY

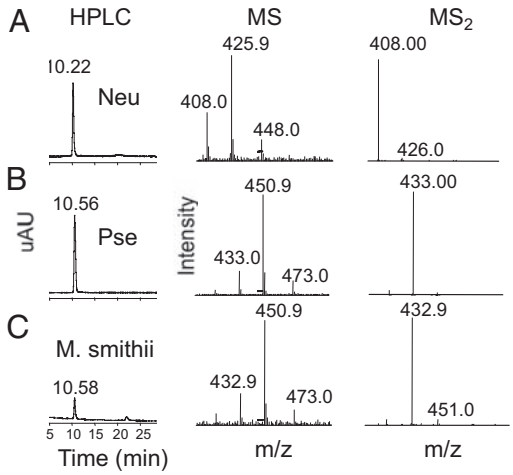


Fig. 5. The human gut methanogen *M. smithii* synthesizes pseudaminic, but not *N*-acetylneuraminic, acid. LCMS analysis of NulO sugars *N*-acetylneuraminic acid (Neu5Ac) (*A*), Pse from *Pseudoalteromonas atlantica* LPS (*B*), or isolated from cell pellets of *M. smithii* (*C*), as described in *Materials* and *Methods*. In all cases, MS data are shown from 400–500 *m/z*.

reverse-phase HPLC with tandem mass spectrometry, and found that Neu and Pse have similar but distinct retention times and can be discriminated on the basis of mass-to-charge (*m/z*) ratio (Fig. 5*A* and *B*). Parallel preparation and analysis of *M. smithii* NulO sugars in parallel with Neu and Pse standards clearly demonstrated a retention time and mass consistent with Pse expression (Fig. 5*C*). This finding again validates the phylogenetic approach to predicting NulO structure, and emphasizes that available methods for Sia detection can be successfully expanded to consider the larger family of NulO sugars.

Prospectus. In summary, our results provide insight into the evolutionary history of Sias by considering them in the larger phylogenetic context of related NulO sugars. We emphasize that the surprisingly wide distribution of NAB pathways among the 3 domains of life (Bacteria, Archaea, and Eukarya) is a reflection of many interwoven evolutionary processes, including gene duplications with functional divergence, gene loss, lateral gene transfer, and more specific adaptations of biosynthetic pathways. Clearly, much remains to be done to understand the biology and evolution of these remarkably common and diverse carbohydrate molecules found at the surfaces of contact between many bacteria and their external environments. These findings serve as a proof of principle for the utility of a phylogenomic/phyloglycomic approach to predicting NulO sugar types and strongly suggest that the expression of Sias and Sia-like sugars by bacteria may be advantageous in a wide range of animal body habitats. Determining the advantages of Sia mimicry in these different host contexts, as well as identifying those factors contributing to lateral dissemination of Sia gene cassettes (i.e., functional clusters) and their incorporation into various LPS or capsular polysaccharide biosynthetic pathways, are important areas for further investigation. These studies should provide a path for future investigation concerning the contributions of Sias and related sugars to the survival and persistence of microbes in both host and environmental reservoirs, as well as in disease pathogenesis.

Materials and Methods

Identification of Physically Clustered NAB Genes. BLASTp, on the ''BLAST with microbial genomes'' webpage at the National Center for Biotechnology Institute (NCBI) website (www.ncbi.nlm.nih.gov/sutils/genom_table.cgi), was used to query 960 complete microbial genomes. Multiple NAB-1 and NAB-2 amino acid sequences were used for genomic query, with an emphasis on NAB enzymes of defined function or those encoded in organisms known to express specific NulO structures. Incomplete genomes also were queried using the nonredundant protein database and were included in the data set if deemed NAB-positive. Accession numbers for homologous NAB sequences were cataloged according to the NCBI taxonomic classification and examined for ''functional clustering'' (20) of NAB-1 and NAB-2 enzymes, as judged by proximal accession numbers. NAB-1 homologs were validated by phylogenetic analysis, and ''contaminating'' CMP-Kdo synthetases were removed; these distant NAB-1 homologs were not found in functional clusters with NAB-2 homologs. Note that Kdo is an 8-carbon α-keto acid that follows a similar biosynthetic pathway involving condensation of a 5-carbon sugar with phosphoenolpyruvate, followed by activation to CMP-Kdo (1). Results of the genomic profiling were compiled, and the proportion of genomes with one or more physical clusters of NAB-1 and NAB-2 genes was expressed as a function of total genomes surveyed in the different microbial taxa (Fig. 2).

NAB-2 Phylogeny for Prediction of NulO Sugar Types. Amino acid sequence comparisons indicated that NAB-2 sequences are better conserved than other NAB enzymes and form the most conclusive basis for prediction of specific NulO sugars in different organisms. NAB-2 amino acid sequences were collected using BLASTp at the NCBI nonredundant protein database and aligned using ClustalQt (Fig. S5) (31). Nexus files from the alignment were uploaded into PAUP* 4.0b10 (32) for exclusion of gaps and domains not found in all sequences, followed by construction of a neighbor-joining tree using the bootstrap/jackknife option with 1,000 replicates. Less-conserved sequences were apparent from visual inspection of the alignment and clustering of branches in a star shape at the base of the phylogenetic tree. Such branches were ''pruned'' from the tree in successive analyses to reveal significant monophyletic clades and improve aspects of the alignment that could better resolve sequence relationships between different clades. Analysis of the same alignment based on parsimony produced a nearly identical phylogenetic pattern as that from the distance-based approach. NulO types were predicted from extrapolations of published biochemical data (1, 12–16, 22–24, 33–35) to other members of clades supported by high bootstrap values (shown in Fig. 3*A*). To determine the distribution of NulO types among different microbial phylotypes, the percentages of NAB-2 clade affiliations (Fig. 3*A*–*G*) were calculated and expressed as a function of the number of genomes surveyed in each NCBI-classified taxonomic group (Fig. 3*B*). ''Pruned'' tree branches representing diverged NAB-2 enzymes were included in the tabulation of NAB-positive genomes as ''unknown'' NulO type (shown in black in Fig. 3*B*) if a nearby NAB-1 homolog was identified. Note that many individual genomes encode multiple NAB pathways (e.g., ε-Proteobacteria, Bacteroidetes; see Table S1), sometimes leading to a ratio of NAB-2 homologs:genomes sampled of >1 in Fig. 3*B*.

Strains and Culture Conditions. The 2 strains of *Photobacterium profundum* with complete genome sequences are close phylogenetic relatives but are adapted to different aquatic ecosystems (36). *P. profundum* SS9 is a piezophilic (pressure-loving) strain that grows optimally under low-temperature, high-pressure conditions. In contrast, *P. profundum* 3TCK is adapted to an aquatic niche much closer to the surface. High-pressure growth of SS9 was performed anaerobically at 16 °C in 2216 medium (Difco) supplemented with 20 mM glucose and 100 mM Hepes buffer (pH 7.5) (Sigma). Late-exponential phase cultures were diluted 500-fold into fresh medium and used to fill 4.5-mL polyethylene transfer pipettes (Samco). Transfer pipettes were heat-sealed with a hand-held heat-sealing clamp (Nalge) and incubated at 30 MPa in a stainless steel pressure vessel (37). 3TCK was cultivated in the same media but incubated at room temperature without shaking.

Three strains of *M. smithii* were obtained from the DSMZ culture collection (2374, 2375, and 11975), 4 strains were isolated from a single human fecal sample by selective culturing, and the sequenced-type strain (PS) was obtained from ATCC. *M. smithii* was grown in supplemented MBC medium under anaerobic conditions for 6 days at 37 °C as described previously (7).

GeneChip-Based Studies of *M. smithii*. Genomic DNA was prepared and hybridized to a custom Affymetrix GeneChip containing probesets that recognize 99% of its 1,795 predicted protein coding genes. Similarly, RNA isolated under different in vitro growth conditions was hybridized to GeneChips as described previously (See *SI Text*) (7).

Chemical Analysis of NulO Acids. *M. smithii* and *P. profundum* were harvested from cultures by centrifugation and washed twice with PBS. NulO residues were released from cells or purified LPS samples containing Leg (29) or Pse (30) acids by mild acid hydrolysis, and low molecular weight fractions were sub-

jected to derivatization with DMB, followed by HPLC analysis and liquid chromatography–mass spectrometry (LCMS) as described previously (see *SI Text*) (38).

ACKNOWLEDGMENTS. We thank Russell F. Doolittle for many helpful discussions, Sandra Diaz for assisting with mass spectrometry, Doug Bartlett for providing *P. profundum*, Emiley Eloe for performing high-pressure cultivation of SS9, and Henning Seedorf for assisting with the GeneChip hybridizations.

We appreciate the efforts of Olga Zagnitko in updating annotations of NAB pathways in The SEED, an annotation/analysis tool provided by the Fellowship for Interpretation of Genomes, as well as insightful comments from Aaron Carlin, Yung-Chi Chang, and Shannon Weiman. This work was funded by generous support from the University of California President's Postdoctoral Fellowship Program (to A.L.L.) and the Gianinni Family Foundation (to A.L.L.), along with National Institutes of Health grants HL057345 (to A.V.), DK30292 (to J.I.G.), and R01-HD051796 (to V.N.).

1. Angata T, Varki A (2002) Chemical diversity in the sialic acids and related alpha-keto acids: An evolutionary perspective. *Chem Rev* 102:439–469.
2. Eylar EH, Doolittle RF, Madoff MA (1962) Sialic acid from blood cells of the lamprey eel. *Nature* 193:1183–1184.
3. Varki A (2007) Glycan-based interactions involving vertebrate sialic acid–recognizing proteins. *Nature* 446:1023–1029.
4. Warren L (1963) The distribution of sialic acids in nature. *Comp Biochem Physiol* 10:153–171.
5. Vimr E, Lichtensteiger C (2002) To sialylate, or not to sialylate: That is the question. *Trends Microbiol* 10:254–257.
6. Kelm S, Schauer R (1997) Sialic acids in molecular and cellular interactions. *Int Rev Cytol* 175:137–240.
7. Samuel BS, et al. (2007) Genomic and metabolic adaptations of *Methanobrevibacter smithii* to the human gut. *Proc Natl Acad Sci USA* 104:10643–10648.
8. Damian RT (1965) Molecular mimicry in biological adaptation. *Science* 147:824.
9. Carlin AF, Lewis AL, Varki A, Nizet V (2007) Group B streptococcal capsular sialic acids interact with siglecs (immunoglobulin-like lectins) on human leukocytes. *J Bacteriol* 189:1231–1237.
10. Ram S, et al. (1998) A novel sialic acid binding site on factor H mediates serum resistance of sialylated *Neisseria gonorrhoeae*. *J Exp Med* 187:743–752.
11. Carlin AF, et al. (2009) Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage neutrophil Siglec-9 and dampen the innate immune response. *Blood* 113:3333–3336.
12. Knirel YA, Shashkov AS, Tsvetkov YE, Jansson PE, Zähringer U (2003) 5,7-Diamino-3,5,7,9-tetradeoxynon-2-ulosonic acids in bacterial glycopolymers: Chemistry and biochemistry. *Adv Carbohydr Chem Biochem* 58:371–417.
13. Glaze PA, Watson DC, Young NM, Tanner ME (2008) Biosynthesis of CMP-N,N′-diacetyllegionaminic acid from UDP-N,N′-diacetylbacillosamine in *Legionella pneumophila*. *Biochemistry* 47:3272–3282.
14. McNally DJ, et al. (2007) Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. *J Biol Chem* 282:14463–14475.
15. Schoenhofen IC, McNally DJ, Brisson JR, Logan SM (2006) Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: Synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. *Glycobiology* 16:8C–14C.
16. McNally DJ, et al. (2006) Functional characterization of the flagellar glycosylation locus in *Campylobacter jejuni* 81–176 using a focused metabolomics approach. *J Biol Chem* 281:18489–18498.
17. Schirm M, et al. (2003) Structural, genetic and functional characterization of the flagellin glycosylation process in *Helicobacter pylori*. *Mol Microbiol* 48:1579–1592.
18. Logan SM, et al. (2009) Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics-based approaches. *FEBS J* 276:1014–1023.
19. Twine SM, et al. (2008) Flagellar glycosylation in *Clostridium botulinum*. *FEBS J* 275:4428–4444.
20. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N (1999) The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA* 96:2896–2901.
21. Bonner CA, et al. (2008) Cohesion group approach for evolutionary analysis of TyrA, a protein family with wide-ranging substrate specificities. *Microbiol Mol Biol Rev* 72:13–53.
22. Chou WK, Dick S, Wakarchuk WW, Tanner ME (2005) Identification and characterization of NeuB3 from *Campylobacter jejuni* as a pseudaminic acid synthase. *J Biol Chem* 280:35922–35928.
23. Hao J, Balagurumoorthy P, Sarilla S, Sundaramoorthy M (2005) Cloning, expression, and characterization of sialic acid synthases. *Biochem Biophys Res Commun* 338:1507–1514.
24. Suryanti V, Nelson A, Berry A (2003) Cloning, over-expression, purification, and characterisation of N-acetylneuraminate synthase from *Streptococcus agalactiae*. *Protein Expr Purif* 27:346–356.
25. Daines DA, Wright LF, Chaffin DO, Rubens CE, Silver RP (2000) NeuD plays a role in the synthesis of sialic acid in Escherichia coli K1. *FEMS Microbiol Lett* 189(2):281–284.
26. Lewis AL, Hensler ME, Varki A, Nizet V (2006) The group B streptococcal sialic acid O-acetyltransferase is encoded by neuD, a conserved component of bacterial sialic acid biosynthetic gene clusters. *J Biol Chem* 281:11186–11192.
27. Salzberg SL, White O, Peterson J, Eisen JA (2001) Microbial genes in the human genome: Lateral transfer or gene loss? *Science* 292:1903–1906.
28. Doolittle RF (2002) Gene transfers between distantly related organisms. *Horizontal Gene Transfer* eds Syvanen M, Kado C (Academic, New York), 2nd Ed, pp 269–275.
29. Knirel YA, Rietschel ET, Marre R, Zähringer U (1994) The structure of the O-specific chain of *Legionella pneumophila* serogroup 1 lipopolysaccharide. *Eur J Biochem* 221:239–245.
30. Perepelov AV, et al. (2005) Structure of an acidic polysaccharide from the agar-decomposing marine bacterium *Pseudoalteromonas atlantica* strain IAM 14165 containing 5,7-diacetamido-3,5,7,9-tetradeoxy-L-*glycero*-L-*manno*-non-2-ulosonic acid. *Carbohydr Res* 340:69–74.
31. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinform* Unit 2.3.
32. Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods) (Sinauer Associates, Sunderland, MA), Version 4.
33. Sundaram AK, et al. (2004) Characterization of N-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. *Biochem J* 383(Pt 1):83–89.
34. Gunawan J, et al. (2005) Structural and mechanistic analysis of sialic acid synthase NeuB from *Neisseria meningitidis* in complex with Mn^{2+}, phosphoenolpyruvate, and N-acetylmannosaminitol. *J Biol Chem* 280:3555–3563.
35. Shashkov AS, et al. (2007) Structure of the O-antigen of *Providencia stuartii* O20, a new polysaccharide containing 5,7-diacetamido-3,5,7,9-tetradeoxy-l-*glycero*-d-*galacto*-non-2-ulosonic acid. *Carbohydr Res* 342:653–658.
36. Campanaro S, et al. (2005) Laterally transferred elements and high pressure adaptation in *Photobacterium profundum* strains. *BMC Genomics* 6:122.
37. Eloe EA, Lauro FM, Vogel RF, Bartlett DH (2008) The deep-sea bacterium *Photobacterium profundum* SS9 utilizes separate flagellar systems for swimming and swarming under high-pressure conditions. *Appl Environ Microbiol* 74:6298–6305.
38. Lewis AL, Nizet V, Varki A (2004) Discovery and characterization of sialic acid O-acetylation in group B *Streptococcus*. *Proc Natl Acad Sci USA* 101:11123–11128.

