WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Evolution, Ecology and Population Biology

Dissertation Examination Committee:
David Queller, Chair
Joan Strassmann, Chair
Justin Fay
Jonathan Myers
Kenneth Olsen

Genomic Signatures of Conflict and Cooperation in Plants and Social Amoebae

by

Katherine Sylvia Geist

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

August 2019
St. Louis, Missouri

TABLE OF CONTENTS

LIST OF PUBLICATIONS

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I. **Geist, K.S.**, Strassmann, J.E. and Queller*, D.C. 2019. Family quarrels in seeds and rapid adaptive evolution in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **116**: https://doi.org/10.1073/pnas.1817733116

II. Noh, S., **Geist, K.S.**, Tian, X., Strassmann, J.E. and Queller*, D.C. 2018. Genetic signatures of microbial altruism and cheating in social amoebas in the wild. *Proc. Natl. Acad. Sci.* **115:** 3096–3101. https://doi.org/10.1073/pnas.1720324115

Papers I-II are reproduced in the Appendix with the permission of the publishers under different permission as stated here:
https://www.Proc. Natl. Acad. Sci..org/page/about/rights-permissions

* Corresponding author.

The contribution of Katherine Sylvia Geist to the papers included in this thesis was as follows:

I. Designed, performed, analyzed the data, and wrote the paper.
II. Designed the molecular evolution analyses, analyzed those data, and wrote the paper.

LIST OF FIGURES

LIST OF TABLES

LIST OF BOXES

ABBREVIATIONS

CMS	cytoplasmic male sterility
r	relatedness between two individuals
DFE	Distribution of Fitness Effects
SFS	Site Frequency Spectrum
ω	ratio of the rate of nonsynonymous fixation between species to the rate of synonymous fixation, also denoted d_N / d_S
Dn	nonsynonymous fixed differences between species (counts)
Ds	synonymous fixed differences between species (counts)
Pn	nonsynonymous within-species differences (counts)
Ps	synonymous within-species differences (counts)
MK test	McDonald-Kreitman (1991) test
α	proportion of nonsynonymous substitutions that are adaptive
ω_a	rate of adaptive fixations between species
GBS	genotype-by-sequencing

ACKNOWLEDGMENTS

First, my deepest gratitude to my advisors, Joan Strassmann and David Queller. They have both encouraged me to continuously improve: as a scientist, as a writer, and even as a person. Despite the times I have challenged their faith in me, they continue to celebrate my every success. It continues to be my goal to make them proud. In their lab, I found an intellectual safe haven.

I would like to thank the members of my committee for their guidance and patience. I thank Justin Fay for his expertise and for providing me my first foray into some of these topics with my rotation in his lab. I thank Ken Olsen for keeping me 'rooted' in the botany as well as for his valuable input over the years. I thank Jonathan Myers for helping me think outside my system and helping me remember the underlying ecology of my study organisms.

I absolutely must thank my lab members through the years, starting with Debbie Brock, who not only trained me in the ways of *D. discoideum* but has continued to be my cheerleader. I thank Tracy Douglas for being a friend and sounding board, and for many fun hours playing with Dicty genomes. I would like to thank jeff smith for our many fruitful conversations, and for James Medina and Shreenidhi Perukkaranai Madabhushi for being such wonderful officemates these last years. Shreenidhi has become such a wonderful friend and colleague. Last, but certainly not least, I would like to thank my friend and collaborator, Suegene Noh. Suegene helped me grow in more ways than she can ever know both scientifically and personally.

Many thanks to the members of my cohort Arthur Porto, Claudia Henriquez, and Jennifer Gruhn for helping me survive the early years. I will always be grateful to Jennifer and Suegene for helping me survive a very difficult time mid-grad school.

I am grateful to DBBS for its support over the years, particularly the staff of student health and my physicians at Barnes. I would not be where I am without them. I would also like to thank members of the Biology department, in particular Carl Hennicke for his assistance with the biology cluster over the years. I would like to acknowledge my funding support for research and tuition through the university, National Science Foundation (DGE-1143954, IOS-1656756, IOS-1753743), and John Templeton Foundation (43667).

Prior to coming to Wash U, I had some amazing folks in my academic life without whom I never would have made it here. I would like to thank my Wyoming research mentors, Alex Buerkle and Tom Parchman, for showing me how exciting bioinformatics research could be. Yaffa Grossman, John Junck, and Ken Yasukawa at Beloit College have continued to be encouraging even from afar. Yaffa was the first botanist in my life and introduced me to biology through statistics. John introduced me to genetics and bioinformatics and encouraged me to become a programmer. Ken opened up a whole new way of thinking about the world through a social lens. I am eternally grateful for their mentorship.

I want to thank my dad for being, well, my dad. Through everything, he's been there for (mostly virtual) hugs and encouragement. If there is anyone on this planet with unfailing belief in me, it's him. He has never thought me capable of failure. My brother and sister have known just what to say when I needed it most. My aunt Lou Anne inspires me daily as she shows me what it means to 'beat the odds'. Finally, there is my best friend Jessica Dybsetter, who chased me down when I didn't return her calls, held my hand from 800 miles away, and above all else made sure I never felt alone.

To those yet unnamed who inspired me along my journey with their tenacity, endurance, or kindness, I admire you. To those unnamed who propelled me forward unintentionally – thank you.

Katherine Geist

Washington University in St. Louis

August 2019

Dedicated to Jessica for always being my kindred spirit.

ABSTRACT OF THE DISSERTATION

Genomic Signatures of Conflict and Cooperation in Plants and Social Amoebae

by
Katherine Sylvia Geist

Doctor of Philosophy in Evolution, Ecology and Population Biology

Washington University in St. Louis, 2019

Professor David Queller, Co-Chairperson

Professor Joan Strassmann, Co-Chairperson

Arms races involve bouts of reciprocal co-adaptation to a social environment. We have a strong sense for how arms races drive the evolution of genes in purely antagonistic contexts, such as host-pathogen or predator-prey. In these systems, conflict that produces arms races between two parties results in positive selection – the fixation of adaptive alleles between species – for both parties. However, we do not have an equal sense for how arms races during cooperative enterprises shape genic evolution. If we assume that arms races affect genic evolution similarly regardless of context – antagonistic or cooperative – then we would expect a signature of positive selection as a hallmark of arms races that have occurred between otherwise cooperating parties.

This dissertation attempted to test this prediction using two different systems, within-family conflicts in the plant genus *Arabidopsis* and between-clone conflicts in the social amoeba *Dictyostelium discoideum*. In Chapter 1, I introduce conflict and cooperation, how arms races drive positive selection, and my study systems in more detail. Because two of my chapters have already been published, they can be found under List of Publications. In Paper I, I used sets of genes predicted by theory to be involved in within-family conflicts over maternal resource

allotment to the developing seed. I tested whether these genes exhibited the telltale signature of an arms race as predicted. I found evidence that strongly supports a mother-offspring conflict scenario: genes enriched in the maternal seed coat and endosperm show elevated rates of adaptation relative to the embryo. This supports mother-offspring conflict because, as the intermediate provisioning tissue for the embryo, the endosperm is predicted to be the seed compartment in conflict with the mother plant, not the embryo. Further, I find that genes enriched in nutrient transfer tissues show elevated rates of adaptation relative to those enriched in non-transfer tissues. This further supports a mother-offspring conflict scenario over maternal resource allocation. I rule out other competing hypotheses including selection for smaller seed size in the *A. thaliana* lineage.

In Chapter 2, I continue to focus on within-family conflict over maternal resource allocation in seeds, this time using genes that have parent-of-origin biased expression (imprinting). The kinship theory of imprinting predicts that imprinted genes are in conflict with the mother plant over maternal resource allotment. Given the coincident mother-offspring conflict over maternal resource allocation I found in Paper I, I test whether imprinted genes experience a selection pressure distinct from that. I test the prediction that an arms race between mother plant and imprinted genes has driven positive selection of genes – here imprinted genes only. If test if the signatures I find are significantly greater than that of the background tissues. I find that imprinted genes show higher rates of adaptive evolution than their background tissues. This suggests that the selection pressure on imprinted genes is specific to their imprinting status. Further, my results are consistent with a conflict scenario over maternal resource allocation.

In Paper II, I switched systems to the social amoeba *D. discoideum* to test whether between-individual conflicts during asexual fruiting body development could lead to arms races.

Using genes identified by my collaborator, I tested whether genes differentially expressed during chimeric mixing showed evidence of an arms race. We would expect a possible arms race during chimeric mixing in order to suppress cheating, or the disproportionate contribution of one genotype to spore at the expense of the other genotype that goes into sterile stalk. Consistent with an arms race scenario, I found that genes differentially expressed – both up- and down-regulated – during chimeric mixing had higher rates of adaptive evolution when compared to the genomic background. This suggests that these genes may be important in the wild for facultative strategies to prevent exploitation by other genotypes.

Overall, these studies examined the effect of conflict in the context of cooperation on genic evolution: is it the same as we see with pure antagonism? This answer is that it appears to be. Not only can we use these kinds of methods to test theory about conflict genes in a robust way, but we can also use these methods to confirm the genes we identify are relevant to our organism in the wild. The latter is especially powerful for organisms like microbes or plants where observing social conflicts is not necessarily as straightforward as in animals. Further, these results suggest a strong role for kin conflict in seed development that has been largely understudied. It is the hope that this dissertation sparks a new set of kin conflict questions for researchers interested in both the proximate and ultimate factors affecting seed development.

INTRODUCTION

1.1 EVOLUTIONARY CONFLICT AND COOPERATION

Ever since Darwin, biologists have focused on conflict as different organisms struggle for the same resources. But it is also important not to neglect the power and importance of cooperation which allows highly successful alliances to form. Cooperation among genes, cells, and organisms was involved in nearly all major transitions in the hierarchy of life, including eukaryotic cells, multicellular organisms, and animal societies (Maynard Smith and Szathmáry 1995, Szathmáry and Maynard Smith 1995). Each of these transitions required that lower-level units cooperate and forego their independent existence to evolve into the higher-level unit. Thus, organisms evolved at each of these levels of biological organization via extensive cooperation – and minimal conflict – among their interacting parts (Queller and Strassmann 2009, Strassmann and Queller 2010).

It follows that we can then think of cooperation not as the absence of conflict but rather as the *mitigation* of conflict. Genetic conflicts can, and do, arise at all levels of biological organization both within and between individuals. A genetic conflict is one where the different parties involved in the conflict have antagonistic effects on one another's reproduction such that one party *benefits* by harming the other party (Werren 2011). They can be powerful selective forces with important evolutionary consequences (Burt and Trivers 2006, Werren 2011, Queller and Strassmann 2018). In fact, if conflict between and within organisms is too great, it will break down or prevent the emergence of higher levels of biological organization (Buss 1987, Maynard Smith and Szathmáry 1995, Queller 1997, Bourke 2011, Queller and Strassmann 2018).

It is important to distinguish actual and potential conflict, where the former is observed strife and the latter is a condition under which strife is predicted to evolve (Ratnieks and Reeve

1992, Ratnieks *et al.* 2006). We can think of potential conflict as when, given sufficient genetic variation, two parties should be selected to drive a shared trait – known as a joint phenotype – in two different directions according to their own interests (Queller 2014, Queller and Strassmann 2018). One example of a joint phenotype between a plant and its herbivore would be the use of plant resources, say stored sugars. Milkweeds (*Asclepias* spp.) benefit from using sugars for their own growth and have evolved toxic compounds like cardenolides, which can be neuro-toxic to some herbivores, as protection (Birnbaum and Abbott 2018). However, herbivores would benefit from hijacking the sugars for themselves. The obligate milkweed aphid, *Aphis nerii*, have evolved post-ingestive enzymes that break down these toxins, whereas other herbivores like the larvae of the monarch butterfly (*Danaus plexippus*) have evolved toxin resistance. It is worth noting that these are not the only examples of counter-defenses that milkweeds and their herbivores have evolved in response to each other.

Although biology is replete with examples of obviously antagonistic conflict, here I focus primarily on conflict that arises during cooperative enterprises. Conflict can evolve during cooperative endeavors because each party gains in inclusive fitness by moving a joint phenotype in a different direction, for example during sexual reproduction. The successful fusion of a zygote and provisioning of the developing embryo by the mother is essential for the fitness of both parties. However, evolutionary conflicts during sexual reproduction can occur at multiple levels: between males and males or between females and females over who mates with whom and when (Arnqvist and Rowe 2005), between mother and father over who cares for the offspring (Houston *et al.* 2005), or between siblings over parental resources (Mock and Parker 1997). These conflicts can also occur at a molecular level, such as between males and females over fertilization (Swanson and Vacquier 2002, Galindo *et al.* 2003), between cytoplasmic

elements and the nuclear genome over the production of male gametes (Burt and Trivers 2006), or between mother and father over the provisioning of the developing offspring (Trivers 1974).

For example, we generally think of the cooperation between eukaryotes and their mitochondria as extensive because mitochondria have largely forfeited their own reproductive rights. They are uniparentally-inherited and do not undergo meiosis. Thus, mitochondria will be selected to favor the offspring sex in which it replicates (Cosmides and Tooby 1981, Burt and Trivers 2006), which is usually the mother (for exceptions see Neale *et al.* 1989; Luo *et al.* 2018). In facultatively hermaphroditic plants like *Arabidopsis thaliana*, a conflict between the mitochondria and the nuclear genome can lead to cytoplasmic male sterility (CMS) (Törjèk *et al.* 2006, Durand *et al.* 2012, Simon *et al.* 2016). Even though most plants are hermaphrodites – producing both female (ovule) and male (pollen) gametes – their mitochondria are inherited only through ovules. Any mitochondrial mutant that prevents the production of pollen should be positively selected as long as it increases female fertility even slightly (Lewis 1941). On the other hand, if a male sterility gene was located in the nuclear genome, it would only spread if it more than doubles female fertility. Thus, there is a strong selection pressure for nuclear genes to counteract CMS, restore male fertility, and CMS is indeed often suppressed by nuclear genes (Simon *et al.* 2016).

One particular conflict that arises from sexual reproduction of particular interest here is conflict within families, starting with the embryo and its conflict with its mother. Conflict may evolve between a mother and her offspring over the joint phenotype of maternal resource allotment (Trivers 1974). A mother is equally related to all her offspring, and so should be selected to balance investment among her current and future offspring all else being equal. She would only be selected to provision one offspring more if the benefit exceeds the cost to her

other offspring. On the other hand, any given offspring shares only one-half to one-fourth of its genes with its full-siblings and half-siblings, respectively. Thus, an offspring will be selected to demand more resources from the mother for itself until the point at which it harms its inclusive fitness through harm on kin. Such a demand on maternal resources could then lead to other types of within-family conflict, such as sibling rivalry (competition among siblings over maternal resources). This could in turn lead to fratricide and reduced brood size. In plants, parent-offspring conflict theory posits increasing conflict with increasing genetic diversity due to outcrossing (Shaanker *et al.* 1988). One prediction born of this is that, as plant taxa shift mating system from outcrossing to self-pollinating, there should be reduced seed abortion. In fact, some of the best evidence of parent-offspring conflict in plants comes from the observation that seed abortion rates increase with increased outcrossing (Shaanker *et al.* 1988).

Conflicts are also very possible and present during asexual (clonal) reproduction, which would include the somatic reproduction of cells within a multicellular individual. Although conflicts within a multicellular organism seem to be better controlled because cellular reproduction is clonal and there is an early separation of the germline and soma (Buss 1983), there are notable exceptions. Genetic conflicts can arise from somatic mutation and subsequent selection on selfish cell lineages. Somatic cancers are a classic example of a selfish cell lineage that can potentially erode the organism through lethality (Aktipis and Neese 2013). The newly arisen cancer mutants are in conflict with the rest of the somatic cells over resource use. Failure to effectively suppress conflict within a multicellular organism, like cancer, can undermine the stability of an organism – here, death.

Conflicts also evolve more readily in organisms that are not strictly clonal, such as those with aggregative multicellularity like the social amoeba *Dictyostelium discoideum* (Strassmann

et al. 2000) or the bacterium *Myxococcus xanthus* (Velicer and Vos 2009). In these systems, the reproductive units (spores) are the result of an aggregation event, in which free-living cells come together to form a fruiting body. This fruiting body lifts the spores above the surface, presumably to aid in dispersion. However, in both systems cheating can arise – some genotypes preferentially create reproductive spores at the expense of the individual cells that create sterile stalk (Strassmann *et al.* 2000; Fiegna and Velicer 2003). High relatedness (*r*) among individual cells, which can be achieved passively through structured growth in the soil environment (Buttery *et al.* 2012, smith *et al.* 2016) or through active means like kin recognition (Strassmann 2016; Wielgloss *et al.* 2018), can keep cheating low in populations. In those cases where relatedness is purposefully kept low, cheating is driven high and destroys the stability of the multicellular fruiting body (Kuzdzal-Fick *et al.* 2011). This would cause failed dispersal and thus lowered fitness.

Queller and Strassmann (2018) recently argued that we seem to have largely solved the problem of how cooperation evolves, but we have fewer general principles about the evolution of conflict. Certainly, this is the case when it comes to our understanding of how conflict affects genic evolution. This gap is what I attempt to address with this dissertation. There seems to be one possible generalization about how antagonistic conflict affects genes: arms races drive positive selection in genes.

1. 2. Evolutionary Arms Races and Genic Evolution

Evolutionary conflict is likely to be an important driver of molecular variation within and between species (Van Dyken and Wade 2012). Some of our best evidence of how conflict affects genic evolution comes from host-pathogen systems, where molecular arms races between the

two parties have resulted in detectable signatures in the gene sequences involved (*e.g.*, see Daugherty and Malik 2012). In these systems, evolutionary arms race dynamics drive patterns of rapid evolution and positive selection through repeated, alternating bouts of adaptation between two parties. First, one party evolves an evasive adaptation. Then, the second party evolves a counteradaptation to restore the *status quo* or even gain its own advantage. Because the two parties continually adapt in response to the other, metaphors like "arms race" or "Red Queen" (Van Valen 1973; Dawkins and Krebs 1979) have been applied. In Lewis Carroll's (1871) *Through the Looking-Glass*, the Red Queen has Alice run faster and faster only to wind up where they began. She then tells Alice that it takes "all the running [Alice] can do to keep in the same place" (Carroll 1871). Initially, "arms race" was used to describe adaptations in one species leading to counteradaptations in another species, such as predator and prey or host and parasite (Van Valen 1973). This concept has since been extended to interacting members of the same species (Dawkins and Krebs 1979).

Brockhurst *et al.* (2014) further subdivided Red Queen dynamics based on their mode of selection and genetic architecture. The type of dynamic I focus on in this dissertation is predicted to be of the 'escalatory' type: positive selection drives escalation of traits that are expected to have a quantitative or polygenic genetic architecture. This is in contrast with Red Queen dynamics that either fluctuate or 'chase'. In the latter, high levels of genetic diversity are maintained rather than the arms race driving species to fix genetic differences. I present several examples of escalatory Red Queen dynamics in Table 1.1.

Studies of escalating arms race dynamics in various systems have elucidated some commonalities in the patterns of molecular variation arising from these interactions (Nielsen 2005, Daugherty and Malik 2012). These patterns suggest that arms races leave a signature of

positive selection on genes, where positive selection is typically identified as departure from neutral evolution. Positive selection can either be identified as a departure from neutral sequence evolution within a species or neutral sequence evolution between species. I will discuss the relevant tests for positive selection in more depth in Section 1.3. Table 1.1 presents key examples of how various arms race scenarios have influenced gene evolution, including host-pathogen, predator-prey, plant-herbivore, and male-female sexual conflict. Despite a variety of analysis types across a diverse array of taxa and systems, a common pattern emerges: *escalating arms races drive positive selection in genes*. This pattern suggests that we would expect to see a signature of positive selection in genes involved in an arms race (Daugherty and Malik 2012).

With the exception of male-female conflict over fertilization, the arms race examples from Table 1.1 are purely antagonistic in nature. Mating is an inherently cooperative enterprise because without that cooperation the successful production of progeny is unlikely. However, just as in any cooperative venture, the two parties can have different genetic interests and conflict between them can evolve. Examples from additional taxa support the idea that arms races between males and females over reproduction can drive positive selection in genes (for example, insects: Panhuis and Willi 2006, Pröschel *et al.* 2007; mice: Dean *et al.* 2009, Kousathanas *et al.* 2014; plants: Gossmann *et al.* 2013, Arunkumar *et al.* 2013). There is also evidence from mother-embryo conflict in mice (Chuong *et al.* 2010) that arms races may drive the positive selection of genes.

Beyond these examples, though, the evidence is sparse for scenarios where conflict has evolved between two cooperating parties. I seek to address this gap in our understanding with this dissertation. Once filled, this gap will help us address the general importance of evolutionary conflict for signatures of positive selection. Further, it will allow us to assess the predictive value

of looking for evidence of rapid, positive selection in the genome to identify putative social conflict genes. I propose that within-family conflicts during sexual reproduction and between-clone conflicts during asexual reproduction provide excellent opportunities to address these questions.

1.3 Detecting Positive Selection

Discussing all possible effects of positive selection in the genome is beyond the scope of this dissertation. However, it is worth mentioning that as we continue to look for adaptive loci there is no single effect of adaptation on genome architecture or evolution of gene sequences (Radwan and Babik 2012). There are a few very specific effects of adaptation that pertain to arms race-driven positive selection from Table 1.1 that I need to elucidate. In nearly all of those examples, the authors found fixed adaptive differences between species across genes and, in some cases, were able to identify single nucleotides under putative positive selection. Our models of adaptive evolution assume that adaptation arises from *de novo* mutations (Orr 2005). Further, because of degeneracy of the genetic code, these mutations are (1) most likely to be nonsynonymous and (2) probably deleterious. Thus, most of the mutations expected to 'survive' are ones that are neutral or nearly neutral (Kimura 1983, Ohta 2002, Nei *et al.* 2010). We thus generally assume that synonymous changes are neutral or nearly so, and that any difference between nonsynonymous and synonymous variation within or between species is a result of purifying or positive selection (Fay and Wu 2001, Nei *et al.* 2010). Adaptive mutations therefore should be driven to fixation between species quickly, something that we can detect with between-species comparisons. Empirical data on bacteria support this, where most fixed mutations are beneficial (Barrick *et al.* 2009, Tenaillon *et al.* 2012) and theory predicts that these fixations will occur quickly (Sniegowski *et al.* 2010). However, in all populations, the

effectiveness of selection is going to depend on both effective population size and recombination rate (Lynch 2007).

Most of the methods that are currently employed to detect positive selection are based either directly or indirectly on these premises (*e.g.,* Nielsen 2005). These methods differ in the type of data, time scale, numbers of genes, and overall power to detect selection. Tests differ fundamentally in whether they are based on the site-frequency spectrum (SFS) or compare interspecific and/or intraspecific variation among genes (Nielsen 2005). The SFS-based methods are considered indirect methods (Fay 2001) and include tests like Tajima's *D* (1989) or Fay and Wu's *H* (2000). These methods look for skews in the frequency distribution of neutral variation linked to a site under positive selection, *i.e.,* a selective sweep. They can identify departures from neutrality that indicate a sweep (*e.g.,* $D < 0$), but can be biased by population structure, demography, or recombination (Templeton 2006). One can distinguish positive selection from demography by doing a genome-wide comparison: whereas positive selection will create a local sweep, demographic effects will create a genome-wide sweep. Other indirect methods for detecting positive selection that compare intraspecific variation include using F_{ST}, which measures population differentiation (Weir and Cockerham 1984). One can look for evidence of population differentiation site-by-site or gene-by-gene using F_{ST}.

Such methods are often considered indirect tests for positive selection (Fay and Wu 2001, Nielsen 2005) because they look for departures from neutral evolution within a population. Direct methods include information about interspecific sequence variation. A method that compares only interspecific variation is dN\dS or ω, which is the rate of nonsynonymous substitution relative to the rate of synonymous substitution (Yang and Bielawski 2000). dS is assumed to give the rate of neutral evolution in protein coding genes, thus a significant $\omega > 1$

implies rapid, positive selection. However, this test is conservative because most genes, even those under positive selection, will also be partly under purifying selection. This can lower ω below one even if positive selection is acting on the gene. To increase the sensitivity to detect positive selection, methods also exist that compare interspecific to intraspecific variation, such as the McDonald-Kreitman (1991) test (MK test). This method compares nonsynonymous and synonymous fixed differences between species, Dn and Ds respectively, to nonsynonymous and synonymous polymorphism, Pn and Ps respectively. Unlike ω which estimates a rate (denoted with lower case *d* in dN/dS) of fixation between nonsynonymous and synonymous mutations, the MK test uses counts. It compares Dn/Ds relative to Pn/Ps for a gene with a 2x2 Fisher-Exact test. A significant Dn/Ds > Pn/Ps is indicative of positive selection. The latter methods, ω and the MK test, are often considered more immune to the effects of population history because they include interspecific data (although see Eyre-Walker 2002 or Messer 2013). Further, because the MK test adds Pn/Ps, it controls not only for neutral evolution but also for purifying selection. From Pn/Ps alone, we can obtain an expected Dn/Ds if the only evolutionary forces on a gene is neutral evolution and purifying selection. The tradeoff is that because ω and the MK test are often calculated on single genes, the power to detect selection can be quite low (Fay and Wu 2001, Eyre-Walker 2002).

All of these analyses have specific limitations and uses, but in the case of the MK test certain extensions have been created to attempt to overcome them. For example, pooling genes can increase detection power (Rand and Kann 1996, Fay and Wu 2001, Stoletzki and Eyre-Walker 2011). Of course, one must know what genes to include in the analysis *a priori*. If one finds a signature of adaptation in the pooled set of genes, it can be inferred that positive selection has operated these genes – or at least a sufficiently large subset of them – to generate this signal.

In this dissertation, I predominantly focus on one metric, α, which is an extension of the MK test and pools information from across a class of genes. The metric α estimates the proportion of adaptive substitutions as $\alpha = 1 - (\frac{Pn}{Ps} / \frac{Dn}{Ds})$, where *Pn* and *Ps* are counts of nonsynonymous and synonymous polymorphisms within species, respectively, and *Dn* and *Ds* are counts of nonsynonymous and synonymous substitutions between a species pair, respectively (Fay and Wu 2001). Thus, an $\alpha = 1$ suggests that 100% of all nonsynonymous substitutions are adaptive, whereas an $\alpha = 0$ suggests that none are. Values of α can range from $-\infty$ to 1, although values of $\alpha < 0$ are generally considered to be difficult to interpret (Eyre-Walker 2002). Negative α values are common in plants (Gossmann *et al*. 2010). A different parameterization of α, the rate of adaptive substitutions ω_a accounts for any effect of effective population size on adaptive allele fixation (Gossmann *et al.* 2012). Here, I focus on the method of Eyre-Walker and Keightley (2009) to estimate α and ω_a which uses the distribution of fitness effects (DFE) and the pairwise ratio of nonsynonymous to synonymous substitutions (ω, Yang 2007). Throughout my dissertation, I employ a custom program that I wrote that interfaces with the DFE-α program and performs all necessary resampling statistics.

1.4 Study Systems

Reproduction, whether sexual or asexual, provides unique opportunities for conflict during an otherwise harmonious endeavor (Queller and Strassmann 2009). If the goal is to ascertain how conflicts during cooperative endeavors affect genic evolution, focusing on conflicts during reproduction is a reasonable starting point. First, there is the growing body of evidence from diverse taxa that arms races from male-female conflicts drive rapid molecular evolution (*e.g.*, Wilburn and Swanson 2016). Second, there are many opportunities during sexual

Type of Arms Race	System	Result	Analyses Used	Citation
Host-Pathogen	(Host side only) *RNASEL* gene in humans confers viral resistance through its involvement in the antiviral and apoptotic actions of interferons. Authors look for evidence of positive selection across primates and within humans (*H. sapiens*). Authors also look for any population association with prostate cancer.	Across primates, evidence of site-specific positive selection in both the C-terminal and N-terminal ankyrin repeats of *RNASEL*. These protein domains interact directly with viruses. Within humans, evidence of positive selection, specifically a particular allelic variant that confers greater viral resistance. This same allele is also negative associated with prostate cancer incidence world-wide.	Interspecific Methods (dN/dS); SFS-based methods (π, Tajima's D); Haploltype Analysis	Jin *et al.* 2012
	(Host and Pathogen sides) Viral capsid protein of HIV and other immunodeficiency viruses in primates are required for entry to host cells and subsequent retroviral replication. TRIM5α is a protein expressed on host cells that recognizes capsid proteins upon viral entry to restrict virus replication within the cell.	Evidence of positive selection in both the viral capsid protein (pathogen side) and TRIM5α (host side) suggests an ongoing arms race between them. Site-specific positive selection has been shown in the capsid-binding domain of TRIM5α. Capsid proteins also show evidence of positive selection likley to evade recognition by TRIM5α or other host proteins.	Interspecific Methods (dN/dS)	Sawyer *et al.* 2005; Bozek and Lengauer 2010
	For a more comprehensive review, see Daugherty and Malik 2012			
Predator-Prey	(Predator side only) Members of the Didelphidae family of opossums are known predators of pit vipers, which produce a hemorrhagic venom. One target of the venom is a hemostatic blood protein, von Willebrand factor (vWF).	Evidence of positive selection at specific sites in the gene encoding vWF, but only in the Didelphid clade of opossums. Amino acid changes affect the net charge and hydrophobicity of vWF, which is hypothesized to affect venom binding and confer resistance to the pit viper toxin.	Interspecific Methods (dN/dS)	Jansa and Voss 2011
	For a more comprehensive review, see Arbuckle *et al.* 2017			

Plant-Herbivore	(Plant side only) *Populus tremula* wound-inducible protease inhibitor genes. Protease inihibitors inhibit herbivory from some predators.	Evidence for selective sweeps in 5 of 6 protein inhibitor genes examined. Evidence of balancing selection in one protease gene, *T11*.	SFS-based methods (π, Tajima's *D*); Interspecific Methods (dN/dS, McDonald-Kreitman Test); Haplotype Analysis	Ingvarsson 2005
	(Herbivore side only) Brazzein taste receptor *TAS1R3* in *primates*. Brazzein is thought to be a "mimic" that tastes sweet yet underlies little caloric value to aid in seed dispersal.	Evidence of positive selection in the receptor gene *TAS1R3* in all primate lineages that can taste brazzein. One lineage, *Gorilla* spp., did not show evidence of positive selection in *TAS1R3*. This lineage does not appear to taste brazzein nor have Gorillas been observed eating brazzein-producing plants in the wild.	Interspecific Methods (dN/dS)	Guevara *et al.* 2016
	For a more comprehensive review, see Jander 2018			
Sexual Conflict	(Both Sides) Sperm lysin from abalone species (*Haliotis* spp.) interact with a receptor on the egg envelope called vitelline envelope receptor for lysin (VERL) in a species-specific manner.	Early studies found evidence of positive selection in sperm lysin only, but closer analysis of the eveolution of two repeats in the VERL gene. So, whereas the majority of VERL evolves neutrally, these repeats evolve rapidly. Authors suggest this may be evidence of a chase dynamic where these repeats are highly evolving, and sperm lysin evolves to 'keep up'.	Interspecific Methods (dN/dS)	Swanson *et al.* 2001; Galindo *et al.* 2003
	(Female side only) In humans, egg envelope (also known as the zona pelludica) proteins bind to sperm in a species-specific pattern.	The two species-specific sperm-binding regions of these proteins show strong evidence of positive selection.	Interspecific Methods (dN/dS)	Swanson *et al.* 2001
	For a more comprehensive review, see Wilburn and Swanson 2016			

TABLE 1.1. BRIEF SURVEY OF ARMS-RACE DRIVEN MOLECULAR EVOLUTION OF GENES. The overwhelming pattern is that putative arms races drive the selection of genes as identified by either inter- or intraspecific methods or both. Nearly all systems are purely antagonistic with the exception of sexual conflict which involves a cooperative enterprise between males and females for the successful reproduction of both parties. Evidence come from various taxa, and in most cases evidence of an arms-race driven conflict is found for both sides, for example, both hosts and pathogens or both sperm and egg. This evidence is necessary to conclude that coadaptation is occurring in both parties. It also suggests that we would predict positive selection in the conflict genes of both parties.

and asexual reproduction to study the evolution of conflict. During sexual reproduction, genes are usually distributed to gametes via a "fair" meiosis, so any deviation from this Mendelian view of life could reflect conflict (Burt and Trivers 2006, Daugherty and Malik 2012). Male and female genomes form a union to make progeny, but they may have competing interests in particular after fertilization. In systems with parental resource provisioning, males and females – often through the progeny and mother, respectively – may disagree over the amount and timing of maternal resources allotted.