Supporting Information

Lewis et al. 10.1073/pnas.0902431106

SI Materials and Methods

GeneChip-Based Studies of *M. smithii*. This custom GeneChip was also used for whole genome transcriptional profiling of the type strain. Cells were grown at 37 °C, with or without agitation (100 rpm), in 125-mL serum bottles containing 15 mL of supplemented MBC medium (7) under an atmosphere of H_2 and CO_2 (4:1) that was replenished every 6 h, and harvested during the log or stationary phase (log phase: OD_{600} of 1.10 and 0.36 for agitated and static cultures, respectively; stationary phase: OD_{600} of 3.14 and 0.57, respectively). RNA was isolated, and cDNAs were prepared and then hybridized to GeneChips as described previously (7) (n = 9–13 GeneChips/condition). GeneChip-wide normalization (to an intensity of 500) was carried out with an Affymetrix MAS5. The significance of observed differences in gene expression was determined using a 2-tailed Student t test.

Isolation and Characterization of NulOs. After resuspension in 2N acetic acid, cell suspensions were incubated at 80 °C for 3 h to release cell surface NulO residues. Insoluble cell debris was pelleted at maximum speed on a tabletop centrifuge, and material released into the soluble fraction was passed over a 10-K molecular weight cutoff filtration unit (Centricon). Purified LPS samples containing Leg (25) or Pse (26) acids were processed similarly by mild acid hydrolysis and filtration. Low molecular weight fractions or commercially available Neu were derivatized with DMB, and NulO–DMB adducts were resolved by HPLC using a reverse-phase C18 column (Varian) eluted isocratically at a rate of 0.9 mL/min over 50 min using 85% MQ water, 7% methanol, and 8% acetonitrile. DMB-derivatized extracts or individually isolated HPLC peaks were analyzed by LCMS using a Finnigan-MAT HPLC system with a tandem LCQ mass spectrometer (46). Detection of fluorescently labeled NulO sugars was achieved at excitation and emission wavelengths of 373 nm and 448 nm, respectively.

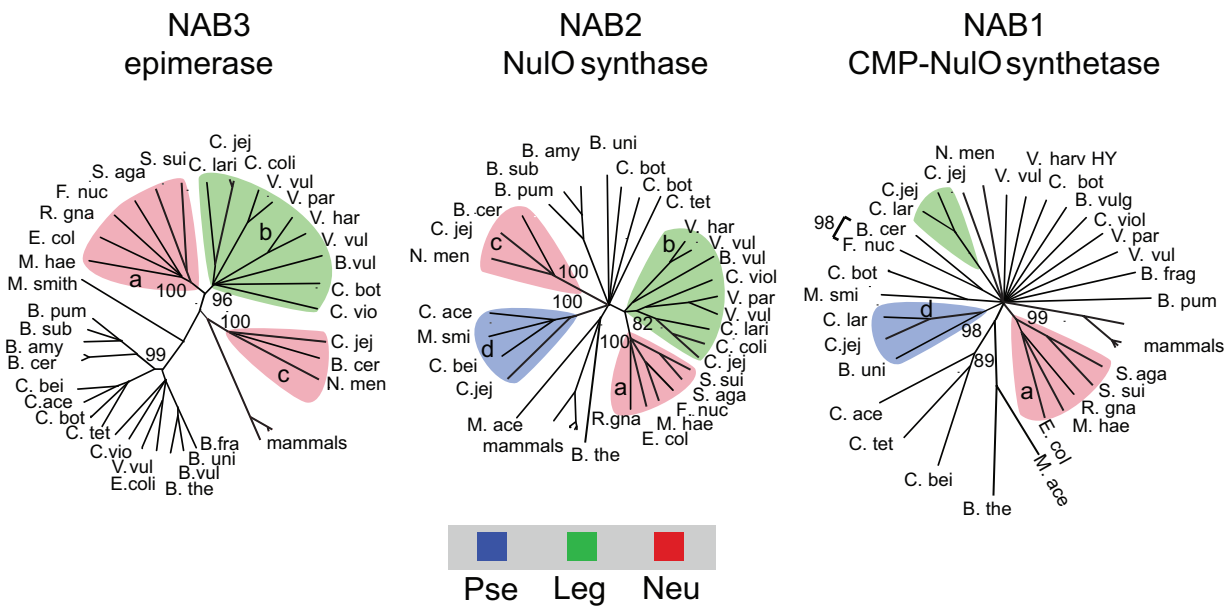


Fig. S1. Phylogenies of NAB pathways reveal distinct bacterial innovations of Sia mimicry. Phylogenetic trees were constructed based on NAB-1, NAB-2, and NAB-3 amino acid sequences collected from pathogenic bacteria represented in clades ''a'' and ''c,'' as well as a defined subset of organisms representative of other phylogenetic clades shown in Fig. 3. Colored branches indicate published biochemical data for specific NulO residues, as shown in the color key. Shading reflects monophyletic clades with high bootstrap support.

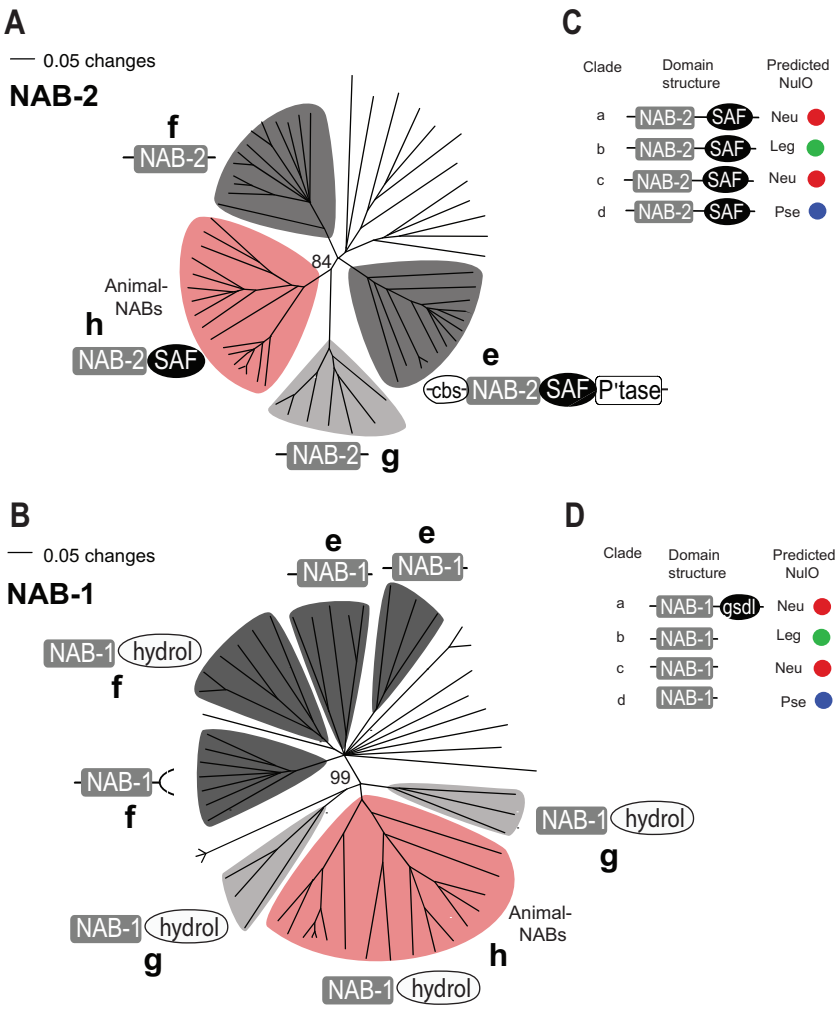


Fig. S2. Phylogenetic analysis of animal and animal-like NAB-1 and NAB-2 amino acid sequences are consistent with domain structure. (*A* and *B*), NAB-1 (*A*) and NAB-2 (*B*) amino acid sequences from organisms represented in Fig. 3, phylogenetic clades ''e''–''h,'' were collected by BLASTp and subjected to a phylogenetic analysis that included only the NAB-1 or NAB-2 domains common to all sequences in the alignment. Protein domain organizations for clades based on the Pfam database (47), and our amino acid alignments were overlaid onto the tree. Red shading highlights the animal taxa (clade ''h'') that express Sias on their cell surfaces. All other clades (shown in gray) reflect novel phylogenetic classes of animal-like NAB enzymes for which no biochemical data currently exist. Clade ''g'' is shown in a lighter gray than other animal-like NABs to emphasize its closer phylogenetic relationship with NABs from animals as indicated by bootstrap values (shown). As in other analyses, NAB-1 appears less well conserved but shares similar phylogenetic features with functionally clustered NAB-2 sequences from the same organisms. Note that the ancestral architecture of NAB-2 likely included the C-terminal SAF domain, and that most of the animal-like NAB-1 sequences (similar to animal NAB-1 and in contrast to the sequences in clades ''a''–''d'') contain a C-terminal domain with predicted hydrolase activity.(*C*) and *D*, Protein domain organization for clades ''a''–''d'' shown for comparison. Pfam numbers for domains are as follows: NAB-1, PF02348; NAB-2, PF03102; SAF, PF08666; CBS, PF00571; P'tase, PF01261; GSDL, PF00657; Hydrol, PF00702.