I first investigate opportunities for conflict during sexual reproduction in the angiosperm genus *Arabidopsis* to examine the effects on genic evolution of conflict-driven arms races during cooperation (Paper I and Chapter 2). Because the unit of selection is the gene, I first needed sets of genes predicted to be under conflict in these scenarios to test these ideas. From there, I used the proportion of adaptive substitutions, α, and other tests for selection to compare how they these conflict genes evolve relative to other genes from the genome or the tissue in which they are found. I continue my investigation on the effects of conflict during cooperative enterprises in the social amoeba, *D. discoideum* (Paper II). Though asexual propagation may appear even more congenial, organisms with facultative multicellularity may contend with 'cheaters' that try to overrepresent themselves in the next generation. I use this other system to examine the effect of conflict on genic evolution in a completely different type of cooperative system to gain a fuller picture of the phenomenon. In all three cases, I have chosen systems where identifiable groups of genes could be hypothesized to be under greater conflict, which provides me with clear, testable predictions.

Here, I hypothesize that ***conflict-driven arms races drive adaptive DNA divergence in genes*** because of reciprocal coevolution between the two parties involved. The two parties need

not be of different species, like host-pathogen or predator-prey arms races, but can be of the same species like in male-female arms races. Further, the conflict driving these arms races can be found in ***cooperative endeavors.*** I predict that these conflict-driven arms races ***affect genic evolution in cooperative contexts similarly to in antagonistic ones****,* although kinship could reduce the degree of conflict**.** I focus on conflicts during sexual seed development in the genus *Arabidopsis*, leveraging the extensive knowledge of developmental timing, trajectory, and genes for *A. thaliana* specifically. I also test this hypothesis in putative conflict genes differentially expressed in chimeras during multicellular fruiting body development of the social amoeba, *D. discoideum.*

1.4.1 Reproduction and Seed Development in *Arabidopsis*

Seeds allow plants to reproduce even in dry environments and broadly disperse their offspring to colonize new environments. Because of this, seed production was a major innovation in the evolution of vascular plants that has been tightly linked with the success of all seed plants, particularly angiosperms (reviewed in Linkies *et al.* 2010). Whether a seed makes it to the next generation can depend on numerous traits that affect interactions with their external environment. If a seed makes it to the soil, it must lie dormant until conditions are opportune for germination, then grow viable, competitive, and reproductively mature plants. Thus, seeds may have evolved ecological adaptations that enhance survival through dormancy, dispersal, herbivory and predatory defenses, germination, seed coat permeability, or light and water response (Baskin and Baskin 2014). Seed size is another trait that determines seedling survival and establishment (Thompson *et al.* 1993, Moles and Westoby, 2004; 2006).

While the seed is still developing in its seedpod, other opportunities for adaptation exist within the seed itself. Seeds may have evolved adaptations to their internal environment because of conflict generated by differences in relatedness among three seed compartments. The angiosperm seed is comprised of three usually genetically different tissues - embryo, endosperm, and maternal seed coat. In many angiosperm species, endosperm is the nutritive tissue of the embryo, providing nutrition throughout seed ontogeny. It can also provide long-term reserves for the embryo during seed dormancy and germination. The endosperm is essential for embryogenesis, even in species that reproduce via asexual seeds (Costa *et al.* 2004). The endosperm is functionally analogous to other tissues that provision embryos, like the fetal placenta in mammals. Unlike the offspring-derived placenta, the endosperm arises from the second of two genetically unequal fertilization events. During fertilization in diploid angiosperms, the pollen tube delivers two identical haploid male gametes to the ovule (Fig 1). The first fertilization occurs when one pollen nucleus fuses with the nucleus of the egg cell, a haploid female gamete, to become the zygote. This fusion results in an embryo that is made of equal maternal and paternal genomic contributions. During the second fertilization, another identical pollen nucleus usually fuses with two identical mother-derived nuclei in the central cell (Figure 1.1). In about 70% of angiosperm species, including *Arabidopsis,* the endosperm (Figure 1.2 B) is the triploid product of a second fertilization event (most other species also have double fertilization but show different endosperm ploidy levels) (Crepet and Niklas 2009). In *Arabidopsis,* this makes the endosperm genetically identical to its embryo but with a double dose of maternal genes. Unlike the embryo and endosperm, the seed coat is genetically identical to the mother plant (Figure 1.2). Thus, there are three genetic parties in the seed that could be in conflict. The embryo represents the next generation. The seed coat represents the mother plant

that protects and nourishes its developing offspring via the endosperm that surrounds the embryo. Lastly, the endosperm is an intermediate party that plays a critical role in nutrient transfer but, in *Arabidopsis*, is nearly absorbed by the embryo by seed maturation (Costa *et al.* 2004).

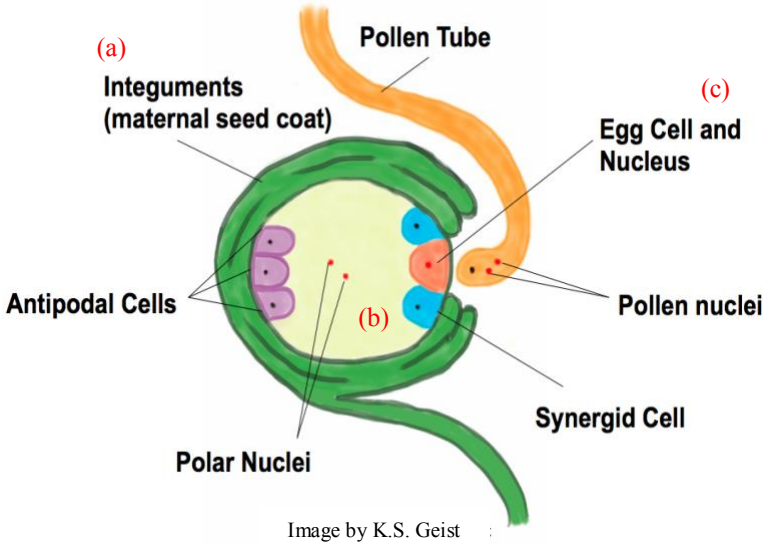


FIGURE 1.1. DOUBLE FERTILIZATION IN ANGIOSPERMS. In angiosperms, two types of haploid, multicellular gametophytes are produced by meiosis followed by subsequent rounds of mitosis: **pollen** (♂) and the **embryo sac** (♀). The angiosperm ovule is comprised of diploid maternal tissues called integuments (a) that later give rise to the maternal seed coat and surround the embryo sac. The embryo sac is typically seven-celled and eight-nucleate, with a large central cell (b) containing two polar nuclei, three antipodal cells, two synergid cells, and one egg cell (c). The opening in the integuments near the egg cell is the **micropyle**, through which pollen enters the ovule. A mature pollen grain contains two cells and three nuclei, a tube cell with a single nucleus and a generative cell with two nuclei. During fertilization, the pollen grain germinates, and the tube cell grows toward the micropyle. When it penetrates the ovule, one pollen nucleus fuses with the egg cell nucleus to form the diploid zygote and the

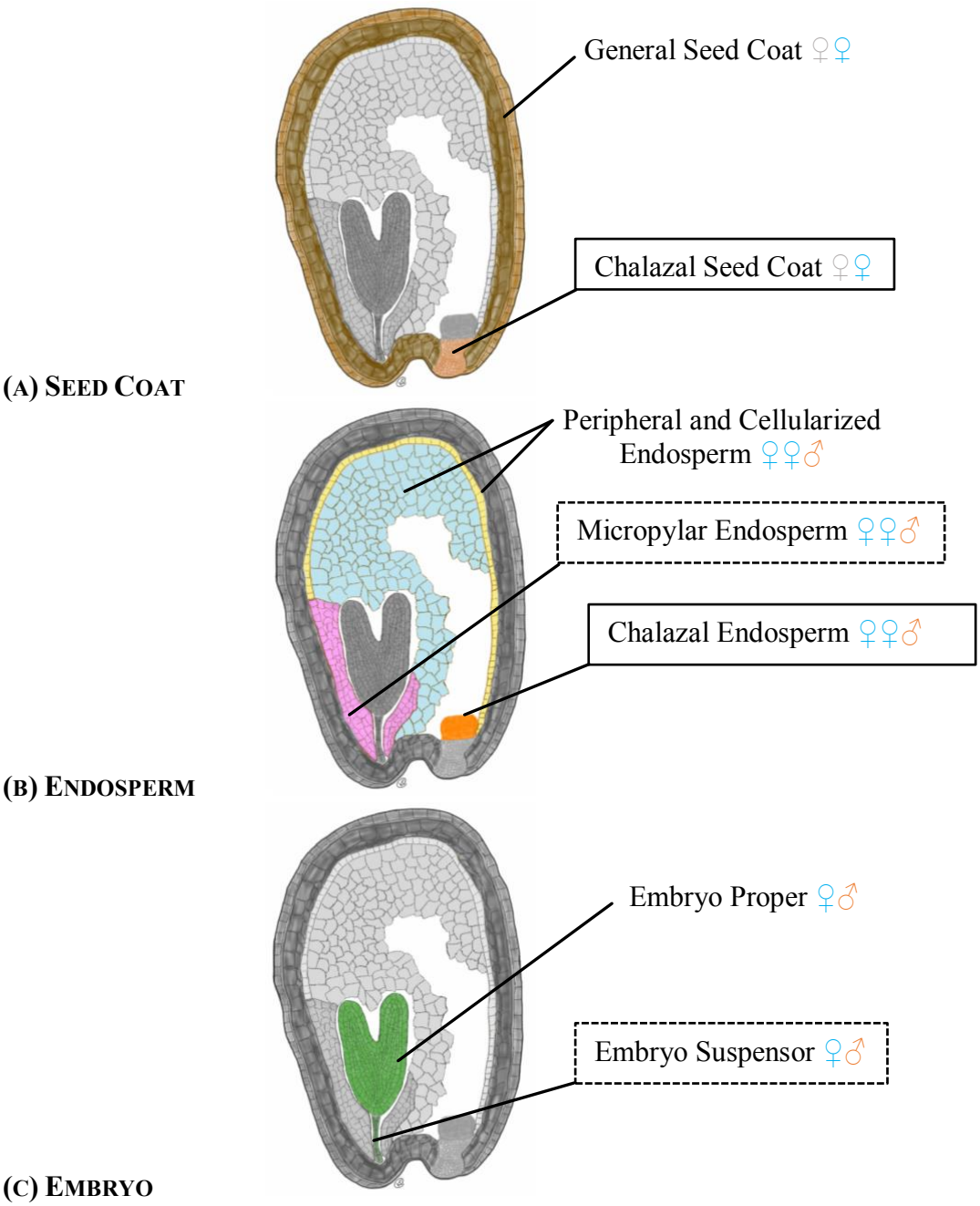


FIGURE 1.2. ANGIOSPERM SEED DEVELOPMENT OCCURS BETWEEN THREE GENETICALLY DISTINCT TISSUES. After fertilization, three genetic individuals in the seed are formed. The maternal seed coat is the diploid maternal tissue nourishes the developing seed via the triploid endosperm. The endosperm acts as a nurse tissue for the diploid embryo. Each compartment is comprised of subtissues, some of which function (solid black outline) or putatively function (dashed outline) in nutrient resource transfer from the mother plant; those subtissues without are not involved directly in nutrient acquisition from the mother. ♀ indicates one of two haploid genomes of the mother plant. ♀ indicates the other haploid genome of the mother plant that is inherited by her offspring. ♂ indicates the haploid genome from the father plant that the offspring inherits.

1.4.1. PARENT-OFFSPRING CONFLICT IN *ARABIDOPSIS* SEEDS

Although angiosperm development requires a high level of constraint among these three seed parties (reviewed in Ingram 2010), opportunities for conflict among the parties exist. Embryos or their nourishing tissue, the endosperm, could be selected to acquire more resources from the mother plant, which could in turn select for resistance from the mother (Westoby and Rice 1982, Queller 1983;1984;1989, Haig 1987). This could be considered a type of parent-offspring conflict with the addition of a third party, the endosperm (Trivers 1974; Queller 1983).

Under the original view of parent-offspring conflict as posited by Trivers (1974), conflict could arise because of different maternal resource investment optima for mothers versus their offspring. From a genetic perspective, mothers will be selected to balance investment among current and future offspring because they are equally related to offspring (relatedness $r = 0.5$). In contrast, an offspring will be selected to garner resources for itself at the expense of its siblings because it is more related to itself ($r = 1$) than to a sibling ($r = 0.5$). This creates a disparity where any given offspring will value the benefit of maternal investment twice as highly as their mothers, shifting its optimum level of investment greater than its mother's. Multiple paternity only increases this disparity because offspring are only related to half-siblings by $r = 0.25$, meaning that they will value the benefit of investment four times as highly as their mothers. Thus, the intensity of conflict is inversely proportional to relatedness between the parties.

Queller (1983, 1984) extended this theory to seed plants and the three parties: maternal seed coat, endosperm, and embryo. Using a kin selection (Hamilton 1964a;1964b) framework as Trivers (1974) did, Queller (1983;1984) argued that the different relatednesses among these three seed parties generate different maternal investment optima. In brief, we compare the benefit to cost ratio, *b/c*, from a generalized form of Hamilton's Rule (Box 1) to ratios of relatedness between the two individuals. The relatedness ratios for the three seed tissues are given in Table 1.2, which give us the relative genetic interests of the seed tissues (Queller 1983). The ratio of relatedness can be interpreted as the strength with which a tissue will favor its own embryo relative to other embryos (Westoby and Rice 1982). It can also be thought of as the point at which the tissue will no longer favor its own embryo at the expense of other embryos (Queller 1983). We can use these relatedness ratios to find the investment optima for each seed party, where each line represents a party's optimum (Figure 1.3). Any space between the lines represents potential conflict over the joint phenotype of maternal resource allocation - values of the benefit-cost ratio over which two parties should disagree. What becomes clear is that the greatest potential conflict exists between the embryo and maternal seed coat, with the

BOX 1.1. HAMILTON'S RULE

$$rb - c > 0$$

Hamilton's (1964a;1964b) original formulation has been rearranged in a number of different mathematically equivalent forms. One form breaks down relatedness for each of the two parties (West-Eberhard 1975). This form includes the genetic viewpoints of other individuals by explicitly stating the relatedness of donor and recipient:

$$br_b - cr_c > 0$$

As in Hamilton's original formulation, b is the benefit gained by the recipient and c is the cost borne by the actor. r_b is the relatedness of the individual to the recipient and r_c is the relatedness of the individual to the donor. This can then be arranged to a linear form:

$$b/c > r_c / r_b$$

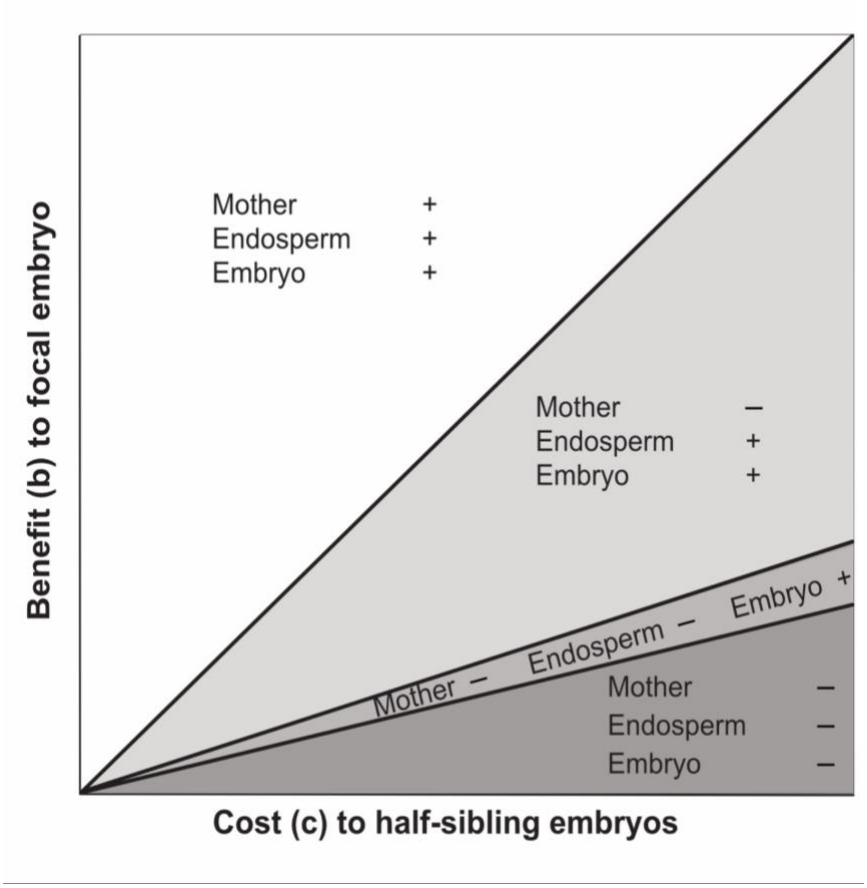
r_c / r_b is the ratio of relatedness calculated in Table 1.2. We can determine where there are differences in optima in resource provisioning between mothers and offspring if

endosperm falling intermediate (Figure 1.3), although the optimum for the endosperm falls closer to that of the embryo than mother. Thus, even though the endosperm is quantitatively weighted toward the mother in terms of maternal genomic contributions, it will still be selected to provision its embryo because it has the same genes as its embryo. It is important to note that even though we present the case of complete outcrossing with many pollen parents (Figure 1.3), the differences are qualitatively the same for a mixture of half- and full-siblings as well (Table 1.2). Full-siblings would align endosperm and embryo interests.

Seed Tissue	Embryo θ	Full-Sibling Embryo θ	Half-Sibling Embryo θ	Ratio of Relatedness $\theta_{Embryo} / \theta_{Full\text{-}Sib}$	Ratio of Relatedness $\theta_{Embryo} / \theta_{Half\text{-}Sib}$
Maternal Seed Coat	1/4	1/4	1/4	1	1
Endosperm	1/2	1/4	1/6	2	3
Embryo	1/2	1/4	1/8	2	4

TABLE 1.2. COEFFICIENTS OF KINSHIP AND RELATEDNESS RATIOS BETWEEN SEED TISSUES AND EMBRYOS. The coefficient of relatedness, *r*, is the proportion of alleles expected to be shared between two relatives due to identity by descent. Relatedness can also be expressed as the coefficient of kinship (θ), the probability that two alleles chosen randomly from each of the relatives are identical by descent (Falconer and Mackay 1996). The relationship between the two measures is $r = 2\theta$. Queller (1983) calculated the coefficient of kinship (θ) for each seed to three types of embryos: (a) its own embryo in the seed, (b) a full-sibling embryo in a different ovule but an identical pollen father, and (c) a half-sibling embryo in a different ovule with a different pollen father. The ratio of relatedness was calculated between a tissue's own embryo and either a full-sibling or half-sibling embryo from the kinship coefficients.

It is possible that because the endosperm's function is to nourish the embryo and their interests are quite similar, most if not all of the potential conflict between embryo and mother is handled by the endosperm instead. Further, the embryo must develop precisely whereas the endosperm's only function is provisioning. This would effectively align the interests of mother



Party *x*	$r_{embryo\text{-}x}$	$r_{halfsib\text{-}x}$	ratio = slope
Mother	**1/2**	**1/2**	**1**
Endosperm	**1**	**1/3**	**1/3**
Embryo	**1**	**1/4**	**1/4**

FIGURE 1.3. ZONES OF CONFLICT BETWEEN GENES OF DIFFERENT SEED PARTIES. From Box 1.1 and Table 1.2, we can derive the optima for each seed tissue (maternal seed coat, endosperm, and embryo) given relatednesses (*r*) calculated between the embryo and itself or the embryo and its half-sibling. If an allele expressed by one of these tissues (*x* = seed coat, endosperm, embryo) increases nutrient flow to the seed, it increases the focal embryo's fitness by *b* (y-axis) at the expense of current or future maternal half-siblings, *c* (x-axis). From Hamilton's Rule (1964a;1964b), this allele will be favored when $r_{embryo\text{-}x} \cdot b - r_{halfsib\text{-}x} \cdot c > 0$, such that *r* is the relatedness of the embryo or the half-sibling embryo to the tissue *x* (Table 1.2, Queller 1983). This suggests that each tissue will favor increased nutrient flow (+) when $b > (r_{halfsib\text{-}x} / r_{embryo\text{-}x}) \cdot c$, and disfavor it (−) when the inequality is reversed. If *b/c* is high enough (white zone), all seed parties favor nutrient transfer to the focal embryo and we would expect no conflict to evolve. In *b/c* is low enough (dark gray), none of seed parties favor nutrient transfer to the focal embryo, and conflict would again not be expected to evolve. Our zone of conflict lies in the two intermediate regions between the optimum for mother and endosperm (lighter gray) and endosperm and embryo (middle gray). Note that the zone for conflict is much larger between mother plant and endosperm than it is between an endosperm and embryo. Reproduced from Paper I.

and embryo more closely than if there were no endosperm at all. Thus, if there were a potential conflict-driven arms race between the mother plant and her offspring over the joint phenotype of maternal resource investment, we might only see it with the endosperm rather than the embryo.

The aim of Paper I was to use the patterns of molecular evolution, specifically α, to test if the conflicts predicted by Queller (1983, 1984; Figure 3) have created an arms race between the mother plant and endosperm. Potential conflict between maternal seed coat, endosperm, and embryo is an ideal study system because theory predicts differential investment optima (Queller 1983;1984, Westoby and Rice 1982). Recently generated, publicly available data from *A. thaliana* gives us the opportunity to identify the putative genes involved in this conflict (Harada *et al.* 2012). I can thus test the prediction that conflict-driven arms races over maternal resource allocation has created signatures of positive selection in only those genes predicted to be involved in the conflict. This is one of the only tests of positive selection on parent-offspring conflict genes (see also Chuong *et al.* 2010). It is also one of a limited number of tests of kin selection in plants (see also Shaanker *et al.* 1988 and Dudley 2015).

Last, this study system is important given a lack of direct evidence of parent-offspring conflict in plants. Empirical evidence of parent-offspring conflict in animal systems is now quite abundant, including: Soay sheep *(Ovis aries,* Wilson *et al.* 2005), turtles (*Apalone mutica*, *Chelydra serpentina*, and *Chrysemys picta;* Janzen and Warner 2009), domestic pigs (*Sus scrofa*, Drake *et al.* 2008), placental fish (*Heterandria formosa*, Schrader and Travis 2009), social insects (*Bombus terrestris* and *Formica truncorum*, Ratnieks *et al.* 2006), and numerous bird species (*e.g.*, reviewed in Kilner and Hinde 2012). In contrast, the evidence of parent-offspring conflict in plants has been quite indirect because it is challenging to observe directly (e.g., Queller 1983, Westoby and Rice 1982, Shaanker *et al.* 1988). Given this limited empirical

evidence, any molecular evolution support of these predictions would provide critical evidence of the phenomenon in *Arabidopsis*.

What I find is several lines of evidence that support possible mother-offspring conflict in seeds. I predicted that if the conflict between mother plant and endosperm was great enough it could generate arms-race driven positive selection in seeds. I find evidence that supports this prediction at the exclusion of other alternative explanations. First, despite the developmental constraint we would expect on the seed, I find that genes of the seed are generally evolving with more adaptive substitutions than genes from other plant parts like the floral bud, leaf, stem, and root.

Among the three seed parties – maternal seed coat, endosperm, and embryo – I find significantly more adaptive substitutions in the maternal and endosperm tissues than the embryo. If this had been the result of abiotic or germination-driven selection on the seed, we would expect to see the signature of positive selection in the seed coat only. Given the narrow zone of conflict in Figure 1.3 between endosperm and embryo (medium gray region), I suggest that the embryo is not as likely to evolve conflict with the mother plant. Instead, the endosperm has perhaps taken over the role of conflict on the embryo's behalf, which has been proposed as an explanation for the evolution of the endosperm in angiosperms (Friedman 1995).

Most importantly, for all of the known or putative nutrient transfer tissues of the maternal seed coat (chalazal), endosperm (chalazal and micropylar), and embryo (suspensor) (Figure 1.2), *all* show an elevated α relative to their non-transfer counterparts. For example, in the endosperm, the chalazal and micropylar genes show a signature of positive selection that the peripheral/cellularized endosperm does not. The cellularized/peripheral endosperm is thought to function solely as a storage subtissue (Costa *et al.* 2004).

One alternative explanation for the elevated α in nutrient-transfer genes that I excluded with my analysis was that of increased selection in the *A. thaliana* lineage due to selection for smaller seeds. Although we generally think 'bigger is better' when it comes to seeds, there are some ecological conditions which favor smaller seed size (Moles and Westoby 2004; 2006). The seeds of *A. thaliana* have evolved to be smaller than their congeners (Al-Shehbaz and O'Kane 2002). Although one hypothesis to explain this is that seeds evolved to be smaller in *A. thaliana* as a result of mating system shift from obligate to facultative outbreeding (deJong *et al.* 2005), that explanation would not predict the pattern of positive selection I see in seed tissues. However, if seeds evolved to be smaller in the *A. thaliana* lineage for an ecological reason that confers a selective advantage, this could cause the patterns we see. For example, in some locally adapted populations of *A. thaliana,* smaller seeds are better able to float, which is thought to improve their dispersal in flood-prone environments (Saez-Aguayo *et al.* 2014). Selection for smaller seeds would likely have targeted nutrient transfer genes, as those genes affect seed filling. Thus, it could have generated a similar pattern of positive selection. To test this, I included not only comparisons from one species pair, *A. thaliana* and its sister species *A. lyrata*, which have differently sized seeds, also between *A. lyrata* and *A. halleri*, which have similarly sized seeds. However, I find no evidence to support selection for smaller seed size. I see a similar pattern of α for the *thaliana-lyrata* and *lyrata-halleri* comparisons. Controlling for selection for smaller seed size in *A. thaliana* is equally important for Chapter 2.

Thus, the explanation that best fits the pattern of α I observed is conflict over resource transfer from the mother plant to the endosperm/embryo. On the side of the mother (chalazal seed coat), selection could serve to inhibit transfer whereas on the side of endosperm (chalazal and micropylar) and embryo suspensor, selection could serve to elicit transfer. Although it was

not within the scope of this dissertation to test these functional predictions outright, we do have a hint from the gene ontology analysis that is consistent with this prediction. Some of the most enriched gene ontology terms in the analysis I performed are for intra- and extracellular communication. Thus, while this does not provide specific gene functions or phenotypic effects as yet, this provides researchers with a place to start. These candidate genes could be of particular interest to researchers studying the control of seed size or intra-seed signaling.

For example, one immediate avenue of inquiry could be to harness the large number of re-sequenced genomes of *A. thaliana* (some of which I used here) to identify candidate loci underlying natural variation in seed size. This type of genome-wide association could be readily undertaken using the natural variation in seed sizes of *A. thaliana* ecotypes and associating that variation with genomic variation. These candidates could be cross-referenced with the putative conflict genes from this study, which would generate good candidates for phenotyping. It is also highly likely that some of these conflict genes are linked with variation in other life history traits, such as flowering time, as seed size genes have previously been found to be (Alonso-Blanco *et al.* 1999). As much interest as there is in seed size and the genes underlying seed size, this is a yet underexplored area (except see Alonso-Blanco *et al.* 1999 and Moore *et al.* 2013).

A second particularly interesting follow-up study could also look at potential tradeoffs in seed size and brood yield in *A. thaliana* with these conflict genes in mind. Compared to congeners *A. lyrata* and *A. lyrata, A. thaliana* forms more seed pods per plant with a higher number of its smaller-sized seeds (Krämer 2015). This could provide *A. thaliana* an advantage for its smaller seed size even if it that reduced seed size initially resulted from a transition to selfing (deJong *et al.* 2005). This opens up yet another opportunity for conflict with the mother plant that may or may not be underpinned by the conflict loci identified in this study.

1.4.2. IMPRINTING CONFLICT IN *ARABIDOPSIS* SEEDS

The conflict among the three seed members, seed coat, endosperm, and embryo, is not the only potential conflict over the joint phenotype of resource allocation that is predicted in the seed (Haig 2000). There is another potential conflict between males and females over resource allocation that may be embedded in, but distinct from, this mother-offspring conflict. Genomic imprinting, or parent-of-origin specific gene expression, results from a chemical memory that identifies an allele's parental origin (Gehring 2013). This occurs via epigenetic modification of the chromatin or DNA at the locus or modifier loci. Importantly, maternally-expressed (matrigene) and paternally-expressed (patrigene) imprinted genes can be identical in sequence and differ only in expression levels. Often, an allele from only one parent is expressed. Many imprinted loci are known to play integral roles in offspring development and mother-offspring interactions, including maternal resource provisioning (Wilkins and Haig 2003). In plants, imprinting can have phenotypic effects on seed size, seed abortion, embryo development, and endosperm cellularization (Gehring 2013).