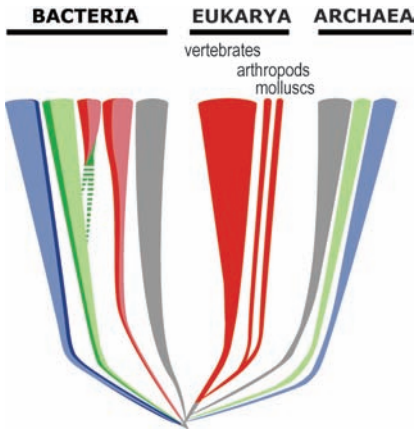


Fig. S3. A model of NulO evolution based on phylogenomic evidence. Based on phylogenetic and genomic evidence, we suggest that an early cellular diversification of NulO sugar structures resulted in the wide variety and distribution of NulO sugars that we find today (darker colors reflect published data; lighter colors indicate phylogenetic predictions). At least 3 distinct semiconvergent evolutionary paths for de novo biosynthesis of Sias are supported by the phylogenetic and biochemical data (i.e., in animals and 2 different groups of microbes often found in close association with Sia-expressing animals).

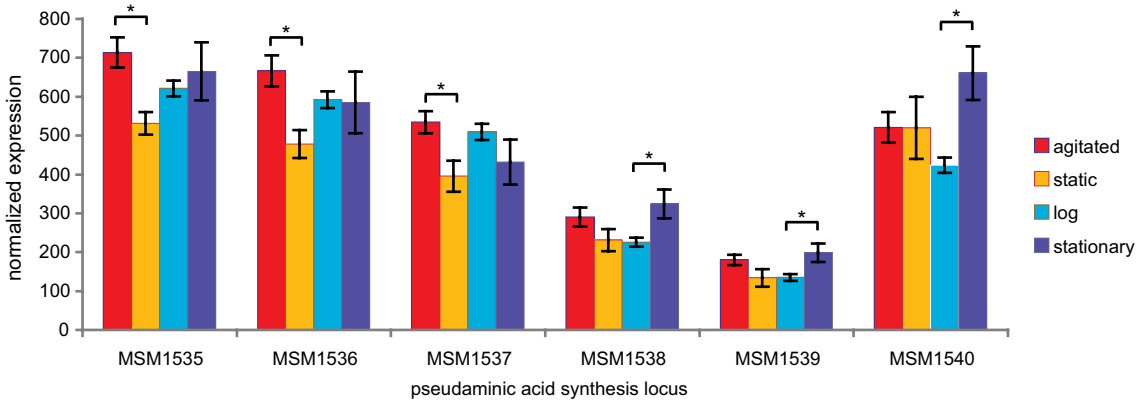


Fig. S4. Expression of predicted *M. smithii* Pse synthesis genes during growth from log phase to stationary phase in standard medium, with or without agitation. Mean values for GeneChip probeset intensities ± SEM are plotted ($n = 9–13$ GeneChips/condition). Asterisks indicate statistically significant differences ($P < .05$ by the Student *t* test). Original annotations of these genes were dTDP-D-glucose 4,6 dehydratase (MSM1535), acylneuraminate cytidylyltransferase (MSM1537), CMP-sialic acid synthetase (MSM1538), and a sialic acid synthase (MSM1539) (7). MSM1536 and MSM1540 encode a pleiotropic regulatory protein DegT and glycerol-3-phosphate dehydrogenase, respectively.

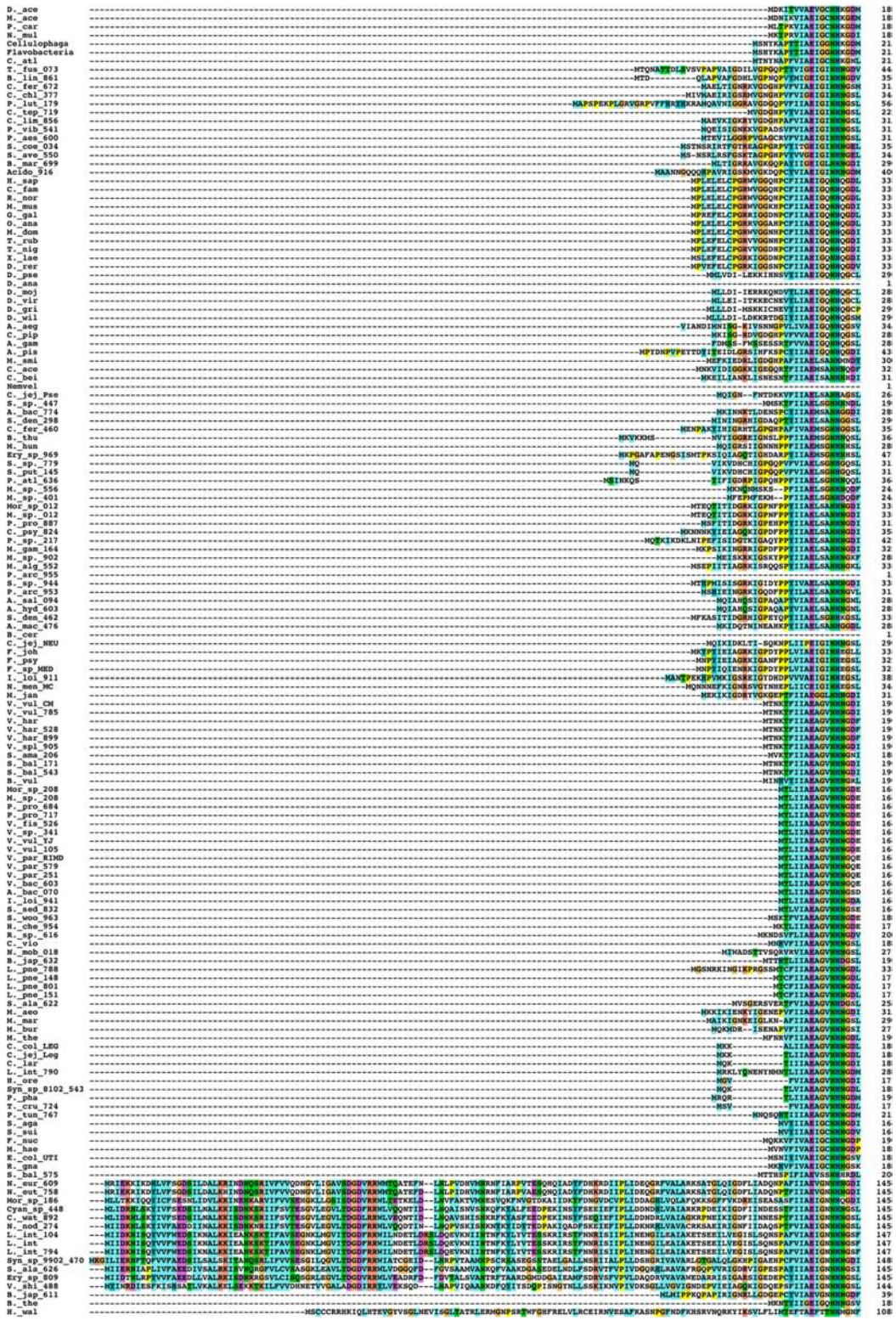


Fig. S5. Multiple sequence alignment of NAB-2 amino acid sequences used for construction of the phylogenetic tree in Fig. 3. Clustal Qt alignment of NAB-2 sequences. With the exception of the more limited sequences from sea anemone (*Nematostella*) and *P. arctica*, all gaps in the alignment were excluded from the phylogenetic analysis.

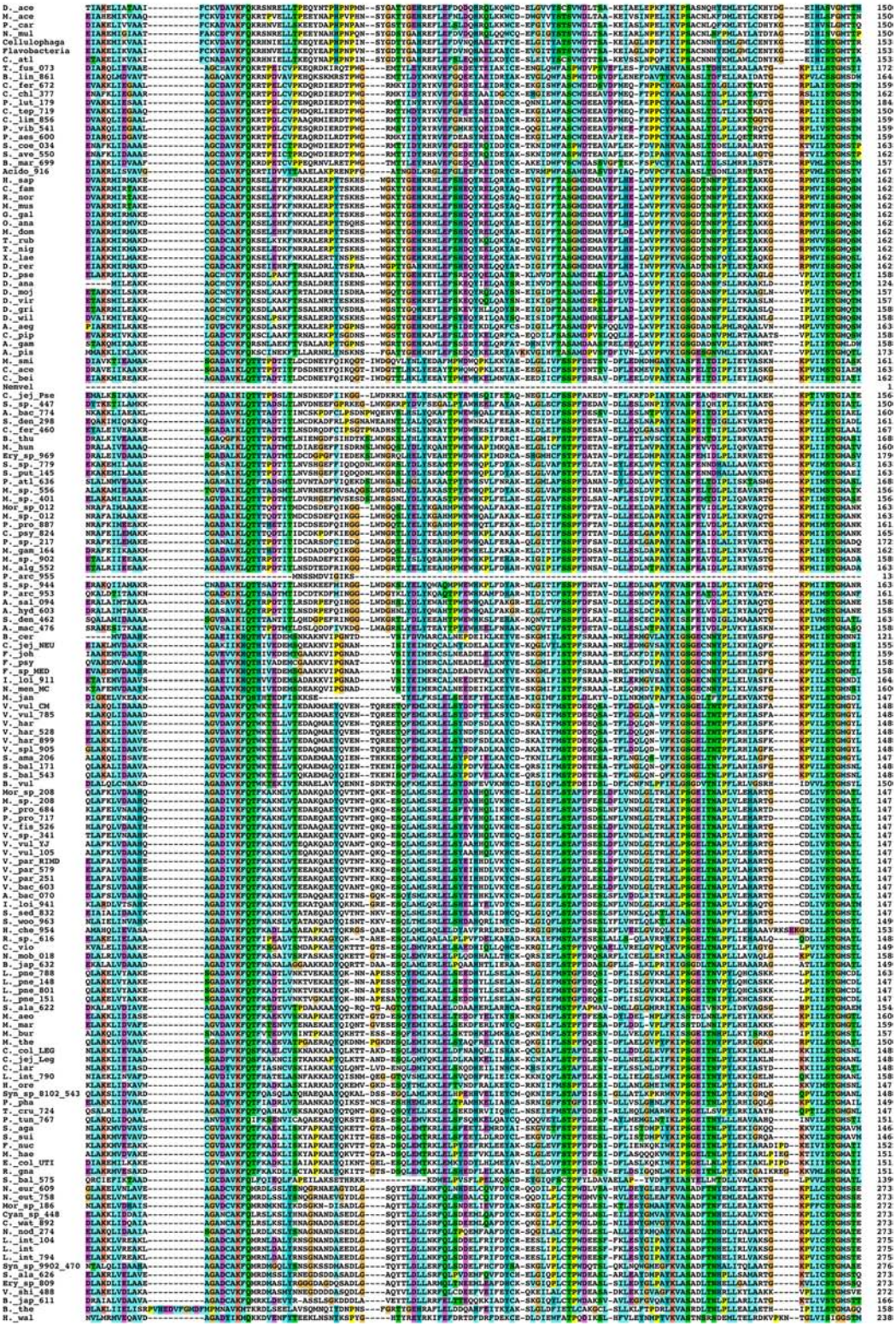


Fig. S5. Continued.

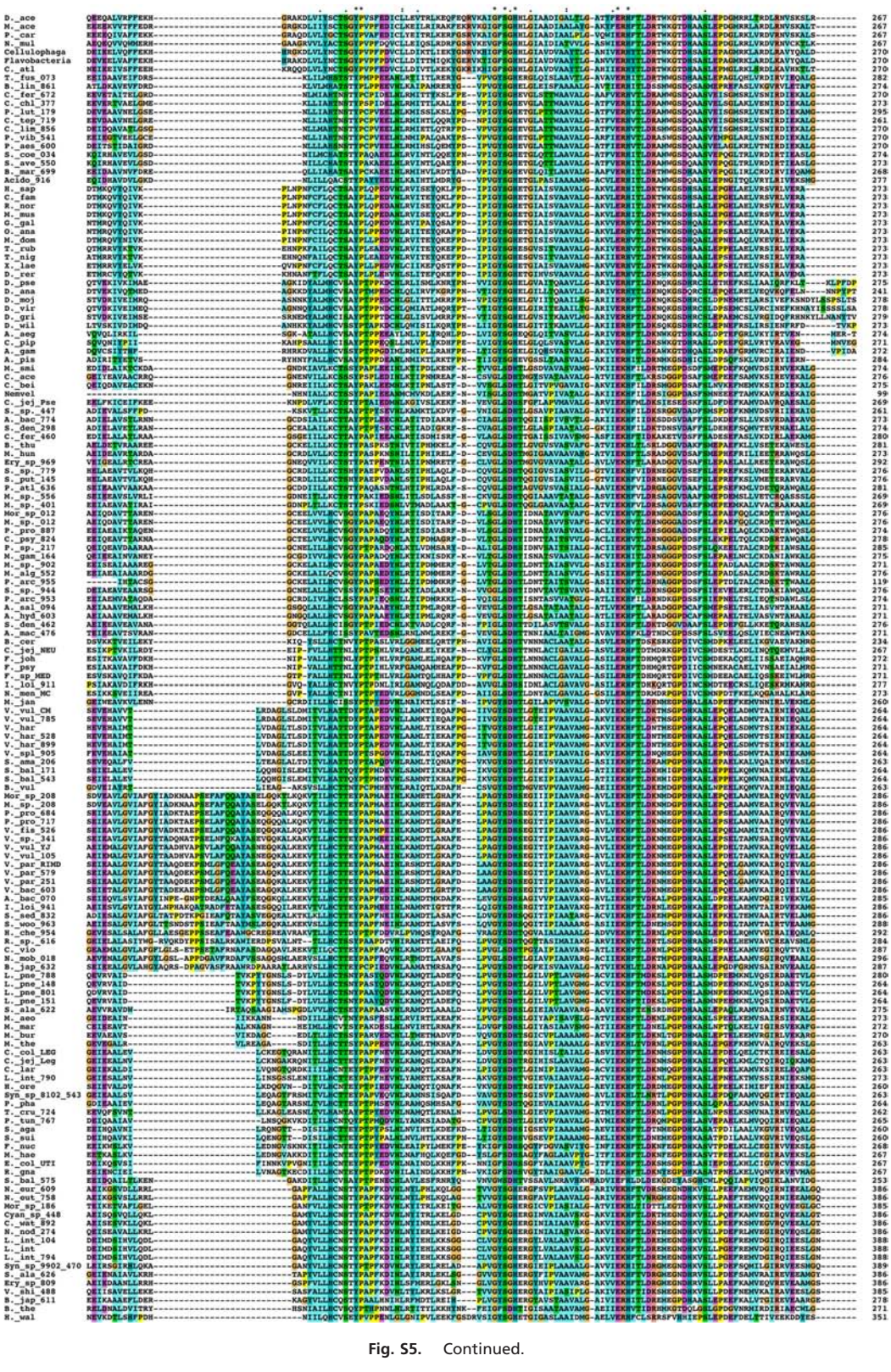


Fig. S5. Continued.