The proximate causes of imprinting have been well-elucidated (Gehring 2013), and there has been mostly theoretical (*e.g.,* see Haig 2000 or Spencer and Clark 2014) with some empirical (*e.g.,* see Tuteja *et al.* 2019) work on the ultimate causes as well. Among all the ultimate theories for imprinting, only one is conflict-based: the kinship or parental-conflict theory (Haig 1997;2000, Spencer and Clark 2014). It argues that genomic imprinting is an evolutionary consequence of conflict between maternal and paternal genomes over maternal investment because of relatedness asymmetries (Patten *et al.* 2014). Just as with parent-offspring conflict in plants, these relatedness asymmetries can then lead to different investment optima (Figure 2.1). From a genetic perspective, if offspring are half-siblings, matrigenic interests are more closely

aligned with the mother plant although this is not an exact alignment because each offspring matrigene is only one of the mother's two gene copies. On the other hand, patrigenes have no alignment whatsoever (compare the matrigene and patrigene lines in Figure 2.1). For cases where siblings are full or a mix of full- and half-siblings, patrigene interests will begin to align with the embryo.

The kinship theory is arguably one of the most recognized imprinting theories to the point that some authors have complained of a lack of consideration of other theories (Spencer and Clark 2014, but see Moore and Mills 2008 for those with an opposite view). The theory also successfully explains the observed distribution and phenotypic effects of imprinted loci (Haig 2000, Moore and Mills 2008). For example, imprinting has been found in nearly all of the taxa in which it was predicted on the basis of extended maternal care (placental mammals: Monk 2015, marsupials: Renfree *et al.* 2008, plants: Köhler and Weinhofer-Molisch 2010, social insects: Kocher *et al.* 2015) though not in placental fishes (Lawton *et al.* 2005). Further, patrigenes in mammals favor prenatal growth, as predicted, while matrigenes inhibit growth (Haig 2004). Similarly, *Arabidopsis* patrigenes promote endosperm growth while matrigenes inhibit (Köhler and Weinhofer-Molisch 2010).

Here, I test whether the conflict experienced by imprinted genes increase the rate of adaptive evolution above any generated by mother-offspring conflict in the seed. The advantage of the study system here is the decades of research on imprinted genes in *Arabidopsis* in addition to the putative mother-offspring conflict genes used in Paper I. This positions us to test just how much conflict may drive the evolution of imprinted genes in *A. thaliana* and whether it is consistent with the arms race prediction of the kinship theory.

Compared with previous studies of genic evolution in imprinted genes of *Arabidopsis,* I find that imprinted genes are rapidly evolving compared to the rest of the genome (see Wolff *et al.* 2011, Tuteja *et al.* 2019). I use a more complete set of both 'confirmed' (known phenotypes) and 'candidate' (identified through high-throughput screens) than either of those previous studies. Further, whereas those studies concentrated on ω, which uses interspecific information only to infer selection, I used α to have more power to pool groups of genes. I find that both the confirmed and candidate imprinted genes have similarly high proportions and rates of adaptive substitutions.

More importantly, despite the mother-offspring conflict of the seed, I find imprinted genes are evolving with a higher α and ω_a than their respective background tissues. This is important because most imprinted genes have been identified in the endosperm, which I showed in Paper I to have a high rate of adaptive evolution. Thus, any signature of positive selection in imprinted genes could have simply reflected positive selection on the seed over mother-offspring conflict and not the separate source of selection predicted by the kinship theory.

I further find that patrigenes have a higher α and ω_a than matrigenes, which is consistent with previous findings in both *A. thaliana* (Tuteja *et al.* 2019) and less conclusively *Capsella rubella* (Hatorangan *et al.* 2016). What is different here is that I test patrigenes and matrigenes of the endosperm and embryo, respectively, to test for any evidence of haploid selection on imprinted loci. In brief, imprinted genes are effectively haploid, which would expose matrigenes to greater selection (both purifying and positive) in their heterozygous state in the endosperm only. This is because, unlike the diploid embryo, the triploid endosperm has a double dose of imprinted matrigene. Thus, we would predict a higher α in matrigenes of the endosperm (but not embryo) if haploid selection were the proximate cause of elevated α of imprinted loci. Instead, I

find the opposite pattern. I find, again, that patrigenes have experienced greater adaptive evolution than matrigenes, regardless of whether they are expressed in the endosperm or embryo. So, while this does not rule out that haploid selection could function on imprinted genes, it is not a sufficient explanation for the patterns we see.

Overall, the kinship theory is the only theory to predict conflict over maternal resource allocation and an arms race between imprinted genes and the maternal plant. My results strongly support kin conflict as a primary driver of evolution of imprinted gene sequences while ruling out other hypotheses never before tested. Further, we show that because α and ω_a are so high, imprinting conflict appears to be a very strong selective force. Such a signature of selection could possibly be used to confirm the imprinting status of future candidate genes.

In Paper I, I show that there is very minimal overlap between the mother-offspring conflict genes and imprinted genes. In the endosperm specifically – again, where most imprinting takes place in *A. thaliana* – the largest portion of imprinted genes were in the cellularized/peripheral endosperm – the tissue evolving with lower α. I found 8.8% of cellularized/peripheral endosperm were known imprinted genes, compared to 0.5% and 2.2% imprinted in the micropylar or chalazal regions, respectively. Future directions might look to see if the mother-offspring conflict genes are not interacting with imprinted genes even if they are not themselves imprinted. Elucidating these networks, as well as how these gene networks are affected by environment, could also shed light on both the proximate and ultimate controls of seed size in plants (Gutierrez-Marcos *et al.* 2012, Costa *et al.* 2012).

A further area of follow-up could be to compare the evolution of imprinted genes from partial vs. fully exalbuminous species (angiosperms whose embryos have absorbed all or nearly all endosperm at maturation). Whereas *Arabidopsis* is only partly exalbuminous, legumes are

fully exalbuminous. In both cases, the result is an embryo that is relatively large at maturation (as compared to embryos from albuminous species). However, in the Fabaceae, at the end of maturation the embryo is acquiring nutrients directly from the mother plant because the endosperm is gone (Zhang *et al.* 2007). Because imprinting is predominantly found in the endosperm, it would be interesting to know whether imprinting exists in the Fabaceae, although the methylome of soybean, *Glycine max*, is similar to that of *A. thaliana* (Lin *et al.* 2017). The fully exalbuminous nature of legumes like soybean is not likely to remove imprints, if they already exist, but it could change where imprinting is found (endosperm vs. embryo) and it could also create stronger selection on imprinted genes of the embryo. In *Arabidopsis,* I found a possible trend of higher α in imprinted genes of the embryo (Figure 2.5). Could this possibly reflect a shift at maturation of the embryo 'dealing with' the mother plant more directly via its imprinted genes? Our finding of a higher embryo α was both surprising and puzzling, but perhaps it reflects a maturation strategy that is the result of competition between matrigenes and patrigenes (Sakai 2010).

Lastly, more work is needed to test Haig's (2000; 2013) prediction that, if the kinship theory is correct, imprinting should be under stronger selection in outbreeding plants than inbreeding. Although we have a small hint with this work that this could be the case, it is far from conclusive. We do find that in the two populations of the obligately outcrossing *A. lyrata* (with the outcrossing *A. halleri* providing divergence data), not only do all imprinted genes have significantly high α in both populations but patrigenes do too (Figure 2.6). However, some populations of *A. thaliana* populations (*A. lyrata* outgroup) show this pattern of higher patrigenic α as well. Thus, our results only give mild support to kinship theory's prediction about higher imprinting evolution in outcrossers. A more robust test is needed.

1.4.3. ASEXUAL SPORE DEVELOPMENT IN *DICTYOSTELIUM*

The social amoeba, *Dictyostelium discoideum*, has both unicellular and multicellular life stages (Kessin 2010). It lives a predominantly solitary existence as single cells in the leaf litter of forest floors where, upon starvation, it becomes transiently multicellular. As a predator of bacteria, depletion of local food supply triggers a cascade of extracellular signaling to recruit neighboring cells. These cells then aggregate to form a multicellular, motile slug that migrates to a suitable location thought to be above the soil line for dispersal by arthropods (Bonner 2009, smith *et al.* 2014). Upon reaching its destination, the slug forms a fruiting body: spores are housed inside a sorus and held aloft with a rigid stalk composed of cells that have foregone reproduction. The composition of a mature fruiting body is roughly 80% reproductive spores and 20% somatic stalk cells (Bonner 1967, Raper 1984, Kessin 2010).

Cellular conflicts in *D. discoideum* are well-known and measurable in the lab (Gilbert *et al.* 2007, Santorelli *et al.* 2008, Strassmann and Queller 2011, Kuzdzal-Fick *et al.* 2011). Relatedness within fruiting bodies is high in a natural populations of *D. discoideum*, which would protect against harmful cheaters (Gilbert *et al.* 2007). Yet, some wild clones cheat each other by overrepresenting themselves in spore versus stalk (Strassmann *et al.* 2000). Many mutations appear to lead to cheating in *D. discoideum* (Santorelli *at el.* 2008), and these genes show a signature of balancing selection compared to other genes in the genome (Ostrowski *et al.* 2015). Cheating can be very harmful to cooperation, with some cheaters increasing rapidly in low relatedness conditions (Kuzdzal-Fick *et al*. 2011). Even if cheating is not apparent, there appear to be costs of chimerism, including reduced slug migration (Foster *et al.* 2002). Numerous mechanisms appear to have evolved to avoid cheating, including kin recognition where clones sort by adhesion loci *tgrB* and *tgrC* (Mehdiabadi *et al.* 2006, Benabentos *et al.*

2009, Hirose *et al.* 2011, Ho *et al.* 2013). Some populations can co-evolve resistance in response to cheating (Khare *et al.* 2009, Hollis 2012, Levin *et al.* 2015). There are also indirect ways cheating can be avoided, such as by maintaining high relatedness (Kuzdzal-Fick *et al*. 2011), which can be produced by local colonization (Buttery *et al.* 2012, smith *et al.* 2016). Pleiotropy is another indirect method that can control cheating, for example at the *dimA* locus (Foster et al. 2004). The *dimA* locus is required to differentiate into prestalk cells, which could make it a good target for cheating. However, lack if *dimA* by amoeba cells results in exclusion from spores, thereby mitigating any potential for cheating at that particular locus.

The relevance of cheating is in the wild has been questioned (Tarnita 2017), but the molecular history that can be gleaned from coding sequences may disagree. If cheating is common in nature and can result in the co-evolution of cheating resistance (Khare *et al.* 2009, Hollis 2012, Levin *et al*. 2015), this conflict could in turn lead to conflict and increased selection. Ostrowski *et al*. (2015) found evidence of balancing selection at characterized cheater loci, which they concluded to result from stalemate conflict.

One potential drawback of analysis on these genes is that they were isolated from cheaters that had evolved from a single genetic background (Santorelli *et al.* 2008; Ostrowski *et al.* 2015). In fact, most cheating assays of *D. discoideum* have been conducted in a uniclonal social context. Nearly all of the studies done to date have been performed on ancestors of a single natural clone, NC4 (for example, Santorelli *et al.* 2008, Kuzdzal-Fick *et al.* 2011). Thus, if the goal is to identify whether cheater genes may be subject to conflict-driven arms races in *D. discoideum,* genes isolated during chimeric mixing would be a more appropriate. Although *D. discoideum* readily forms chimeras in the lab (Strassmann *et al.* 2000) and chimeras can be found in the wild (Gilbert *et al.* 2007), the developmental trajectory is not fixed. Throughout

recruitment, aggregation, and even slug migration, individual cells can potentially 'opt' to abandon the multicellular unit (Kessin 2010). Perhaps more importantly, cells can influence other cells as to whether they go to the sterile stalk or the reproductive spore. Chimeric mixing is the context at which cheating would be adaptive, because cheaters could join the aggregate and thereby ultimate overrepresent themselves in the next generation via overallocation to spore. Thus, genes differentially expressed during chimeric mixing would make good candidates for facultative cheating or cheating resistance (Paper II). Under high relatedness, these genes would not be expressed or as expressed; but under lower relatedness (chimeric) conditions, they could be precisely those genes to confer cheating ability or resistance.

Thus, in Paper II, I test genes identified in chimeric mixing between wild clones of *D. discoideum* for signatures of conflict-driven arms races, again using the statistic α. These genes were identified at the tight-aggregate stage of social development in *D. discoideum*, which is a critical point for spore-stalk differentiation (Parikh *et al.* 2010). The advantage of this study system is that it allows us to use natural gene histories to ascertain the importance of cheating in the wild, something that is nearly if not impossible to study directly. Further, this project sets precedent for how one might study other conflict-driven arms races in otherwise intractable systems like microbes.

In Paper II, I performed all molecular evolution analyses after the differentially expressed genes were identified by RNA-seq. What I found was that both up- and down-regulated genes differentially expressed during chimeric mixing had a significantly high α and ω_a when compared to other genes in the genome. This was in contrast to the evidence of stalemate conflict (balancing selection) Ostrowski *et al.* (2015) found, which I did not. One possible reason for this is that these are genes critical for spore-stalk differentiation. The cheater genes used by

Ostrowski *et al.* (2015) – those originally identified by Santorelli *et al.* 2008 – are facultative cheater genes, as would be the genes identified here. The key difference is likely that of genomic background: the genes from the previous study were all identified in a single background. On the other hand, these were identified when mixing wild-collected clones, and are thus more likely to represent possible cheating or cheating avoidance strategies in nature.

What is interesting is that both the up- and down-regulated genes showed a signature of adaptation, although the up-regulated genes were higher. This could again reflect alternative strategies. Perhaps a clone mixes with another genotype and it downregulates a particular signal or public good to avoid mixing with that genotype. In contrast, maybe the perceived antagonist is escalating a particular signal to try to coerce the first genotype into partnership (so that it can force it into sterile stalk).

Such possibilities make this system ripe for follow-up. The signature of positive selection suggests that these genes are adaptive in the wild, and thus these genes likely underpin strategies employed by all wild *D. discoideum*. What is unclear is what phenotypes are associated with these changes in expression. Gene knockouts could possibly be employed to test for phenotypic effect, although it is possible that these genes are 'necessary' for development. In other words, knocking them out could break multicellularity all together. A more interesting question would perhaps then be one of mixing in different proportions, as the mixture used in Paper II is 50:50. Is the direction or magnitude of expression altered as mix proportions are shifted away from equality? Perhaps there are obvious cheating phenotypes that would emerge from these pairings or perhaps there is a cheating hierarchy, as has been observed in previous studies (Fortunato *et al.* 2003, Buttery *et al.* 2009). The results of this system also encourage us to move into other cooperative microbe systems, such as *Myxococcus xanthus,* to perform similar experiments. It is

yet unclear when in these systems one would predict arms races that drive positive selection in cheating/cheating avoidance genes versus when one might expect stalemate evolution. However, what is clear from this study is that regardless of how lab-adapted a microbe might seem (Tarnita 2017), we can use molecular evolution to answer questions about the relevance of genes – and ultimately phenotypes – in microbes' natural environments.

1.5. LITERATURE CITED

Aktipis, C.A. and Nesse, R.M. 2013. Evolutionary foundations for cancer biology. *Evol Appl* **6**: 144–159.

Alonso-Blanco, C., Vries, H.B., Hanhart, C.J. and Koornneef, M. 1999. Natural allelic variation at seed size loci in relation to other life history traits of *Arabidopsis thaliana*. *PNAS* **96**: 4710–4717.

Al-Shehbaz, I.A. and O'Kane, S.L. 2002. Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *Arabidopsis Book* **1**.