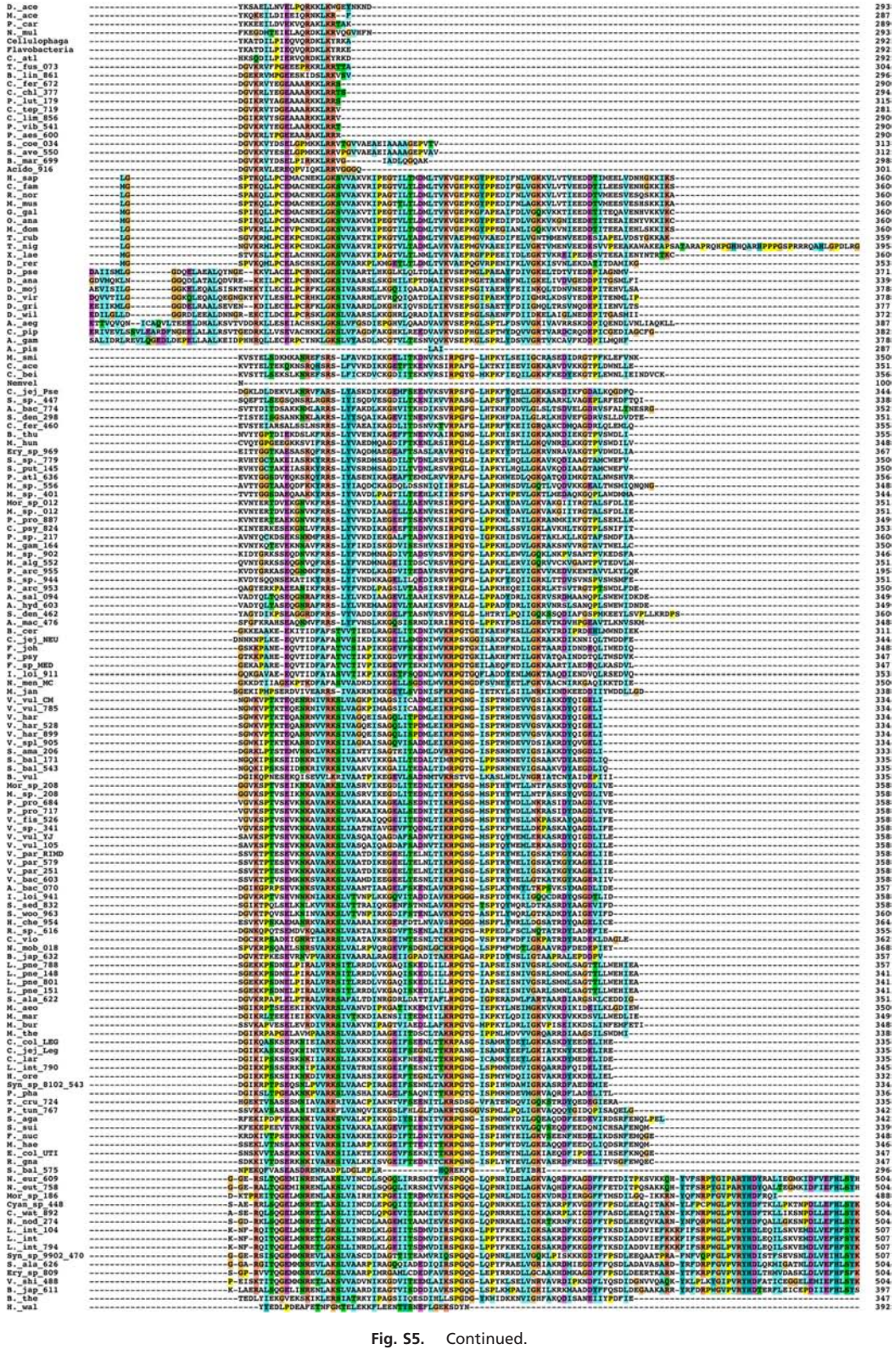


Fig. S5. Continued.

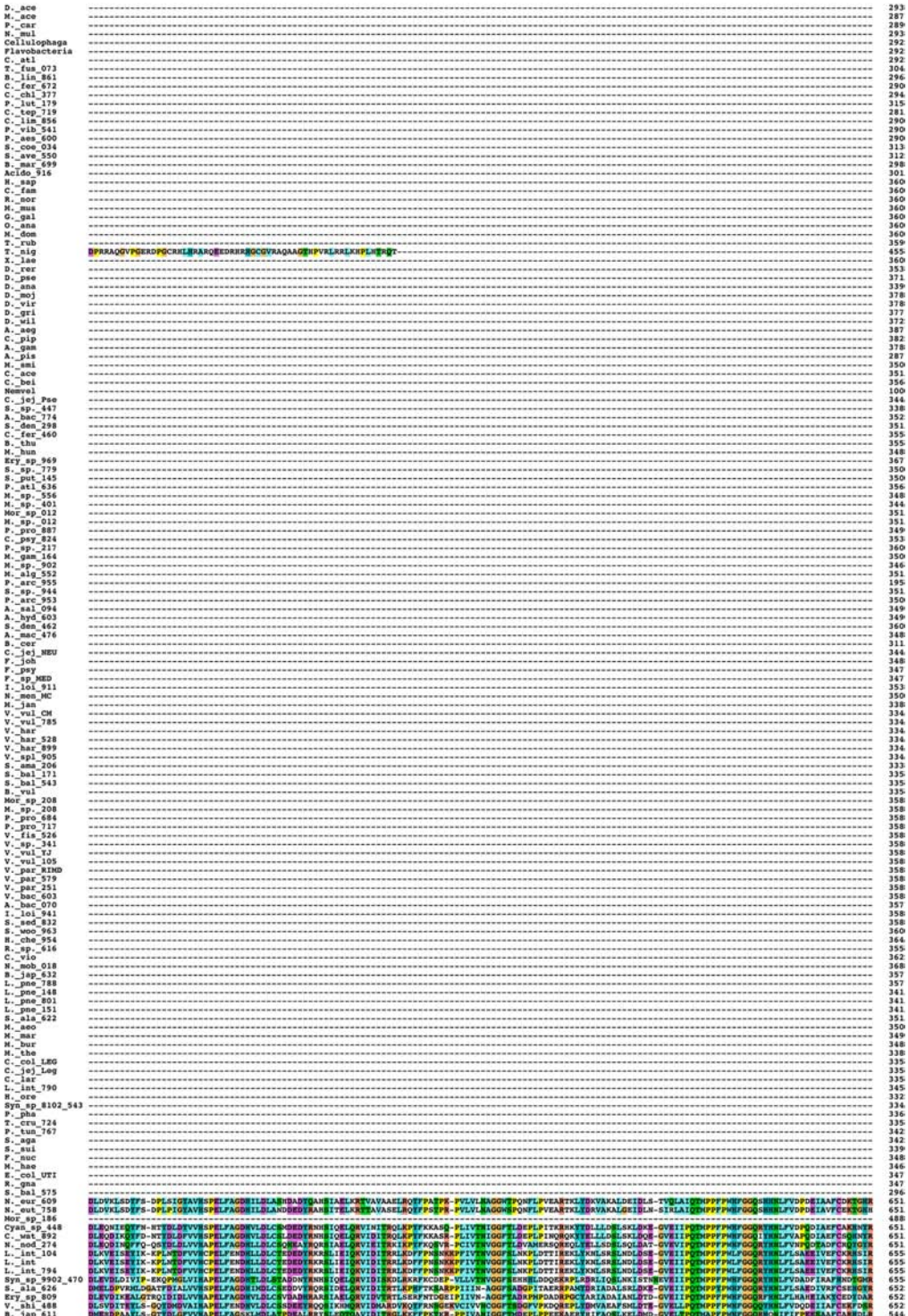


Fig. S5. Continued.

Other Supporting Information

Table S1

Table S1. Distribution of NAB homologs among sequenced bacteria and archaea

NCBI accession numbers for NAB-1 and NAB-2 homologs are color coded according to Fig. 3A predictions, **Neu**; **Leg**; **Pse**; **'animal-like'**, Strains are in alphabetical order with NAB-negative genomes enclosed by double lines below each NCBI class of NAB-positive genomes.

ε-Proteobacteria	**NAB-1**			**NAB-2**		
Arcobacter butzleri RM4018	YP_001491128			YP_001491129		
Caminibacter mediatlanticus TB-2	ZP_01871752	ZP_01872215		ZP_01871757	ZP_01872218	
Campylobacter coli RM2228	ZP_00367365	ZP_00367381	ZP_00368095	ZP_00367371	ZP_00367377	
Campylobacter concisus 13826	YP_001467628				YP_001467632	
Campylobacter curvus 525.92	YP_001408965			YP_001408968		
Campylobacter fetus subsp. fetus 82-40	YP_892682			YP_892689		
Campylobacter jejuni RM1221	YP_179491	YP_179505		YP_179497	YP_179501	
Campylobacter jejuni subsp. doylei 269.97	YP_001397604	YP_001397584	YP_001397865	YP_001397596	YP_001397590	
Campylobacter jejuni subsp. jejuni 260.94	ZP_01069032	ZP_01069114		ZP_01068970		ZP_01069267
Campylobacter jejuni subsp. jejuni 81116	YP_001482908			YP_001482827		YP_001482914
Campylobacter jejuni subsp. jejuni 81-176	YP_001000986	ZP_02271650	YP_001000821	YP_001000992	ZP_02271488	YP_001000818
Campylobacter jejuni subsp. jejuni 84-25	ZP_01099325	ZP_01099767	ZP_01099742	ZP_01099311	ZP_01099299	ZP_01099657
Campylobacter jejuni subsp. jejuni CF93-6	ZP_01068249	ZP_01068280	ZP_01068549	ZP_01068260	ZP_01068224	ZP_01068514
Campylobacter jejuni subsp. jejuni CG8486	ZP_01810694	ZP_01810708		ZP_01810700	ZP_01810704	ZP_01809838
Campylobacter jejuni subsp. jejuni HB93-13	ZP_01070901		ZP_01070836	ZP_01070873		ZP_01071011
Campylobacter jejuni subsp. jejuni NCTC 11168	NP_282457	NP_282477	NP_282291	NP_282463	NP_282473	NP_282289
Campylobacter lari RM2100	ZP_00369430	ZP_00369519		ZP_00368650	ZP_00369513	
Campylobacter upsaliensis RM3195	ZP_00371764	ZP_00371495	ZP_00371144	ZP_00370335	ZP_00371499	
Helicobacter acinonychis str. Sheeba	YP_665018	YP_664764		YP_664200		YP_665020
Helicobacter hepaticus ATCC 51449	NP_860431			NP_860439		
Helicobacter pylori 26695	NP_207124			NP_206977		
Helicobacter pylori HPAG1	YP_627070			YP_626916		
Helicobacter pylori J99				NP_222887		
Sulfurimonas denitrificans DSM 1251	YP_393113				YP_393109	
Wolinella succinogenes DSM 1740	NP_908204			NP_908207		
Campylobacter hominis ATCC BAA-381	*Sulfurovum sp. NBC37-1*					
Nitratiruptor sp. SB155-2	*Thiomicrospira denitrificans ATCC 33889*					
Bacteroidetes	**NAB-1**			**NAB-2**		
Algoriphagus sp. PR1	ZP_01717386			ZP_01717383		
Bacteroides capillosus ATCC 29799	ZP_02036000					
Bacteroides fragilis NCTC 9343	YP_211444					
Bacteroides fragilis YCH46	YP_101321	YP_099504	YP_099012			
Bacteroides ovatus ATCC 8483	ZP_02066715					
Bacteroides thetaiotaomicron VPI-5482	NP_810626.1			NP_810627		
Bacteroides uniformis ATCC 8492	ZP_02070812	ZP_02071546	ZP_02071964	ZP_02071542	ZP_02070811	
Bacteroides vulgatus ATCC 8482	YP_001301263	YP_001301261	YP_001299934	YP_001301264		
Cellulophaga sp. MED134	ZP_01051503	ZP_01049060		ZP_01051504		
Chlorobium chlorochromatii CaD3	YP_379437	YP_378636				YP_379377

chlorobium ferrooxidans DSM 13031	ZP_01385741	ZP_01385461	ZP_01385231	ZP_01385460		ZP_01385672
Chlorobium limicola DSM 245	ZP_00511412	ZP_00513057				ZP_00512609
Chlorobium phaeobacteroides DSM 266	YP_911960	YP_912515				YP_911647
Chlorobium tepidum TLS	NP_662044	NP_662709				NP_661719
Croceibacter atlanticus HTCC2559	ZP_00950410	ZP_00951263	ZP_00950767		ZP_00950411	
Cytophaga hutchinsonii ATCC 33406	YP_679353				YP_679350	
Flavobacteria bacterium BAL38	ZP_01733186	ZP_01735244		ZP_01733184		
Flavobacteria bacterium BBFL7	ZP_01202896	ZP_01201921	ZP_01203049	ZP_01202899	ZP_01201920	
Flavobacteriales bacterium HTCC2170	ZP_01105687	ZP_01106328		ZP_01105691		
Flavobacterium johnsoniae UW101	YP_001192662	YP_001196831		YP_001192660		
Flavobacterium psychrophilum JIP02/86	YP_001296151	YP_001295888		YP_001296150		
Flavobacterium sp. MED217	ZP_01058929	ZP_01061526		ZP_01058928		
Gramella forsetii KT0803	YP_860622	YP_862055	YP_861860		YP_860619	
Pedobacter sp. BAL39	ZP_01883717				ZP_01883719	
Pelodictyon luteolum DSM 273	YP_374880					YP_375179
Pelodictyon phaeoclathratiforme BU-1	ZP_00590637	ZP_00588167	ZP_00588267	ZP_00588274	ZP_01717383	ZP_00588729
Polaribacter irgensii 23-P	ZP_01118271					
Prosthecochloris aestuarii DSM 271	ZP_00592355	ZP_00590933				ZP_00591854
Prosthecochloris vibrioformis DSM 265	YP_001130443	YP_001129928				YP_001130541
Psychroflexus torquis ATCC 700755	ZP_01254358	ZP_01253608				
Robiginitalea biformata HTCC2501	ZP_01120753					
Salinibacter ruber DSM 13855	YP_444751			YP_444748		
Tenacibaculum sp. MED152	ZP_01052754	ZP_01052336		ZP_01052757		
unidentified eubacterium SCB49	ZP_01889752	ZP_01890888				

Bacteroides caccae ATCC 43185	*Chlorobium phaeobacteroides BS1*	*Parabacteroides merdae ATCC 43184*
Candidatus Sulcia muelleri str. Hc (Homalodisca coagulata)	*Parabacteroides distasonis ATCC 8503*	*Porphyromonas gingivalis W83*

Cyanobacteria	**NAB-1**			**NAB-2**		
Acaryochloris marina MBIC11017	YP_001516521					
Anabaena variabilis ATCC 29413	YP_323517	YP_323463				
Crocosphaera watsonii WH 8501	ZP_00517669			ZP_00514892		
Cyanothece sp. CCY0110-	ZP_01729244			ZP_01731448		
Gloeobacter violaceus PCC 7421	NP_926721					
Lyngbya sp. PCC 8106	ZP_01622329	ZP_01622482				
Nodularia spumigena CCY9414	ZP_01631002	ZP_01631569		ZP_01632274		
Nostoc punctiforme PCC 73102	ZP_00108272	ZP_00110422		ZP_00109033		
Nostoc sp. PCC 7120	NP_485331	NP_485019				
Prochlorococcus marinus str. AS9601	YP_001009733			YP_001009815		
Prochlorococcus marinus str. MIT 9211	YP_001551127	YP_001551172		YP_001551130		
Prochlorococcus marinus str. MIT 9215	YP_001484662	YP_001484645		YP_001484680	YP_001484665	
Prochlorococcus marinus str. MIT 9301	YP_001091637	YP_001091645		YP_001091632		
Prochlorococcus marinus str. MIT 9303	YP_001016135			YP_001016130		

Prochlorococcus marinus str. MIT 9312	YP_397845	YP_397837	YP_397842	YP_397836
Prochlorococcus marinus str. MIT 9313	NP_893935		NP_893939	
Prochlorococcus marinus str. NATL1A	YP_001014697		YP_001014695	
Synechococcus elongatus PCC 7942	YP_401307			
Synechococcus sp. BL107	ZP_01468600	ZP_01468925	ZP_01469027	ZP_01468604
Synechococcus sp. CC9311	YP_729407	YP_729419	YP_729411	YP_729418
Synechococcus sp. CC9605	YP_382511		YP_382515	
Synechococcus sp. CC9902-	YP_376474	YP_376118	YP_376470	YP_376102
Synechococcus sp. RS9916	ZP_01471749		ZP_01471755	
Synechococcus sp. RS9917	ZP_01078981		ZP_01078980	
Synechococcus sp. WH 8102-	NP_896493	NP_896539	NP_896543	NP_896489
Synechocystis sp. PCC 6803	NP_441379		NP_441367	
Trichodesmium erythraeum IMS101	YP_723938		YP_723990	
Leptolyngbya valderiana BDU 20041	*Synechococcus elongatus PCC 6301*		*Synechococcus sp. WH 5701*	
Prochlorococcus marinus str. NATL2A	*Synechococcus sp. JA-2-3B'a(2-13)*		*Synechococcus sp. WH 7803*	
Prochlorococcus marinus ss. marinus CCMP1375	*Synechococcus sp. JA-3-3Ab*		*Synechococcus sp. WH 7805*	
Prochlorococcus marinus ss. pastoris CCMP1986	*Synechococcus sp. RCC307*		*Thermosynechococcus elongatus BP-1*	
δ-Proteobacteria	**NAB-1**		**NAB-2**	
Bdellovibrio bacteriovorus HD100	NP_968563		NP_968560	
delta proteobacterium MLMS-1	ZP_01287684		ZP_01288484	
Desulfovibrio desulfuricans G20	YP_389746		YP_389745	YP_390169
Desulfovibrio vulgaris subsp. vulgaris DP4	YP_968030		YP_968032	YP_968074
Desulfovibrio vulgaris ss. vulgaris Hildenborough			YP_009573	YP_012219
Desulfuromonas acetoxidans DSM 684	ZP_01312106	YP_386919	ZP_01312107	ZP_01311516
Geobacter bemidjiensis Bem			ZP_01773432	
Geobacter lovleyi SZ	YP_001951964			
Geobacter metallireducens GS-15	YP_384236		YP_383423	
Geobacter sulfurreducens PCA	NP_953021		NP_953019	
Geobacter uraniireducens Rf4	YP_001232805		YP_001232802	
Lawsonia intracellularis PHE/MN1-00	YP_595364			YP_965814
Myxococcus xanthus DK 1622	YP_629360			
Pelobacter carbinolicus DSM 2380	YP_356697		YP_356696	YP_356559
Syntrophus aciditrophicus SB	YP_463211			
Anaeromyxobacter dehalogenans 2CP-C	*Desulfotalea psychrophila LSv54*		*Plesiocystis pacifica SIR-1*	
Anaeromyxobacter sp. Fw109-5	*Desulfovibrio vulgaris subsp. vulgaris DP4*		*Stigmatella aurantiaca DW4/3-1*	
Candidatus Desulfococcus oleovorans Hxd3	*Pelobacter propionicus DSM 2379*		*Syntrophobacter fumaroxidans MPOB*	
α-Proteobacteria	**NAB-1**		**NAB-2**	
Bradyrhizobium japonicum USDA 110	NP_772636	NP_772612	NP_772632	NP_772611
Erythrobacter sp. NAP1	ZP_01040808	ZP_01038973	ZP_01040809	ZP_01038969
Fulvimarina pelagi HTCC2506	ZP_01440484		ZP_01440497	
Loktanella vestfoldensis SKA53	ZP_01002614		ZP_01002612	
Magnetospirillum magneticum AMB-1	YP_419451	YP_419454	YP_419443	YP_419455
Magnetospirillum magnetotacticum MS-1	ZP_00056632.2|		ZP_00056630	ZP_00054178
Methylobacterium chloromethanicum CM4	ZP_02059631		ZP_02059639	YP_420076
Nitrobacter sp. Nb-311A	ZP_01045625		ZP_01045622	ZP_01045506
Nitrobacter winogradskyi Nb-255	YP_319003		YP_319002	YP_317265

Oceanicaulis alexandrii HTCC2633	ZP_00952361		ZP_00952353	
Paracoccus denitrificans PD1222	YP_913900	YP_917316	YP_917322	
Rhodopseudomonas palustris BisA53	YP_783144		YP_783141	
Rhodopseudomonas palustris HaA2	YP_485157		YP_485155	
Roseobacter denitrificans OCh 114	YP_684331		YP_684329	
Roseobacter sp. AzwK-3b	ZP_01901532	ZP_01901542	ZP_01901536	ZP_01901525
Roseobacter sp. CCS2	ZP_01750795		ZP_01749990	ZP_01750793
Roseobacter sp. SK209-2-6	ZP_01756746	ZP_01756754	ZP_01756749	ZP_01756734
Sphingopyxis alaskensis RB2256	YP_616624	YP_616627	YP_616622	YP_616626
Stappia aggregata IAM 12614	ZP_01545477	ZP_01550520	ZP_01545475	

Sinorhizobium meliloti 1021	*Methylobacterium extorquens PA1*	*Rickettsia felis URRWXCal2*
Agrobacterium tumefaciens str. C58	*Methylobacterium sp. 4-46*	*Rickettsia massiliae MTU5*
alpha proteobacterium HTCC2255	*Neorickettsia sennetsu str. Miyayama*	*Rickettsia prowazekii str. Madrid E*
Anaplasma marginale str. St. Maries	*Nitrobacter hamburgensis X14*	*Rickettsia rickettsii str. 'Sheila Smith'*
Anaplasma phagocytophilum HZ	*Novosphingobium aromaticivorans DSM 12444*	*Rickettsia sibirica 246*
Bartonella bacilliformis KC583	*Oceanicola batsensis HTCC2597*	*Rickettsia typhi str. Wilmington*
Bartonella henselae str. Houston-1	*Oceanicola granulosus HTCC2516*	*Roseobacter sp. MED193*
Bartonella quintana str. Toulouse	*Ochrobactrum anthropi ATCC 49188*	*Roseovarius nubinhibens ISM*
Brucella abortus biovar 1 str. 9-941	*Orientia tsutsugamushi Boryong*	*Roseovarius sp. 217*
Brucella melitensis 16M	*Parvibaculum lavamentivorans DS-1*	*Roseovarius sp. HTCC2601*
Brucella melitensis biovar Abortus 2308	*Parvularcula bermudensis HTCC2503*	*Roseovarius sp. TM1035*
Brucella ovis ATCC 25840	*Rhizobium etli CFN 42*	*Sagittula stellata E-37*
Brucella suis 1330	*Rhizobium leguminosarum bv. viciae 3841*	*Silicibacter pomeroyi DSS-3*
Candidatus Pelagibacter ubique HTCC1002	*Rhodobacter sphaeroides 2.4.1*	*Silicibacter sp. TM1040*
Candidatus Pelagibacter ubique HTCC1062	*Rhodobacter sphaeroides ATCC 17025*	*Sinorhizobium medicae WSM419*
Caulobacter crescentus CB15	*Rhodobacter sphaeroides ATCC 17029*	*Sphingomonas sp. SKA58*
Caulobacter sp. K31	*Rhodobacterales bacterium HTCC2150*	*Sphingomonas wittichii RW1*
Ehrlichia canis str. Jake	*Rhodobacterales bacterium HTCC2654*	*Sulfitobacter sp. EE-36*
Ehrlichia chaffeensis str. Arkansas	*Rhodopseudomonas palustris BisB18*	*Sulfitobacter sp. NAS-14.1*
Ehrlichia chaffeensis str. Sapulpa	*Rhodopseudomonas palustris BisB5*	*Wolbachia endosymbiont of Drosophila ananassae*
Ehrlichia ruminantium str. Gardel	*Rhodopseudomonas palustris CGA009*	*Wolbachia endosymbiont of Drosophila melanogaster*
Ehrlichia ruminantium str. Welgevonden	*Rhodospirillum rubrum ATCC 11170*	*Wolbachia endosymbiont of Drosophila simulans*
Granulibacter bethesdensis CGDNIH1	*Rickettsia africae ESF-5*	*Wolbachia endosymbiont of Drosophila willistoni TSC#14030-0811.24*
Hyphomonas neptunium ATCC 15444	*Rickettsia akari str. Hartford*	*Wolbachia endosymbiont strain TRS of Brugia malayi*
Jannaschia sp. CCS1	*Rickettsia bellii OSU 85-389*	*Xanthobacter autotrophicus Py2*
Maricaulis maris MCS10	*Rickettsia bellii RML369-C*	*Zymomonas mobilis subsp. mobilis ZM4*
Mesorhizobium loti MAFF303099]	*Rickettsia canadensis str. McKiel*	
Mesorhizobium sp. BNC1	*Rickettsia conorii str. Malish 7*	