Arbuckle, K., Rodríguez de la Vega, R.C. and Casewell, N.R. 2017. Coevolution takes the sting out of it: Evolutionary biology and mechanisms of toxin resistance in animals. *Toxicon* **140**: 118–131.

Arnqvist, G. and Rowe, L. 2013. *Sexual Conflict*. Princeton University Press, Princeton, NJ.

Arunkumar, R., Josephs, E.B., Williamson, R.J. and Wright, S.I. 2013. Pollen-specific, but not sperm-specific, genes show stronger purifying selection and higher rates of positive selection than sporophytic genes in *Capsella grandiflora*. *Mol. Biol. Evol.* **30**: 2475–2486.

Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., *et al.* 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**: 1243–1247.

Baskin, C.C. and Baskin, J.M. 2014. *Seeds: Ecology, Biogeography, and, Evolution of Dormancy and Germination*, 2nd edition. Academic Press, San Diego, CA.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E.A., *et al.* 2009. Polymorphic members of the *lag* gene family mediate kin discrimination in *Dictyostelium*. *Curr. Biol.* **19**: 567–572.

Birnbaum, S.S.L. and Abbot, P. 2018. Insect adaptations toward plant toxins in milkweed-herbivores systems – a review. *Entomologia Experimentalis et Applicata* **166**: 357–366.

Bonner, J.T. 1967. *Cellular Slime Molds*, 2nd revised edition. Princeton University Press, Princeton, NJ.

Bonner, J.T. 2009. *The Social Amoebae: The Biology of Cellular Slime Molds*. Princeton University Press, Princeton, NJ.

Bourke, A. 2011. *Principles of Social Evolution*. Oxford University Press, Oxford, United Kingdom.

Bożek, K. and Lengauer, T. 2010. Positive selection of HIV host factors and the evolution of lentivirus genes. *BMC Evolutionary Biology* **10**: 186.

Brockhurst, M.A., Chapman, T., King, K.C., Mank, J.E., Paterson, S. and Hurst, G.D.D. 2014. Running with the Red Queen: the role of biotic conflicts in evolution. *Proceedings of the Royal Society of London B: Biological Sciences* **281**: 20141382.

Burt, A. and Trivers, R.L. 2006. *Genes in Conflict: The Biology of Selfish Genetic Elements*. Belknap, Cambridge, MA.

Buss, L.W. 1983. Evolution, development, and the units of selection. *PNAS* **80**: 1387–1391.

Buttery, N.J., Rozen, D.E., Wolf, J.B. and Thompson, C.R.L. 2009. Quantification of social behavior in *D. discoideum* reveals complex fixed and facultative strategies. *Curr. Biol.* **19**: 1373–1377.

Buttery, N.J., Jack, C.N., Adu-Oppong, B., Snyder, K.T., Thompson, C.R.L., Queller, D.C., *et al.* 2012. Structured growth and genetic drift raise relatedness in the social amoeba *Dictyostelium discoideum*. *Biol. Lett.* **8**: 794–797.

Carroll, L. 1871. *Through the Looking-Glass, and What Alice Found There*. Macmillan, London.

Chuong, E.B., Tong, W. and Hoekstra, H.E. 2010. Maternal-fetal conflict: rapidly evolving proteins in the rodent placenta. *Molecular Biology and Evolution* **27**: 1221–1225.

Cosmides, L.M. and Tooby, J. 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**: 83–129.

Costa, L.M., Gutièrrez-Marcos, J.F. and Dickinson, H.G. 2004. More than a yolk: the short life and complex times of the plant endosperm. *Trends in Plant Science* **9**: 507–514.

Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H. and Gutierrez-Marcos, J.F. 2012. Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Current Biology* **22**: 160–165.

Crepet, W.L. and Niklas, K.J. 2009. Darwin's second "abominable mystery": Why are there so many angiosperm species? *Am. J. Bot.* **96**: 366–381.

Daugherty, M.D. and Malik, H.S. 2012. Rules of engagement: molecular insights from host-virus arms races. *Annual Review of Genetics* **46**: 677–700.

Dawkins R. and Krebs, J.R. 1979. Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **205**: 489–511.

De Jong, T.J., Van Dijk, H. and Klinkhamer, P.G.L. 2005. Hamilton's rule, imprinting and parent-offspring conflict over seed mass in partially selfing plants. *J. Evol. Biol.* **18**: 676–682.

Dean, M.D., Clark, N.L., Findlay, G.D., Karn, R.C., Yi, X., Swanson, W.J., *et al.* 2009. Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Mol. Biol. Evol.* **26**: 1733–

1743.

Drake, A., Fraser, D. and Weary, D.M. 2008. Parent-offspring resource allocation in domestic pigs. *Behavioral Ecology and Sociobiology* **62**: 309–319.

Dudley, S.A. 2015. Plant cooperation. *AoB PLANTS* **7**.

Durand, S., Bouché, N., Perez Strand, E., Loudet, O. and Camilleri, C. 2012. Rapid establishment of genetic incompatibility through natural epigenetic variation. *Current Biology* **22**: 326–331.

Eyre-Walker, A. 2002. Changing effective population size and the McDonald-Kreitman test. *Genetics* **162**: 2017–2024.

Falconer, D.S. and Mackay, T.F.C. 1996. *Introduction to Quantitative Genetics*, 4th edition. Benjamin Cummings.

Fay, J.C. and Wu, C.-I. 2000. Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.

Fay, J.C. and Wu, C.-I. 2001. The neutral theory in the genomic era. *Current Opinion in Genetics and Development* **11**: 642–646.

Fiegna, F. and Velicer, G.J. 2003. Competitive fates of bacterial social parasites: persistence and self-induced extinction of *Myxococcus xanthus* cheaters. *Proc. Biol. Sci.* **270**: 1527–1534.

Fortunato, A., Queller, D.C. and Strassmann, J.E. 2003. A linear dominance hierarchy among clones in chimeras of the social amoeba *Dictyostelium discoideum*. *J. Evol. Biol.* **16**: 438–445.

Foster, K.R., Fortunato, A., Strassmann, J.E. and Queller, D.C. 2002. The costs and benefits of being a chimera. *Proc Biol Sci* **269**: 2357–2362.

Foster, K.R., Shaulsky, G., Strassmann, J.E., Queller, D.C. and Thompson, C.R.L. 2004. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**: 693–696.

Friedman, W.E. 1995. Organismal duplication, inclusive fitness theory, and altruism: understanding the evolution of endosperm and the angiosperm reproductive syndrome. *Proc Natl Acad Sci U S A* **92**: 3913–3917.

Galindo, B.E., Vacquier, V.D. and Swanson, W.J. 2003. Positive selection in the egg receptor for abalone sperm lysin. *PNAS* **100**: 4639–4643.

Gehring, M. 2013. Genomic imprinting: insights from plants. *Annual Review of Genetics* **47**: 187–208.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E. and Queller, D.C. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater

mutants. *PNAS* **104**: 8913–8917.

Gossmann, T.I., Song, B.-H., Windsor, A.J., Mitchell-Olds, T., Dixon, C.J., Kapralov, M.V., *et al.* 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant dpecies. *Mol Biol Evol* **27**: 1822–1832.

Gossmann, T.I., Keightley, P.D. and Eyre-Walker, A. 2012. The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes. *Genome Biol Evol* **4**: 658–667.

Gossmann, T.I., Schmid, M.W., Grossniklaus, U. and Schmid, K.J. 2013. Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol* **31**: 574–583.

Guevara, E.E., Veilleux, C.C., Saltonstall, K., Caccone, A., Mundy, N.I. and Bradley, B.J. 2016. Potential arms race in the coevolution of primates and angiosperms: brazzein sweet proteins and gorilla taste receptors. *Am. J. Phys. Anthropol.* **161**: 181–185.

Gutierrez-Marcos, J.F., Constância, M. and Burton, G.J. 2012. Maternal to offspring resource allocation in plants and mammals. *Placenta* **33**: e3–e10.

Haig, D. 1987. Kin conflict in seed plants. *Trends in Ecology and Evolution* **2**: 337–340.

Haig, D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* **264**: 1657–1662.

Haig, D. 2000. The kinship theory of genomic imprinting. *Annual Review of Ecology and Systematics* **31**: 9–32.

Haig, D. 2004. Genomic imprinting and kinship: how good is the evidence? *Annual Review of Genetics* **38**: 553–585.

Haig, D. 2013. Kin conflict in seed development: an interdependent but fractious collective. *Annual Review of Cell and Developmental Biology* **29**: 189–211.

Haig, D. 2014. Coadaptation and conflict, misconception and muddle, in the evolution of genomic imprinting. *Heredity* **113**: 96–103.

Hamilton, W.D. 1964a. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* **7**: 1–16.

Hamilton, W.D. 1964b. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* **7**: 17–52.

Harada, J.J. and Pelletier, J. 2012. Genome-wide analyses of gene activity during seed development. *Seed Science Research* **22**: S15–S22.

Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. 2016. Rapid evolution of

genomic imprinting in two species of the Brassicaceae. *The Plant Cell* **28**: 1815–1827.

Hirose, S., Benabentos, R., Ho, H.-I., Kuspa, A. and Shaulsky, G. 2011. Self-recognition in social amoebae is mediated by allelic pairs of tiger genes. *Science* **333**: 467–470.

Ho, H.-I., Hirose, S., Kuspa, A. and Shaulsky, G. 2013. Kin recognition protects cooperators against cheaters. *Curr. Biol.* **23**: 1590–1595.

Hollis, B. 2012. Rapid antagonistic coevolution between strains of the social amoeba *Dictyostelium discoideum*. *Proc Biol Sci* **279**: 3565–3571.

Houston, A.I., Székely, T. and McNamara, J.M. 2005. Conflict between parents over care. *Trends in Ecology and Evolution* **20**: 33–38.

Ingram, G.C. 2010. Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195–214.

Ingvarsson, P.K. 2005. Molecular population genetics of herbivore-induced protease inhibitor genes in European aspen (*Populus tremula* L., Salicaceae). *Mol. Biol. Evol.* **22**: 1802–1812.

Jander, G. 2018. Revisiting plant-herbivore co-evolution in the molecular biology era. *Annual Plant Reviews*, pp. 361–384.

Jansa, S.A. and Voss, R.S. 2011. Adaptive evolution of the venom-targeted vWF protein in opossums that eat pitvipers. *PLOS ONE* **6**: e20997.

Janzen, F.J. and Warner, D.A. 2009. Parent–offspring conflict and selection on egg size in turtles. *Journal of Evolutionary Biology* **22**: 2222–2230.

Jin, W., Wu, D.-D., Zhang, X., Irwin, D.M. and Zhang, Y.-P. 2012. Positive selection on the gene RNASEL: correlation between patterns of evolution and function. *Mol. Biol. Evol.* **29**: 3161–3168.

Kessin, R.H. 2010. *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity*, reissue edition. Cambridge University Press, Cambridge, United Kingdom.

Khare, A., Santorelli, L.A., Strassmann, J.E., Queller, D.C., Kuspa, A. and Shaulsky, G. 2009. Cheater-resistance is not futile. *Nature* **461**: 980–982.

Kilner, Rebecca M. and Hinde, Camilla A. 2012. Parent-offspring conflict. *The Evolution of Parental Care* (N. J. Royle, Smiseth, Per T., and Kölliker, Mathias, eds), pp. 119–132. Oxford University Press, Oxford, United Kingdom.

Kimura, M. 1983. *Neutral Theory Molecular Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Kocher, S.D., Tsuruda, J.M., Gibson, J.D., Emore, C.M., Arechavaleta-Velasco, M.E., Queller, D.C., *et al.* 2015. A search for parent-of-origin effects on honey bee gene expression. *G3* **5**: 1657–1662.

Köhler, C. and Weinhofer-Molisch, I. 2010. Mechanisms and evolution of genomic imprinting in plants. *Heredity* **105**: 57–63.

Kousathanas, A., Halligan, D.L. and Keightley, P.D. 2014. Faster-X adaptive protein evolution in house mice. *Genetics* **196**: 1131–1143.

Krämer, U. Planting molecular functions in an ecological context with *Arabidopsis thaliana*. *eLife* **4**: e06100.

Kuzdzal-Fick, J.J., Fox, S.A., Strassmann, J.E. and Queller, D.C. 2011. High relatedness is necessary and sufficient to maintain multicellularity in *Dictyostelium*. *Science* **334**: 1548–1551.

Lawton, B.R., Sevigny, L., Obergfell, C., Reznick, D., O'Neill, R.J. and O'Neill, M.J. 2005. Allelic expression of IGF2 in live-bearing, matrotrophic fishes. *Dev. Genes Evol.* **215**: 207–212.

Levin, S.R., Brock, D.A., Queller, D.C. and Strassmann, J.E. 2015. Concurrent coevolution of intra-organismal cheaters and resisters. *J. Evol. Biol.* **28**: 756–765.

Lewis, D. 1941. Male sterility in natural populations of hermaphrodite plants. The equilibrium between females and hermaphrodites to be expected with different types of inheritance. *The New Phytologist* **40**: 56–63.

Lin, J.-Y., Le, B.H., Chen, M., Henry, K.F., Hur, J., Hsieh, T.-F., *et al.* 2017. Similarity between soybean and *Arabidopsis* seed methylomes and loss of non-CG methylation does not affect seed development. *PNAS* **114**: E9730–E9739.

Linkies, A., Graeber, K., Knight, C. and Leubner-Metzger, G. 2010. The evolution of seeds. *New Phytologist* **186**: 817–831.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., *et al.* 2018. Biparental inheritance of mitochondrial DNA in humans. *PNAS* **115**: 13039–13044.

Lynch, M. 2007. *The Origins of Genome Architecture*, 1st edition. Sinauer Associates Inc., Sunderland, MA.

Maynard Smith, J. and Szathmáry, E. 1995. *The Major Transitions in Evolution*. W. H. Freeman, Oxford.

McDonald, J.H. and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

Mehdiabadi, N.J., Jack, C.N., Farnham, T.T., Platt, T.G., Kalla, S.E., Shaulsky, G., *et al.* 2006.

Social evolution: kin preference in a social microbe. *Nature* **442**: 881–882.

Messer, P.W. and Petrov, D.A. 2013. Frequent adaptation and the McDonald–Kreitman test. *PNAS* **110**: 8615–8620.

Mock, D.W. and Parker, G.A. 1997. *The Evolution of Sibling Rivalry*. Oxford University Press, Oxford, United Kingdom.

Moles, A.T. and Westoby, M. 2006. Seed size and plant strategy across the whole life cycle. *Oikos* **113**: 91–105.

Moles, A.T. and Westoby, M. 2004. Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology* **92**: 372–383.

Monk, D. 2015. Genomic imprinting in the human placenta. *Am. J. Obstet. Gynecol.* **213**: S152-162.

Moore, C.R., Gronwall, D.S., Miller, N.D. and Spalding, E.P. 2013. Mapping quantitative trait loci affecting *Arabidopsis thaliana* seed morphology features extracted computationally from images. *G3* **3**: 109–118.

Moore, T. and Mills, W. 2008. Evolutionary theories of imprinting – enough already! *Genomic Imprinting* (J. F. Wilkins, ed), pp. 116–122. Springer, New York, NY.

Neale, D.B., Marshall, K.A. and Sederoff, R.R. 1989. Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 9347–9349.

Nei, M., Suzuki, Y. and Nozawa, M. 2010. The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet* **11**: 265–289.

Nielsen, R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics* **39**: 197–218.