γ-Proteobacteria	**NAB-1**	**NAB-2**
Aeromonas hydrophila subsp. hydrophila ATCC 7966	YP_858601	YP_858603
Aeromonas salmonicida subsp. salmonicida A449	YP_001140096	YP_001140094
Alkalilimnicola ehrlichei MLHE-1	YP_743154	YP_743157
Alteromonadales bacterium TW-7	ZP_01612074	ZP_01612070
Alteromonas macleodii 'Deep ecotype'	ZP_01110485	ZP_01110476
Azotobacter vinelandii AvOP	ZP_00416273	
Colwellia psychrerythraea 34H	YP_268821	YP_268824

Escherichia coli APEC O1	YP_854393		YP_854394	
Escherichia coli UTI89	YP_542349		YP_542350	
Francisella tularensis subsp. holarctica			YP_513685	
Haemophilus ducreyi 35000HP	NP_873215			
Haemophilus influenzae 22.1-21	ZP_01785141			
Haemophilus influenzae 22.4-21	ZP_01786684			
Haemophilus influenzae 3655	ZP_01788996			
Haemophilus influenzae 86-028NP	YP_249309			
Haemophilus influenzae PittAA	ZP_01790101			
Haemophilus influenzae PittEE	YP_001290615			
Haemophilus influenzae PittGG	YP_001291778			
Haemophilus influenzae PittHH	ZP_01792523			
Haemophilus influenzae PittII	ZP_01794452			
Haemophilus influenzae R2846	ZP_00154934.2|			
Haemophilus influenzae R2866	ZP_00157363.2|			
Haemophilus influenzae R3021	ZP_01797675	ZP_01797708		
Haemophilus influenzae Rd KW20	NP_439432			
Haemophilus somnus 129PT	YP_718918			
Haemophilus somnus 2336	YP_001784447			
Hahella chejuensis KCTC 2396	YP_435951	YP_436364	YP_435954	YP_436365
Idiomarina loihiensis L2TR	YP_154910	YP_154944	YP_154941	YP_154911
Legionella pneumophila str. Corby	YP_001251802		YP_001251801	YP_001251785
Legionella pneumophila str. Lens	YP_126150		YP_126151	YP_126168
Legionella pneumophila str. Paris	YP_123147		YP_123148	YP_123162
Legionella pneumophila subsp. pneumophila str. Philadelphia 1	YP_094787		YP_094788	YP_094804
marine gamma proteobacterium HTCC2207			ZP_01225164	
Marinobacter algicola DG893			ZP_01894552	
Marinobacter sp. ELB17	ZP_01738899		ZP_01738902	
Marinomonas sp. MED121	ZP_01075554		ZP_01075556	
Marinomonas sp. MWYL1	YP_001342399		YP_001342401	
Moritella sp. PE36	ZP_01898014	ZP_01896204&187	ZP_01896208	ZP_01896186
Nitrococcus mobilis Nb-231	ZP_01127539	ZP_01126021	ZP_01126018	ZP_01126018
Oceanobacter sp. RED65	ZP_01306785		ZP_01306787	
Oceanospirillum sp. MED92	ZP_01167341		ZP_01167812	ZP_01167342
Pasteurella multocida subsp. multocida str. Pm70	NP_245124			
Photobacterium profundum 3TCK	ZP_01218688	ZP_01218721	ZP_01218684	ZP_01218717
Photobacterium profundum SS9	YP_130889		YP_130887	
Pseudoalteromonas atlantica T6c	YP_662638		YP_662636	
Pseudoalteromonas tunicata D2	ZP_01132761		ZP_01132767	
Pseudomonas entomophila L48	YP_607270			
Pseudomonas fluorescens Pf-5	YP_258749		YP_258751	
Pseudomonas fluorescens PfO-1	YP_347253		YP_347255	
Pseudomonas putida F1	YP_001269217		YP_001266824	
Pseudomonas putida GB-1	YP_001667719			
Pseudomonas putida KT2440	NP_743946			
Pseudomonas putida W619	YP_001750560		YP_001750558	
Pseudomonas stutzeri A1501	YP_001174303		YP_001174300	

Psychrobacter arcticus 273-4	YP_263949		YP_263953	YP_263955
Psychromonas sp. CNPT3	ZP_01215215		ZP_01215217	
Reinekea sp. MED297	ZP_01114619		ZP_01114616	
Shewanella amazonensis SB2B	YP_928211		YP_928206	
Shewanella baltica OS155			YP_001051309	
Shewanella baltica OS185	YP_001367175		YP_001367171	
Shewanella baltica OS195	YP_001555547		YP_001555543	
Shewanella baltica OS223			ZP_01840575	
Shewanella denitrificans OS217	YP_562297		YP_562298	YP_564113
Shewanella frigidimarina NCIMB 400	YP_751516			
Shewanella loihica PV-4	YP_001093456		YP_001093457	
Shewanella oneidensis MR-1			NP_718815	
Shewanella pealeana ATCC 700345	YP_001499916		YP_001499913	
Shewanella putrefaciens CN-32	YP_001184147		YP_001184145	
Shewanella sediminis HAW-EB3	YP_001474836		YP_001474832	
Shewanella sp. ANA-3	YP_868949		YP_868944	
Shewanella sp. MR-7	YP_737443		YP_737447	
Shewanella sp. W3-18-1	YP_962777		YP_962779	
Shewanella woodyi ATCC 51908	YP_001759959		YP_001759963	
Thiomicrospira crunogena XCL-2	YP_391722		YP_391724	
Vibrio fischeri ES114	YP_203530		YP_203526	
Vibrio harveyi ATCC BAA-1116	YP_001443906		YP_001443899	
Vibrio harveyi HY01	ZP_01987522		ZP_01987528	
Vibrio parahaemolyticus AQ3810	ZP_01993064	ZP_01992254	ZP_01992251	
Vibrio parahaemolyticus RIMD 2210633	NP_796582		NP_796579	
Vibrio shilonii AK1	ZP_01867489		ZP_01867488	
Vibrio sp. Ex25	ZP_01475338	ZP_01475320	ZP_01475341	
Vibrio splendidus 12B01	ZP_00989910		ZP_00989905	
Vibrio vulnificus CMCP6	NP_759780		NP_759785	
Vibrio vulnificus YJ016	NP_933109		NP_933105	
Vibrionales bacterium SWAT-3	ZP_01813606		ZP_01813603	

Acinetobacter baumannii ATCC 17978	*Francisella tularensis ss. novicida GA99-3549*	*Shewanella putrefaciens 200*
Acinetobacter sp. ADP1	*Francisella tularensis ss. novicida U112*	*Shewanella sp. MR-4*
Actinobacillus pleuropneumoniae serovar 1 str. 4074	*Francisella tularensis ss. tularensis FSC033*	*Shigella boydii CDC 3083-94*
Actinobacillus succinogenes 130Z	*Francisella tularensis subsp. tularensis FSC198*	*Shigella boydii Sb227*
Aeromonas salmonicida subsp. salmonicida A449	*Francisella tularensis ss. tularensis SCHU S4*	*Shigella dysenteriae 1012*
Alcanivorax borkumensis SK2	*Francisella tularensis ss. tularensis WY96-3418*	*Shigella dysenteriae Sd197*
Baumannia cicadellinicola str. Hc (Homalodisca coagulata)	*gamma proteobacterium KT 71*	*Shigella flexneri 2a str. 2457T*
Beggiatoa sp. PS	*Halorhodospira halophila SL1*	*Shigella flexneri 2a str. 301*
Buchnera aphidicola str. APS (Acyrthosiphon pisum)	*Idiomarina baltica OS145*	*Shigella flexneri 5 str. 8401*
Buchnera aphidicola str. Bp (Baizongia pistaciae)	*Klebsiella pneumoniae subsp. pneumoniae MGH 78578*	*Shigella sonnei Ss046*
Buchnera aphidicola str. Cc (Cinara cedri)	*Mannheimia haemolytica PHL213*	*Sodalis glossinidius str. 'morsitans'*
Buchnera aphidicola str. Sg (Schizaphis graminum)	*Mannheimia succiniciproducens MBEL55E*	*Vibrio alginolyticus 12G01*
Candidatus Blochmannia floridanus	*marine gamma proteobacterium HTCC2080*	*Vibrio angustum S14*
Candidatus Blochmannia pennsylvanicus str. BPEN	*marine gamma proteobacterium HTCC2143*	*Vibrio cholerae 1587*
Candidatus Carsonella ruddii PV	*Marinobacter aquaeolei VT8*	*Vibrio cholerae 2740-80*

Candidatus Ruthia magnifica str. Cm (Calyptogena magnifica)	*Methylococcus capsulatus str. Bath*	*Vibrio cholerae 623-39*
Candidatus Vesicomyosocius okutanii HA	*Neptuniibacter caesariensis*	*Vibrio cholerae AM-19226*
Chromohalobacter salexigens DSM 3043	*Nitrosococcus oceani ATCC 19707*	*Vibrio cholerae B33*
Citrobacter koseri ATCC BAA-895	*Photobacterium sp. SKA34*	*Vibrio cholerae MAK 757*
Coxiella burnetii Dugway 5J108-111	*Photorhabdus luminescens subsp. laumondii TTO1*	*Vibrio cholerae MO10*
Coxiella burnetii Dugway 7E9-12	*Pseudoalteromonas haloplanktis TAC125*	*Vibrio cholerae MZO-2*
Coxiella burnetii 'MSU Goat Q177'	*Pseudomonas aeruginosa 2192*	*Vibrio cholerae MZO-3*
Coxiella burnetii RSA 331	*Pseudomonas aeruginosa C3719*	*Vibrio cholerae NCTC 8457*
Coxiella burnetii RSA 334	*Pseudomonas aeruginosa PA7*	*Vibrio cholerae O1 biovar eltor str. N16961*
Coxiella burnetii RSA 493	*Pseudomonas aeruginosa PACS2*	*Vibrio cholerae O395*
ctinobacillus pleuropneumoniae L20	*Pseudomonas aeruginosa PAO1*	*Vibrio cholerae RC385*
Dichelobacter nodosus VCS1703A	*Pseudomonas aeruginosa UCBPP-PA14*	*Vibrio cholerae V51*
Endoriftia persephone 'Hot96_1+Hot96_2'	*Pseudomonas mendocina ymp*	*Vibrio cholerae V52*
Enterobacter sakazakii ATCC BAA-894	*Pseudomonas syringae pv. phaseolicola 1448A*	*Vibrio sp. MED222*
Enterobacter sp. 638	*Pseudomonas syringae pv. syringae B728a*	*Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis*
Erwinia carotovora subsp. atroseptica SCRI1043	*Pseudomonas syringae pv. tomato str. DC3000*	*Yersinia bercovieri ATCC 43970*
Escherichia coli 101-1	*Psychrobacter cryohalolentis K5*	*Yersinia enterocolitica subsp. enterocolitica 8081*
Escherichia coli 536	*Psychrobacter sp. PRwf-1*	*Yersinia frederiksenii ATCC 33641*
Escherichia coli 53638	*Psychromonas ingrahamii 37*	*Yersinia intermedia ATCC 29909*
Escherichia coli B	*Rickettsiella grylli*	*Yersinia mollaretii ATCC 43969*
Escherichia coli B171	*Saccharophagus degradans 2-40*	*Yersinia pestis Angola*
Escherichia coli B7A	*Salmonella enterica subsp. enterica serovar 4,[5],12:i:- str. CVM23701*	*Yersinia pestis Antiqua*
Escherichia coli CFT073	*S. enterica subsp. enterica serovar Agona str. SL483*	*Yersinia pestis biovar Antiqua str. B42003004*
Escherichia coli E110019	*S. enterica subsp. enterica serovar Choleraesuis str. SC-B67*	*Yersinia pestis biovar Antiqua str. E1979001*
Escherichia coli E22	*S. enterica subsp. enterica serovar Dublin str. CT_02021853*	*Yersinia pestis biovar Antiqua str. UG05-0454*
Escherichia coli E24377A	*S. enterica subsp. enterica serovar Heidelberg str. SL476*	*Yersinia pestis biovar Mediaevalis str. K1973002*
Escherichia coli F11	*S. enterica subsp. enterica serovar Javiana str. GA_MM04042433*	*Yersinia pestis biovar Microtus str. 91001*
Escherichia coli HS	*S. enterica subsp. enterica serovar Kentucky str. CDC 191*	*Yersinia pestis biovar Microtus str. 91001*
Escherichia coli O157:H7 EDL933	*S. enterica subsp. enterica serovar Kentucky str. CVM29188*	*Yersinia pestis biovar Orientalis str. F1991016*
Escherichia coli O157:H7 str. Sakai	*S. enterica subsp. enterica serovar Newport str. SL254*	*Yersinia pestis biovar Orientalis str. IP275*
Escherichia coli SECEC SMS-3-5	*S. enterica subsp. enterica serovar Newport str. SL317*	*Yersinia pestis biovar Orientalis str. MG05-1020*
Escherichia coli str. K-12 substr. MG1655	*S. enterica subsp. enterica serovar Paratyphi A str. ATCC 9150*	*Yersinia pestis CA88-4125*
Escherichia coli W3110	*S. enterica subsp. enterica serovar Saintpaul str. SARA23*	*Yersinia pestis CO92*
Francisella tularensis subsp. holarctica 257	*S. enterica subsp. enterica serovar Saintpaul str. SARA29*	*Yersinia pestis FV-1*
Francisella tularensis subsp. holarctica FSC022	*S. enterica subsp. enterica serovar Schwarzengrund str. CVM19633*	*Yersinia pestis KIM*
Francisella tularensis subsp. holarctica FSC200	*S. enterica subsp. enterica serovar Schwarzengrund str. SL480*	*Yersinia pestis Nepal516*
Francisella tularensis subsp. holarctica FTA	*S. enterica subsp. enterica serovar Typhi str. CT18*	*Yersinia pestis Pestoides F*
Francisella tularensis subsp. holarctica FTNF002-00	*S. enterica subsp. enterica serovar Typhi Ty2*	*Yersinia pseudotuberculosis IP 31758*
Francisella tularensis subsp. holarctica OSU18	*Salmonella typhimurium LT2*	*Yersinia pseudotuberculosis IP 3295*
Francisella tularensis subsp. novicida GA99-3548	*Serratia proteamaculans 568*	

β-Proteobacteria	**NAB-1**	**NAB-2**

Burkholderia cenocepacia MC0-3	YP_001765854				
Burkholderia phymatum STM815	YP_001859225			YP_001859224	
Burkholderia pseudomallei S13	ZP_01329806				
Chromobacterium violaceum ATCC 12472	NP_903698			NP_903701	
Dechloromonas aromatica RCB	YP_284470			YP_284473	
Herminiimonas arsenicoxydans				YP_001099432	
Methylophilales bacterium HTCC2181	ZP_01552361	NP_841608		NP_841609	
Neisseria meningitidis FAM18	YP_974195			YP_974194	
Neisseria meningitidis MC58	NP_273133			NP_273132	
Nitrosomonas europaea ATCC 19718	NP_841608			NP_841609	
Nitrosomonas eutropha C91	YP_747757			YP_747758	
Nitrosospira multiformis ATCC 25196	YP_411110			YP_411108	
Ralstonia pickettii 12D				ZP_02007588	
Verminephrobacter eiseniae EF01-2	YP_999560			YP_999561	

Acidovorax avenae subsp. citrulli AAC00-1	*Burkholderia oklahomensis EO147*	*Burkholderia thailandensis E264*
Acidovorax sp. JS42	*Burkholderia phytofirmans PsJN*	*Burkholderia thailandensis TXDOH*
Azoarcus sp. BH72	*Burkholderia pseudomallei 1106a*	*Burkholderia ubonensis Bu*
Azoarcus sp. EbN1	*Burkholderia pseudomallei 1106b*	*Burkholderia vietnamiensis G4*
Bordetella bronchiseptica RB50	*Burkholderia pseudomallei 112*	*Burkholderia xenovorans LB400*
Bordetella parapertussis 12822	*Burkholderia pseudomallei 14*	*Comamonas testosteroni KF-1*
Bordetella pertussis Tohama I	*Burkholderia pseudomallei 1655*	*Delftia acidovorans SPH-1*
Burkholderia ambifaria AMMD	*Burkholderia pseudomallei 1710a*	*ethylibium petroleiphilum PM1*
Burkholderia ambifaria MC40-6	*Burkholderia pseudomallei 1710b*	*Janthinobacterium sp. Marseille*
Burkholderia cenocepacia AU 1054	*Burkholderia pseudomallei 305*	*Limnobacter sp. MED105*
Burkholderia cenocepacia HI2424	*Burkholderia pseudomallei 381*	*Methylobacillus flagellatus KT*
Burkholderia dolosa AUO158	*Burkholderia pseudomallei 406e*	*Neisseria gonorrhoeae FA 1090*
Burkholderia mallei 2002721280	*Burkholderia pseudomallei 668*	*Neisseria meningitidis Z2491*
Burkholderia mallei ATCC 10399	*Burkholderia pseudomallei 7894*	*Polaromonas naphthalenivorans CJ2*
Burkholderia mallei ATCC 23344	*Burkholderia pseudomallei 9*	*Polaromonas sp. JS666*
Burkholderia mallei FMH	*Burkholderia pseudomallei 91*	*Polynucleobacter sp. QLW-P1DMWA-1*
Burkholderia mallei GB8 horse 4	*Burkholderia pseudomallei B7210*	*Ralstonia eutropha H16*
Burkholderia mallei JHU	*Burkholderia pseudomallei BCC215*	*Ralstonia eutropha JMP134*
Burkholderia mallei NCTC 10229	*Burkholderia pseudomallei DM98*	*Ralstonia metallidurans CH34*
Burkholderia mallei NCTC 10247	*Burkholderia pseudomallei K96243*	*Ralstonia pickettii 12J*
Burkholderia mallei PRL-20	*Burkholderia pseudomallei NCTC 13177*	*Ralstonia solanacearum GMI1000*
Burkholderia mallei SAVP1	*Burkholderia pseudomallei Pasteur 52237*	*Ralstonia solanacearum UW551*
Burkholderia multivorans ATCC 17616	*Burkholderia sp. 383*	*Rhodoferax ferrireducens T118*
Burkholderia oklahomensis C6786	*Burkholderia thailandensis Bt4*	*Thiobacillus denitrificans ATCC 25259*

Firmicutes	**NAB-1**			**NAB-2**	
Alkaliphilus oremlandii OhILAs	YP_001514040			YP_001514037	
Bacillus amyloliquefaciens FZB42				YP_001423072	
Bacillus cereus G9241	ZP_00236340	ZP_00239762		ZP_00236342	
Bacillus pumilus SAFR-032	YP_001487128			ZP_01724639	YP_001488643
Bacillus sp. B14905	ZP_01724636			ZP_01724639	
Bacillus subtilis subsp. subtilis str. 168				NP_391666	
Bacillus thuringiensis serovar israelensis ATCC 35646				ZP_00741540	
Clostridium acetobutylicum ATCC 824					NP_348805
Clostridium beijerinckii NCIMB 8052	YP_001311347				YP_001311344
Clostridium botulinum A str. ATCC 19397	YP_0013849	YP_001384963			YP_001384922

	23			YP_001384962		
Clostridium botulinum A str. ATCC 3502	YP_001255180	YP_001255220		YP_001255219	YP_001255179	
Clostridium botulinum A str. Hall	YP_001388392	YP_001388433		YP_001388391	YP_001388432	
Clostridium botulinum F str. Langeland	YP_001392001	YP_001391979		YP_001392003	YP_001391978	
Clostridium difficile QCD-32g58	ZP_01801682					
Clostridium kluyveri DSM 555	YP_001395849	YP_001395014		YP_001395855	YP_001395492	
Clostridium novyi NT	YP_877856			YP_877857	YP_877964	
Clostridium tetani E88	NP_782308		NP_781273	NP_782309		
Clostridium thermocellum ATCC 27405	YP_001039031	YP_001038625		YP_001039033	YP_001038624	
Desulfitobacterium hafniense DCB-2				ZP_01372474		
Desulfitobacterium hafniense Y51	YP_519540			YP_520419		
Enterococcus faecium DO	ZP_00604822			ZP_00604816		
Geobacillus kaustophilus HTA426	YP_148973			YP_148976		
Halothermothrix orenii H 168				ZP_01188511	ZP_01188516	
Lactobacillus johnsonii NCC 533		NP_965637				
Lactococcus lactis ssp. cremoris SK1		YP_808727				
Lactococcus lactis subsp. lactis Il140		NP_266866				
Moorella thermoacetica ATCC 39073	YP_429619			YP_429617		
Ruminococcus gnavus ATCC 29149	ZP_02041060			ZP_02041061		
Streptococcus agalactiae 18RS21	ZP_00781880	ZP_00780163	ZP_00781966	ZP_00781900		
Streptococcus agalactiae 2603V/R	NP_688167	NP_688113		NP_688170		
Streptococcus agalactiae 515	ZP_00788998	ZP_00789247		ZP_00789031		
Streptococcus agalactiae A909	YP_329861	YP_329810		YP_329864		
Streptococcus agalactiae CJB111	ZP_00788538	ZP_00786919		ZP_00788547		
Streptococcus agalactiae COH1	ZP_00785707	ZP_00785426		ZP_00785697		
Streptococcus agalactiae H36B	ZP_00782978	ZP_00782941		ZP_00782986		
Streptococcus agalactiae NEM316	NP_735677	NP_735617		NP_735680		
Streptococcus mutans UA159		NP_721495				
Streptococcus suis 05ZYH33	YP_001197947			YP_001197944		
Streptococcus suis 89/1591	ZP_00875903	ZP_00875443		ZP_00875906		
Streptococcus suis 98HAH33	YP_001200143			YP_001200140		
Streptococcus thermophilus CNRZ1		YP_141453				
Streptococcus thermophilus LMD-9		YP_820445				
Streptococcus thermophilus LMG 18		YP_139528				
Syntrophomonas wolfei subsp. wolfei str. Goettingen				YP_752933		
Thermoanaerobacter pseudethanolicus ATCC 33223				YP_001664975		
Thermoanaerobacter sp. X514				YP_001663050		
Thermoanaerobacter tengcongensis MB4				NP_622664		
Thermosinus carboxydivorans Nor1	ZP_01667693					

Alkaliphilus metalliredigens QYMF	*Eubacterium dolichum DSM 3991*	*Pelotomaculum thermopropionicum SI*
Anaerostipes caccae DSM 14662	*Eubacterium siraeum DSM 15702*	*Peptostreptococcus micros ATCC 33270*
Bacillus anthracis str. A1055	*Eubacterium ventriosum ATCC 27560*	*Ruminococcus obeum ATCC 29174*
Bacillus anthracis str. A2012	*Exiguobacterium sibiricum 255-15*	*Ruminococcus torques ATCC 27756*
Bacillus anthracis str. Ames	*Faecalibacterium prausnitzii M21/2*	*Staphylococcus aureus RF122*
Bacillus anthracis str. 'Ames Ancestor'	*Geobacillus thermodenitrificans NG80-2*	*Staphylococcus aureus subsp. aureus COL*