Ohta, T. 2002. Near-neutrality in evolution of genes and gene regulation. *PNAS* **99**: 16134–16137.

Orr, H.A. 2005. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**: 119.

Ostrowski, E.A., Shen, Y., Tian, X., Sucgang, R., Jiang, H., Qu, J., *et al.* 2015. Genomic signatures of cooperation and conflict in the social amoeba. *Curr. Biol.* **25**: 1661–1665.

Panhuis, T.M. and Swanson, W.J. 2006. Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. *Genetics* **173**: 2039–2047.

Parikh, A., Miranda, E.R., Katoh-Kurasawa, M., Fuller, D., Rot, G., Zagar, L., *et al.* 2010. Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol.* **11**: R35.

Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R. and Wolf, J.B. 2014. The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**: 119–128.

Pröschel, M., Zhang, Z. and Parsch, J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* **174**: 893–900.

Queller, D.C. 1983. Kin selection and conflict in seed maturation. *Journal of Theoretical Biology* **100**: 153–172.

Queller, D.C. 1984. Models of kin selection on seed provisioning. *Heredity* **53**: 151–165.

Queller, D.C. 1989. The Sociobiology of Plants Plant Reproductive Ecology. *Science* **243**: 244–244.

Queller, D.C. 1997. Cooperators since life began. *The Quarterly Review of Biology* **72**: 184–188.

Queller, D.C. 2014. Joint phenotypes, evolutionary conflict and the fundamental theorem of natural selection. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **369**: 20130423.

Queller, D.C. and Strassmann, J.E. 2009. Beyond society: the evolution of organismality. *Phil. Trans. R. Soc. B* **364**: 3143–3155.

Queller, D.C. and Strassmann, J.E. 2018. Evolutionary conflict. *Annual Review of Ecology, Evolution, and Systematics* **49**: 73–93.

Radwan, J. and Babik, W. 2012. The genomics of adaptation. *Proc Biol Sci* **279**: 5024–5028.

Rand, D.M. and Kann, L.M. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**: 735–748.

Raper, K.B. 1984. *The Dictyostelids*, 1st edition. Princeton Univ Press, Princeton, NJ.

Ratnieks, F.L.W., Foster, K.R. and Wenseleers, T. 2006. Conflict resolution in insect societies. *Annual Review of Entomology* **51**: 581–608.

Ratnieks, F.L.W. and Reeve, H.K. 1992. Conflict in single-queen hymenopteran societies: the structure of conflict and processes that reduce conflict in advanced eusocial species. *Journal of Theoretical Biology* **158**: 33–65.

Renfree, M.B., Ager, E.I., Shaw, G. and Pask, A.J. 2008. Genomic imprinting in marsupial placentation. *Reproduction* **136**: 523–531.

Saez-Aguayo, S., Rondeau-Mouro, C., Macquet, A., Kronholm, I., Ralet, M.-C., Berger, A., *et al.* 2014. Local evolution of seed flotation in *Arabidopsis*. *PLOS Genetics* **10**: e1004221.

Sakai, S. 2010. Evolution of exalbuminous seeds as a result of competition between maternally derived and paternally derived genes. *Evolutionary Ecology Research* **12**: 855–871.

Santorelli, L.A., Thompson, C.R.L., Villegas, E., Svetz, J., Dinh, C., Parikh, A., *et al.* 2008. Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature* **451**: 1107–1110.

Sawyer, S.L., Wu, L.I., Emerman, M. and Malik, H.S. 2005. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proc. Natl. Acad. Sci. U.S.A.* **102**: 2832–2837.

Schrader, M. and Travis, J. 2009. Do embryos influence maternal investment? Evaluating maternal-fetal coadaptation and the potential for parent-offspring conflict in a placental fish. *Evolution* **63**: 2805–2815.

Shaanker, R.U., Ganeshaiah, K.N. and Bawa, K.S. 1988. Parent-Offspring Conflict, Sibling Rivalry, and Brood Size Patterns in Plants. *Annual Review of Ecology and Systematics* **19**: 177–205.

Simon, M., Durand, S., Pluta, N., Gobron, N., Botran, L., Ricou, A., *et al.* 2016. Genomic conflicts that cause pollen mortality and raise reproductive barriers in *Arabidopsis thaliana*. *Genetics* **203**: 1353–1367.

smith, j., Queller, D.C. and Strassmann, J.E. 2014. Fruiting bodies of the social amoeba *Dictyostelium discoideum* increase spore transport by *Drosophila*. *BMC Evolutionary Biology* **14**: 105.

smith, j., Strassmann, J.E. and Queller, D.C. 2016. Fine-scale spatial ecology drives kin selection relatedness among cooperating amoebae. *Evolution* **70**: 848–859.

Sniegowski P.D. and Gerrish P.J. 2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 1255–1263.

Spencer, H.G. and Clark, A.G. 2014. Non-conflict theories for the evolution of genomic imprinting. *Heredity* **113**: 112–118.

Stoletzki, N. and Eyre-Walker, A. 2011. Estimation of the neutrality index. *Mol. Biol. Evol.* **28**: 63–70.

Strassmann, J.E. 2016. Kin discrimination in *Dictyostelium* social amoebae. *Journal of Eukaryotic Microbiology* **63**: 378–383.

Strassmann, J.E. and Queller, D.C. 2010. The social organism: congresses, parties, and committees. *Evolution* **64**: 605–616.

Strassmann, J.E. and Queller, D.C. 2011. Evolution of cooperation and control of cheating in a social microbe. *PNAS* **108**: 10855–10862.

Strassmann, J.E., Zhu, Y. and Queller, D.C. 2000a. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**: 965–967.

Swanson, W.J., Yang, Z., Wolfner, M.F. and Aquadro, C.F. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *PNAS* **98**: 2509–2514.

Swanson, W.J. and Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137–144.

Szathmáry, E. and Smith, J.M. 1995. The major evolutionary transitions. *Nature* **374**: 227–232.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Tarnita, C.E. 2017. The ecology and evolution of social behavior in microbes. *Journal of Experimental Biology* **220**: 18–24.

Templeton, A.R. 2006. *Population Genetics and Microevolutionary Theory*, 1st edition. Wiley-Liss, Hoboken, NJ.

Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., *et al.* 2012. The molecular diversity of adaptive convergence. *Science* **335**: 457–461.

Thompson, K., Band, S.R. and Hodgson, J.G. 1993. Seed size and shape predict persistence in soil. *Functional Ecology* **7**: 236–241.

Törjék, O., Witucka-Wall, H., Meyer, R.C., von Korff, M., Kusterer, B., Rautengarten, C., *et al.* 2006. Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theor. Appl. Genet.* **113**: 1551–1561.

Trivers, R.L. 1974. Parent-offspring conflict. *Amer. Zool.* **14**: 249–264.

Tuteja, R., McKeown, P.C., Ryan, P., Morgan, C.C., Donoghue, M.T.A., Downing, T., *et al.* 2019. Paternally expressed imprinted genes under positive Darwinian selection in *Arabidopsis thaliana*. *Mol Biol Evol* **36**: 1239–1253.

Van Dyken, J.D.V. and Wade, M.J. 2012. Detecting the molecular signature of social conflict: theory and a test with bacterial quorum sensing genes. *The American Naturalist* **179**: 436–450.

Van Valen, L. 1973. A new evolutionary theory. *Evolutionary Theory* **1**: 1–30.

Velicer, G.J. and Vos, M. 2009. Sociobiology of the *Myxobacteria*. *Annu. Rev. Microbiol.* **63**:

599–623.

Weir, B.S. and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Werren, J.H. 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences* **108**: 10863–10870.

West-Eberhard, M.J. 1975. The evolution of social behavior by kin selection. *The Quarterly Review of Biology* **50**: 1–33.

Westoby, M. and Rice, B. 1982. Evolution of the seed plants and inclusive fitness of plant tissues. *Evolution* **36**: 713–724.

Wielgoss, S., Fiegna, F., Rendueles, O., Yu, Y.-T.N. and Velicer, G.J. 2018. Kin discrimination and outer membrane exchange in *Myxococcus xanthus*: a comparative analysis among natural isolates. *Molecular Ecology* **27**: 3146–3158.

Wilburn, D.B. and Swanson, W.J. 2016. From molecules to mating: rapid evolution and biochemical studies of reproductive proteins. *Journal of Proteomics* **135**: 12–25.

Wilkins, J.F. and Haig, D. 2003. What good is genomic imprinting: the function of parent-specific gene expression. *Nature Reviews Genetics* **4**: 359–368.

Wilson, A.J., Pilkington, J.G., Pemberton, J.M., Coltman, D.W., Overall, A.D.J., Byrne, K.A., *et al.* 2005. Selection on mothers and offspring: whose phenotype is it and does it matter? *Evolution* **59**: 451–463.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., *et al.* 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genetics* **7**: e1002126.

Yang, Z. and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**: 496–503.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

Zhang, W.-H., Zhou, Y., Dibley, K.E., Tyerman, S.D., Furbank, R.T. and Patrick, J.W. 2007. Nutrient loading of developing seeds. *Functional Plant Biol.* **34**: 314–331.

KIN CONFLICTS DRIVE RAPID ADAPTATION OF IMPRINTED GENES IN *ARABIDOPSIS*

2.1. ABSTRACT

Reciprocal conflict between two parties can lead to an evolutionary arms race, which can accelerate genic evolution. The Kinship Theory posits that parent-of-origin allelic expression (imprinting) can lead to an arms race between mothers and fathers over maternal resources provided during offspring development because of kin conflicts. However, mother-offspring conflicts over maternal resources also exist in the seed. We use the proportion, α, and rate, ω_a, of adaptive substitutions to test whether imprinted genes' adaptive signature reflects a mother-father conflict distinct from the seed's mother-offspring conflict. In *Arabidopsis*, we show that imprinted loci have a greater adaptive signature than other seed genes, despite a previously demonstrated mother-endosperm arms race. We show that imprinted patrigenes have a greater adaptive signature than matrigenes. We test a proximate selection-exposure hypothesis as an alternative explanation for the rapid adaptation of imprinted genes but find no support for it. Overall, our results suggest *Arabidopsis* imprinted genes are most likely to evolve because of kin-selected conflict.

Keywords: arms race, imprinting, evolutionary conflict, kinship theory, DFE-alpha, proportion of adaptive substitutions

2.2. INTRODUCTION

Factors that affect seed size have a profound impact on human existence because so much of our food comes from the nutrient-rich endosperm of seeds. Genomic imprinting – parent-specific gene expression – is both an evolutionary and mechanistic factor that can drastically alter seed size (Bai and Settles 2015). Imprinting results from epigenetic modification that identifies an allele's parental origin including DNA methylation and histone modification (Köhler *et al.* 2012). We use the term 'matrigene' and 'patrigene' to refer to an allelic copy derived from the maternal or paternal parent, respectively, regardless of imprinting status (Queller 2003).

In angiosperms, genomic imprinting appears to be isolated to the developing seed, particularly the endosperm (Hsieh *et al.*, 2011; Gehring *et al.* 2011; Pires and Grossniklaus 2014) and the developing embryo to a lesser extent (Jahnke and Scholten, 2009; Nodine and Bartel, 2012; Raissig *et al.* 2013; Pignatta et al., 2014). In mammals, the placenta – which transfers nutrients from mother to offspring – is the primary location of imprinted gene expression, though some is found in other tissues **(**Coan *et al.* 2005; Prickett and Oakley 2012). Thus, in both taxa, imprinted loci are predominantly found in extra-embryonic, resource provisioning tissues. This suggests that imprinted genes play integral roles in offspring development and maternal resource provisioning (Wilkins and Haig 2003).

One evolutionary theory that explains the observed distribution and phenotypic effects of imprinted loci is the kinship theory of imprinting (Haig 2000, Moore and Mills 2008). It posits that imprinting results from conflict between maternal and paternal genomes over maternal investment to the embryo (Haig 1997, Haig 2000). However, in angiosperms imprinted genes

will have evolved in the context of mother-offspring conflict between mother and endosperm and to a lesser extent between mother and embryo (Paper I).

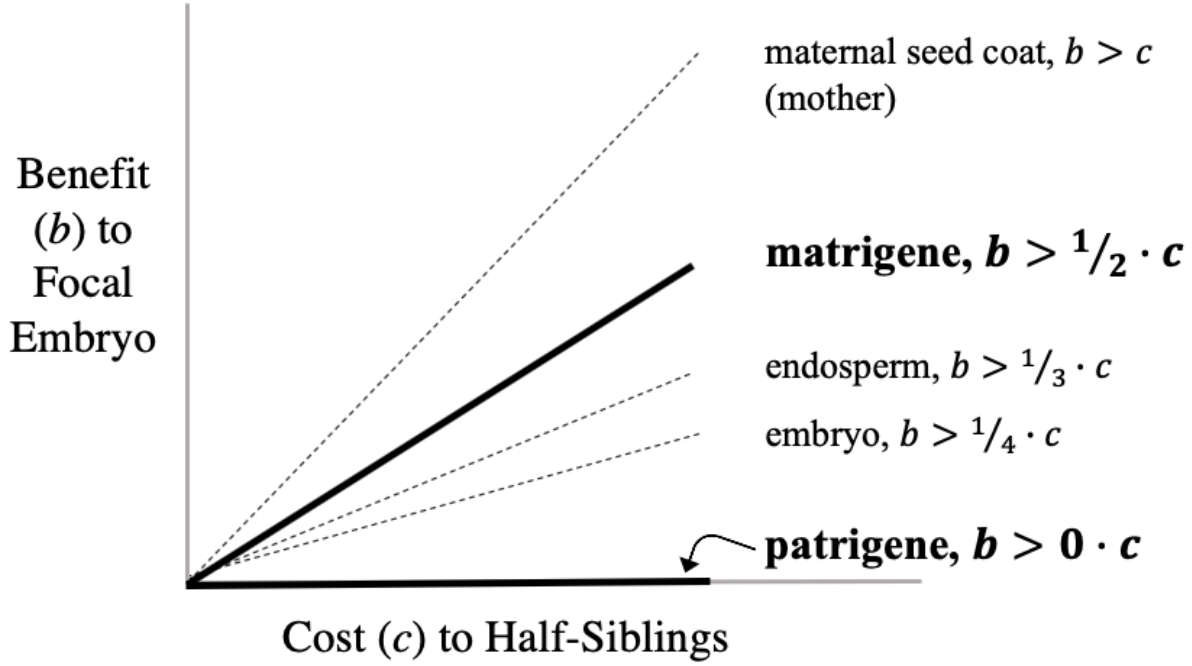


FIGURE 2.1. PREDICTED CONFLICT BETWEEN IMPRINTED GENES AND OTHER SEED GENES. Here we assume an allele of locus *x* (where *x* is the maternal tissue, endosperm, embryo, matrigene, or patrigene) increases flow of nutrients to the focal embryo. It thereby increases the embryo's fitness by *b* (y-axis) at the expense of the total fitness *c* (x-axis) of its current and future half-siblings. We apply Hamilton's rule of kin selection (1964a;1964b, Box 1) such that each party will favor nutrient transfer when $r_{embryo\text{-}x} \cdot b - r_{halfsib\text{-}x} \cdot c > 0$, or $b > (r_{halfsib\text{-}x}/r_{embryo\text{-}x}) \cdot c$, where $r_{embryo\text{-}x}$ and $r_{halfsib\text{-}x}$ are the relatedness coefficients of the party *x* to the focal embryo and its half-siblings, respectively. Conflict occurs when this condition is fulfilled for one party (above its line) but not for another party (below its line) Thus, if we assume that matrigenes and patrigenes are selected independently, patrigenes would never be selected to allow resources to go to maternal half-siblings, while matrigenes sometimes would be.