Bacillus anthracis str. Australia 94	*Lactobacillus acidophilus NCFM*	*Staphylococcus aureus subsp. aureus JH1*
Bacillus anthracis str. CNEVA-9066	*Lactobacillus brevis ATCC 367*	*Staphylococcus aureus subsp. aureus JH9*
Bacillus anthracis str. Kruger B	*Lactobacillus casei ATCC 334*	*Staphylococcus aureus subsp. aureus MRSA252*
Bacillus anthracis str. Sterne	*Lactobacillus delbrueckii subsp. bulgaricus ATCC 11842*	*Staphylococcus aureus subsp. aureus MSSA476*
Bacillus anthracis str. Vollum	*Lactobacillus delbrueckii subsp. bulgaricus ATCC BAA-365*	*Staphylococcus aureus subsp. aureus Mu3*
Bacillus anthracis str. Western North America USA6153	*Lactobacillus gasseri ATCC 33323*	*Staphylococcus aureus subsp. aureus Mu50*
Bacillus anthracis Tsiankovskii-I	*Lactobacillus plantarum WCFS1*	*Staphylococcus aureus subsp. aureus MW2*
Bacillus cereus 03BB108	*Lactobacillus reuteri 100-23*	*Staphylococcus aureus subsp. aureus N315*
Bacillus cereus AH1134	*Lactobacillus reuteri F275*	*Staphylococcus aureus subsp. aureus NCTC 8325*
Bacillus cereus AH187	*Lactobacillus sakei subsp. sakei 23K*	*Staphylococcus aureus subsp. aureus str. Newman*
Bacillus cereus AH820	*Lactobacillus salivarius subsp. salivarius UCC118*	*Staphylococcus aureus subsp. aureus USA300*
Bacillus cereus ATCC 10987	*Lactococcus lactis subsp. cremoris MG1363*	*Staphylococcus epidermidis ATCC 12228*
Bacillus cereus ATCC 14579	*Leuconostoc mesenteroides subsp. mesenteroides ATCC 8293*	*Staphylococcus epidermidis RP62A*
Bacillus cereus B4264	*Listeria innocua Clip11262*	*Staphylococcus haemolyticus JCSC1435*
Bacillus cereus E33L	*Listeria monocytogenes 10403S*	*Staphylococcus saprophyticus subsp. saprophyticus ATCC 15305*
Bacillus cereus H3081.97	*Listeria monocytogenes EGD-e*	*Streptococcus gordonii str. Challis substr. CH1*
Bacillus cereus NVH0597-99	*Listeria monocytogenes F6900*	*Streptococcus pneumoniae D39*
Bacillus cereus subsp. cytotoxis NVH 391-98	*Listeria monocytogenes FSL F2-515*	*Streptococcus pneumoniae R6*
Bacillus cereus W	*Listeria monocytogenes FSL J1-175*	*Streptococcus pneumoniae SP11-BS70*
Bacillus clausii KSM-K16	*Listeria monocytogenes FSL J1-194*	*Streptococcus pneumoniae SP14-BS69*
Bacillus coagulans 36D1	*Listeria monocytogenes FSL J1-208*	*Streptococcus pneumoniae SP18-BS74*
Bacillus halodurans C-125	*Listeria monocytogenes FSL J2-003*	*Streptococcus pneumoniae SP19-BS75*
Bacillus licheniformis ATCC 14580	*Listeria monocytogenes FSL J2-064*	*Streptococcus pneumoniae SP23-BS72*
Bacillus sp. NRRL B-14911	*Listeria monocytogenes FSL J2-071*	*Streptococcus pneumoniae SP3-BS71*
Bacillus sp. SG-1	*Listeria monocytogenes FSL N1-017*	*Streptococcus pneumoniae SP6-BS73*
Bacillus thuringiensis serovar konkukian str. 97-27	*Listeria monocytogenes FSL N3-165*	*Streptococcus pneumoniae SP9-BS68*
Bacillus thuringiensis str. Al Hakam	*Listeria monocytogenes FSL R2-503*	*Streptococcus pneumoniae TIGR4*
Bacillus weihenstephanensis KBAB4	*Listeria monocytogenes HPB2262*	*Streptococcus pyogenes M1 GAS*
Caldicellulosiruptor saccharolyticus DSM 8903	*Listeria monocytogenes J0161*	*Streptococcus pyogenes M49 591*
Carboxydothermus hydrogenoformans Z-2901	*Listeria monocytogenes J2818*	*Streptococcus pyogenes MGAS10270*
Clostridium bolteae ATCC BAA-613	*Listeria monocytogenes LO28*	*Streptococcus pyogenes MGAS10394*
Clostridium botulinum A str. Hall	*Listeria monocytogenes str. 1/2a F6854*	*Streptococcus pyogenes MGAS10750*
Clostridium botulinum Bf	*Listeria monocytogenes str. 4b F2365*	*Streptococcus pyogenes MGAS2096*
Clostridium botulinum C str. Eklund	*Listeria monocytogenes str. 4b H7858*	*Streptococcus pyogenes MGAS315*
Clostridium botulinum G	*Listeria welshimeri serovar 6b str. SLCC5334*	*Streptococcus pyogenes MGAS5005*
Clostridium botulinum NCTC 2916	*Mesoplasma florum L1*	*Streptococcus pyogenes MGAS6180*
Clostridium botulinum str. Iwanei E	*Mycoplasma agalactiae PG2*	*Streptococcus pyogenes MGAS8232*
Clostridium butyricum 5521	*Mycoplasma capricolum subsp. capricolum ATCC 27343*	*Streptococcus pyogenes MGAS9429*
Clostridium cellulolyticum H10	*Mycoplasma gallisepticum R*	*Streptococcus pyogenes SSI-1*
Clostridium difficile 630	*Mycoplasma genitalium G37*	*Streptococcus pyogenes str. Manfredo*
Clostridium difficile QCD-66c26	*Mycoplasma hyopneumoniae 232*	*Streptococcus sanguinis SK36*
Clostridium leptum DSM 753	*Mycoplasma hyopneumoniae 7448*	*Streptococcus thermophilus CNRZ1066*
Clostridium perfringens ATCC 13124	*Mycoplasma hyopneumoniae J*	*Symbiobacterium thermophilum IAM 14863*
Clostridium perfringens B str. ATCC 3626	*Mycoplasma mobile 163K*	*Ureaplasma parvum serovar 1*
Clostridium perfringens C str. JGS1495	*Mycoplasma mycoides subsp. mycoides LC str. GM12*	*Ureaplasma parvum serovar 14*
Clostridium perfringens CPE str. F4969	*Mycoplasma mycoides subsp. mycoides SC str. PG1*	*Ureaplasma parvum serovar 3*

Clostridium perfringens E str. JGS1987	*Mycoplasma penetrans HF-2*	*Ureaplasma parvum serovar 3 str. ATCC 700970*
Clostridium perfringens NCTC 8239	*Mycoplasma pneumoniae M129*	*Ureaplasma parvum serovar 6*
Clostridium perfringens SM101	*Mycoplasma pulmonis UAB CTIP*	*Ureaplasma urealyticum serovar 10*
Clostridium perfringens str. 13	*Mycoplasma synoviae 53*	*Ureaplasma urealyticum serovar 11*
Clostridium phytofermentans ISDg	*Oceanobacillus iheyensis HTE831*	*Ureaplasma urealyticum serovar 12*
Clostridium sp. L2-50	*Oenococcus oeni ATCC BAA-1163*	*Ureaplasma urealyticum serovar 13*
Desulfotomaculum reducens MI-1	*Oenococcus oeni PSU-1*	*Ureaplasma urealyticum serovar 4*
Dorea longicatena DSM 13814	*Onion yellows phytoplasma OY-M*	*Ureaplasma urealyticum serovar 5*
Enterococcus faecalis V583	*Paenibacillus larvae subsp. larvae BRL-230010*	*Ureaplasma urealyticum serovar 7*
Epulopiscium sp. 'N.t. morphotype B'	*Pasteuria nishizawae str. North American*	*Ureaplasma urealyticum serovar 8*
	Pediococcus pentosaceus ATCC 25745	*Ureaplasma urealyticum serovar 9*

Archaea	**NAB-1**	**NAB-2**
Haloquadratum walsbyi DSM 16790	YP_659191	YP_659192
Methanobrevibacter smithii ATCC 35061	YP_001273517	YP_001274112
Methanocaldococcus jannaschii DSM 2661	NP_247475	NP_248059
Methanococcoides burtonii DSM 6242	YP_566239	YP_566240
Methanococcus aeolicus Nankai-3	YP_001324615	YP_001324618
Methanococcus maripaludis C5	YP_001097036	YP_001097035
Methanosarcina acetivorans C2A	NP_618639	NP_618640
Methanosarcina barkeri str. Fusaro	YP_306910	
Methanospirillum hungatei JF-1	YP_504500	YP_504504

Methanococcus vannielii SB	*Methanococcus maripaludis C7*	*Pyrobaculum calidifontis JCM 11548*
Methanoculleus marisnigri JR1	*Methanococcus maripaludis S2*	*Pyrobaculum islandicum DSM 4184*
Aeropyrum pernix K1	*Methanocorpusculum labreanum Z*	*Pyrococcus abyssi GE5*
Archaeoglobus fulgidus DSM 4304	*Methanopyrus kandleri AV19*	*Pyrococcus furiosus DSM 3638*
Caldivirga maquilingensis IC-167	*Methanosaeta thermophila PT*	*Pyrococcus horikoshii OT3*
Candidatus Methanoregula boonei 6A8	*Methanosarcina mazei Go1*	*Staphylothermus marinus F1*
Ferroplasma acidarmanus fer1	*Methanosphaera stadtmanae DSM 3091*	*Sulfolobus acidocaldarius DSM 639*
Haloarcula marismortui ATCC 43049	*Methanothermobacter thermautotrophicus str. Delta H*	*Sulfolobus solfataricus P2*
Halobacterium sp. NRC-1	*Nanoarchaeum equitans Kin4-M*	*Sulfolobus tokodaii str. 7*
Halorubrum lacusprofundi ATCC 49239	*Natronomonas pharaonis DSM 2160*	*Thermococcus kodakarensis KOD1*
Hyperthermus butylicus DSM 5456	*Picrophilus torridus DSM 9790*	*Thermofilum pendens Hrk 5*
Ignicoccus hospitalis KIN4/I	*Pyrobaculum aerophilum str. IM2*	*Thermoplasma acidophilum DSM 1728*
Metallosphaera sedula DSM 5348	*Pyrobaculum arsenaticum DSM 13514*	*Thermoplasma volcanium GSS1*

Actinobacteria	**NAB-1**	**NAB-2**
Brevibacterium linens BL2	ZP_00379860	ZP_00379861
Mycobacterium gilvum PYR-GCK		YP_001136025
Streptomyces avermitilis MA-4680	NP_824551	NP_824550
Streptomyces coelicolor A3(2)]	NP_629034	NP_629033
Thermobifida fusca YX	YP_288072	YP_288073

Acidothermus cellulolyticus 11B	Frankia sp. CcI3	Mycobacterium tuberculosis CDC1551
Actinomyces odontolyticus ATCC 17982	Frankia sp. EAN1pec	Mycobacterium tuberculosis F11
Arthrobacter aurescens TC1	Janibacter sp. HTCC2649	Mycobacterium tuberculosis H37Ra
Arthrobacter sp. FB24	Kineococcus radiotolerans SRS30216	Mycobacterium tuberculosis H37Rv
Bifidobacterium adolescentis ATCC 15703	Leifsonia xyli subsp. xyli str. CTCB07	Mycobacterium tuberculosis str. Haarlem
Bifidobacterium adolescentis L2-32	*marine actinobacterium PHSC20C1*	*Mycobacterium ulcerans Agy99*
Bifidobacterium longum DJO10A	*Mycobacterium avium 104*	*Mycobacterium vanbaalenii PYR-1*
Bifidobacterium longum NCC2705	*Mycobacterium avium subsp. paratuberculosis K-10*	*Nocardia farcinica IFM 10152*

Clavibacter michiganensis subsp. michiganensis NCPPB 382	*Mycobacterium bovis AF2122/97*	*Nocardioides sp. JS614*
Collinsella aerofaciens ATCC 25986	*Mycobacterium bovis BCG str. Pasteur 1173P2*	*Propionibacterium acnes KPA171202*
Corynebacterium diphtheriae NCTC 13129	*Mycobacterium leprae TN*	*Rhodococcus sp. RHA1*
Corynebacterium efficiens YS-314	*Mycobacterium smegmatis str. MC2 155*	*Rubrobacter xylanophilus DSM 9941*
Corynebacterium glutamicum ATCC 13032	*Mycobacterium sp. JLS*	*Saccharopolyspora erythraea NRRL 2338*
Corynebacterium glutamicum R	*Mycobacterium sp. KMS*	*Salinispora arenicola CNS-205*
Corynebacterium jeikeium K411	*Mycobacterium sp. MCS*	*Salinispora tropica CNB-440*
Frankia alni ACN14a	*Mycobacterium tuberculosis C*	*Tropheryma whipplei str. Twist*
		Tropheryma whipplei TW08/27

Spirochetales	**NAB-1**		**NAB-2**	
Leptospira borgpetersenii serovar Hardjo-bovis JB197	YP_800486		YP_800485	YP_800448
Leptospira borgpetersenii serovar Hardjo-bovis L550	YP_797605		YP_797604	YP_797566
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	YP_002112	YP_002102	YP_002108 no ptase	YP_002104
Leptospira interrogans serovar Lai str. 56601	NP_711786	NP_711796	NP_711790	NP_711794
Treponema denticola ATCC 35405			NP_971570	
Treponema pallidum subsp. pallidum str. Nichols	NP_218729		NP_219001	

Borrelia afzelii ACA-1	*Borrelia burgdorferi B31*	*Borrelia garinii PBi*	
Borrelia afzelii PKo	*Borrelia burgdorferi Bol26*	*Borrelia valaisiana VS116*	
Borrelia burgdorferi 156a	*Borrelia burgdorferi ZS7*		