Thus, to understand how conflict between imprinted genes and the mother plant over resource transfer could arise, we must place it relative to the endosperm and the embryo (Figure 2.1). Let us consider maternal resource transfer as a joint phenotype (Queller and Strassmann 2018) – one that is influenced and may evolve via involvement from two or more parties. This allows us to work out the evolutionary conflicts that drive the evolution of genes after they have become imprinted. In Figure 2.1, each line represents benefit/cost (b/c) ratios where no net kin selection occurs on that party for nutrient transfer. We base these lines on inclusive fitness (Hamilton 1964a;1964b) for the choice of whether to allocate more resources to the focal embryo at a cost to half-sibling embryos (Queller 1983;1984, Paper I). The lines represent a point of indifference; above a party's line, that party should favor the transfer, below they should disfavor it. The space between any two of these indifference lines is a conflict zone where one party benefits from more investment to the focal embryo and the other from less. Importantly, this illustrates the distinct interests of matrigenes and patrigenes from each other and from other parties. For example, this is seen in Figure 2.1 as zones of conflict between the mother plant and matrigenes (smaller region, Figure 2.1) or patrigenes (much larger region).

What also becomes apparent from Figure 2.1 is that, in the case of full outbreeding, a patrigene always benefits from increasing nutrient flow to its own embryo at the expense of other fathers' embryos (because patrigenes are not related to these other embryos). Thus, patrigenes will be favored to promote growth of their own offspring (Haig 2000; 2013) and empirical evidence from plants supports this (Raunsgard *et al.* 2018; Willi 2013). On the other hand, a matrigene is equally related to its maternal half-siblings. It should therefore be selected to forego resources for its own embryo if the cost to half-siblings is too high. Directly in between matrigenes and patrigenes is the embryo (unimprinted genes) because it consists of one

matrigene and one patrigene (Figure 2.1). The triploid endosperm, with two matrigenic copies, lies closer to the matrigene line. Lastly, the mother should be selected to give fewer resources to the focal embryo than is favored by either matrigene or patrigene. This is because, unlike them, the mother is equally related to all of her embryos.

This potential conflict over maternal resource allocation could lead to an arms race between imprinted loci and the mother plant, endosperm, or embryo. This arms race could in turn lead to rapid divergence in the gene sequences of imprinted genes, which appears as an accumulation of nonsynonymous mutations relative to synonymous ones (Patten *et al*. 2014). Most mutations within genes are assumed to be nonsynonymous and thus deleterious because they change the amino acid they encode. They are usually selected against in the population quickly. Conversely, synonymous mutations are assumed to be neutral or nearly neutral. They are not directly selected against and more likely to persist. If a gene shows an excess of nonsynonymous mutations leading to divergence between two species relative to synonymous ones, it implies the excess was fixed by adaptation (Yang and Bielawski 2000).

However, genes vary in their individual mutation or recombination rates and in the strength of purifying selection (Bierne and Eyre-Walker 2004). We can control for the latter by including information on nonsynonymous and synonymous polymorphisms, for example, with a McDonald-Kreitman (MK) test (1991) or its extensions. The MK test uses raw counts of nonsynonymous (Dn) and synonymous (Ds) divergence and nonsynonymous (Pn) and synonymous (Ps) polymorphism. It compares Dn/Ds to Pn/Ps, such that Dn/Ds > Pn/Ps indicates an excess of nonsynonymous divergence fixed by adaptation (McDonald-Kreitman 1991). Pn/Ps gives us a baseline of purifying selection with which we can compare Dn/Ds. Extensions of the MK test include the proportion α (Fay and Wu 2001; Smith and Eyre-Walker 2002; Eyre-Walker

and Keightley 2009) and rate ω_a (Gossmann *et al.* 2010) of adaptive substitutions. Both measures give us enhanced power to detect selection because they allow us to summarize over a set of genes. An elevated α or ω_a for a gene set indicates that, on average, these genes have experienced adaptive evolution. Thus, if imprinted genes have evolved in an arms races over maternal resource allocation with the mother plant, as predicted by the kinship theory, they should exhibit an elevated α or ω_a relative to other genes of the genome.

In angiosperms, tests of molecular evolution have thus far largely supported the arms race predictions of the kinship theory because genes show evidence of positive selection (*e.g.,* Spillane *et al.* 2007, Kawabe *et al.* 2007). A more recent analysis of N=31 *A. thaliana* imprinted genes showed evidence of positive selection measured as an elevated rate of pairwise nonsynonymous divergence between *A. thaliana* and *A. lyrata* (Wolff *et al.* 2011). This elevated rate of nonsynonymous divergence was found on the branch leading to the *A. thaliana* lineage for N=62 imprinted genes using *A. thaliana* and orthologs from 32 plant species (Tuteja *et al.* 2019). Additionally, the *Arabidopsis* relative *Capsella rubella* exhibits an elevated rate of adaptive substitutions (ω_a) in endosperm-imprinted genes relative to other genes in the genome (Hatorangan *et al.* 2016). However, note that a combined N=889 confirmed and putatively imprinted genes have been identified in the endosperm and embryo of *A. thaliana*. Yet, most of these have not been tested for rapid adaptation.

Evidence of rapid, adaptive evolution in plants and particularly *A. thaliana* seem consistent with kin conflict and have largely been interpreted as such (Wolff *et al.* 2011; Hatorangan *et al.* 2016, Tuteja *et al.* 2019). But is this really due to imprinting? The challenge is that imprinted genes are evolving in the context of the larger conflict over maternal resource allocation to the developing embryo (Figure 2.1, Paper I). The seed genes of *Arabidopsis,*

particularly those enriched in nutrient allocation tissues, show signatures of rapid adaptation consistent with conflict over maternal resources (Paper I). The two tissues predicted to be entrenched in conflict, the mother plant and endosperm, have evolved with more adaptive substitutions (higher α) than the embryo. However, imprinting is predominantly found in the endosperm. Here, we test if a conflict-driven arms race over maternal resources between imprinted genes and the mother plant is distinct from that of the mother and endosperm. If so, we expect to see a higher α and ω_a in imprinted genes of both endosperm and embryo relative to their respective backgrounds. On the other hand, elevated α of the endosperm only would suggest selection on imprinted genes is due to the overarching mother-endosperm conflict occurring in the seed.

If imprinted genes have an elevated α relative to other genes of the endosperm and embryo, the question then becomes whether there are any alternative explanations. Other ultimate theories of imprinting do not involve conflict over maternal resource allotment (Table 2.1, see also Spencer and Clark 2014). Thus, they do not predict an elevated α in imprinted genes relative to their background seed tissues. An elevated α for imprinted genes relative to both the endosperm or embryo would support the kin conflict hypothesis and not the other non-conflict ultimate hypotheses.

There is one proximate, or mechanistic, alternative to the kinship theory we need to examine, though. The complete silencing of maternal or paternal alleles results in effective haploidy in heterozygous endosperms. This has the important potential to unmask the effects of recessive alleles and expose them to stronger selection (Spencer and Clark 2014). When a new recessive allele is rare, it will occur mostly in heterozygotes. Thus, imprinting will cause stronger selection against deleterious recessives (Pàl *et al.* 2006, Immler and Otto 2018), which would

decrease Pn/Ps. It will simultaneously cause stronger selection on beneficial recessives, which would increase Dn/Ds. The cumulative effect of these changes is to increase the proportion (α) and rate (ω_a) of adaptive evolution. A similar argument applies to selection on the X chromosome (Meisel *et al.* 2013). This effect should also apply, though to a lesser degree, to partially silenced genes. Thus, a higher α for imprinted genes could be due either to conflict or to greater exposure of imprinted genes to selection. The ultimate kin conflict and the proximate selection-exposure hypotheses make different predictions on adaptive evolution for matrigenes versus patrigenes, which we test here.

Previous evidence of asymmetric selection on patrigenes versus matrigenes in *A. thaliana* shows possible higher adaption of patrigenes. Tuteja *et al.* (2019) found evidence of positive selection in patrigenes but not matrigenes from N=62 total imprinted genes in the branch leading to *A. thaliana*. They interpreted this as a result of reciprocal evolution between patrigenes and the maternal plant, a type of interlocus sexual conflict (Willi 2013; Tuteja *et al.* 2019). It is important to note that this is still a conflict-driven arms race and predicted by Figure 2.1. We agree with the basic logic because the zone of conflict between patrigenes and mother plants is the largest of all (Figure 2.1), but a more complete justification for the prediction should consider the full context of all seed parties.

Any conflict between matrigenes and patrigenes would not occur in isolation; they are in conflict in the context of unimprinted maternal-endosperm-embryo conflict over maternal resource allocation (Paper I). Patrigenes would be in greater conflict with these other players in the seed than matrigenes with the exception of the embryo (Figure 2.1). Because the patrigene line is farther from the maternal and endosperm lines than the matrigene line, selection would be stronger on patrigenes over matrigenes. Matrigenes and the maternal plant are related

Why Hypothesis Does Not Predict Rapid Adaptation	Hypothesis	Reference
It is a **cooperation hypothesis**. Although selection could favor matrigenes and/or patrigenes depending on theory, positive selection would not be rapid.	- Maternal-Fetal Co-adaptation	Wolf and Hager 2006
	- Maternal-Fetal Coordination	Keverne and Curley 2008; Miri and Varmuza 2009
	- Cytonuclear Interactions	Wolf 2009
	- Co-adaptation of Gene Expression	Wolf 2013
Hypothesis **predicts no ongoing conflict** between mothers and fathers over imprinting status. It was presented as an origin hypothesis.	- Ovarian Time Bomb	Varmuza and Mann 1994
	- Prevention of Parthenogenesis	Solter 1988
Hypothesis has very **limited scope.** It is unlikely to apply to all imprinted loci or does not apply to plants.	- Chip Off the Old Block	Spencer and Clark 2006
	- Sex-Linked Segregation Distortion	Úbeda and Haig 2004
	- X-linked Sex-specific Selection	Iwasa and Pomiankowski 1999
Although proposed as an origin hypothesis, it could explain maintenance of imprinting. However, it makes **no prediction about differential selection on matrigenes and patrigenes**.	- Minimization of Variance	Solter 1988
	- Complementation of Variance	Kaneko-Ishino *et al.* 2003
Hypothesis is insufficient on its own to explain selection on imprinted genes. It makes **no predictions relevant to our study**.	- Dominance Modification	Sapienza 1989
	- Host Genome Defense	Barlow 1993; McDonald 1999; McDonald *et al.* 2005
Hypothesis **does not predict conflict-of-interests** between males and females. Rather, it is about fixing a design trade-off conflict such that both parties benefit.	- Intralocus Sexual Conflict	Day and Bonduriansky 2004

TABLE 2.1. NON-CONFLICT ULTIMATE HYPOTHESES OF GENOMIC IMPRINTING AND WHY THEY DO NOT PREDICT RAPID EVOLUTION OF IMPRINTED GENES. Many non-conflict hypotheses have been proposed to explain the origin and maintenance of imprinting. However, none predict rapid evolution (high α or ω_a) even if the hypothesis is a selection hypothesis.

to half-siblings of the focal embryo whereas patrigenes are not. This would create large conflict between patrigenes and the mother plant, whereas matrigenes would be in less conflict (Figure 2.1: compare the region between patrigene and maternal plant, vs. matrigene and maternal plant). The unimprinted endosperm line in Figure 2.1 is also closer to the matrigene line than patrigenes because the endosperm contains a double dose of the maternal genome. Thus, selection on the

mother plant, and to a lesser degree the unimprinted endosperm, will pull the system away from patrigenic interests more often than away from matrigenic interests. This creates a potential for stronger selection on novel, adaptive patrigenic alleles.

In contrast, the selection-exposure hypothesis does not predict higher patrigene α. For imprinted genes expressed in embryos, matrigenes and patrigenes would be equally exposed to selection because silencing one allele in heterozygotes fully exposes the other allele. For endosperms, with their double dose of the maternal gene, it is less clear. Again, a complete silencing of either gene totally exposes the other, but it seems possible that a partial silencing allows greater selection on matrigenes, since they will be expressed more. But in both scenarios, faster evolution of patrigenes is not predicted.

Here we measure the proportion α and rates ω_a of adaptive evolution in *Arabidopsis* imprinted genes to test the predictions of the kinship theory. We try to distinguish our findings among alternative hypotheses. We find that our results best support the kin conflict theory for the evolution of imprinted genes in *Arabidopsis*.

2.3. METHODS

Imprinted Gene Sets

We used all recognized and putative imprinted loci in *A. thaliana,* dividing them into two "confirmed" (empirically validated) and "candidate" (high-throughput). In *A. thaliana*, the majority of imprinted loci have been identified in the endosperm (*e.g.,* Hsieh *et al.* 2011; Gehring *et al.* 2011) (N=625), with a smaller number identified through high throughput methods in the embryo (*e.g.*, Raissig *et al.* 2013; Pignatta *et al.*, 2014) (N=155). We subdivided imprinted loci based on where they were imprinted, as well as if they are a matrigene (N=664) or

patrigene (N=179). We also subdivided matrigenes and patrigenes based on whether they were imprinted in endosperm (matrigene N=545, patrigene N=113) or embryo (matrigene N=138, patrigene N=43).

Because the majority of genes were identified by high-throughput screens, they may be biased by seed coat contamination (Schon and Nodine 2017). We also analyzed the imprinted gene set identified by Schon and Nodine (2017) (N=123, their 'stringent' set), which we also subdivided into matrigenes (N=57) and patrigenes (N=67; one gene is biallelic and overlaps both sets). Because matrigenes and patrigenes identified by that study are imprinted in the endosperm only, it was not possible to make an endosperm vs. embryo comparison on that gene set.

To test each of the predictions of the ultimate and proximate hypotheses, we often include genes from the genomic background or from the seed or compartments of the seed (endosperm or embryo) for comparison. Genes from the genomic background were taken from *A. thaliana* and include all genes except those in our imprinted sets. Genes from the seed, embryo, or endosperm, as appropriate, refer to genes identified by Paper I and also exclude those in our imprinted sets.

Tests for molecular adaptation in imprinted loci

To test for sequence-level adaptation in imprinted genes, we combined inter- and intra-specific sequence data in a modified McDonald-Kreitman (1991) test. We estimated both the proportion (α) and rate (ω_a) of adaptive substitutions (Smith and Eyre-Walker 2002; Eyre-Walker and Keightley 2009; Gossmann *et al.* 2010). Here $\alpha = 1 - (\frac{Pn}{Ps} / \frac{Dn}{Ds})$, where *Pn* and *Ps* are counts of nonsynonymous and synonymous polymorphisms within species, respectively, and *Dn* and *Ds* are counts of nonsynonymous and synonymous substitutions between a species pair,

respectively (Fay and Wu 2001). Under the McDonald-Kreitman test, if the ratio *Dn/Ds* exceeds *Pn/Ps*, the sequence is thought to be under strong positive selection (McDonald and Kreitman 1991). This is predicated on the assumption that any adaptive mutations will fix quickly between species, thus contributing to between-species divergence beyond the level of within-species polymorphism, which is determined primarily by mutation, drift, and purifying selection. However, the McDonald-Kreitman test has a low power to detect adaptation when it is performed on a single gene. The metric α combines data from a class of genes, thereby leveraging increased sample size to improve the power to detect adaptive fixation of mutations (Eyre-Walker and Keightley 2009). Since α represents a proportion and not a rate, we also report the rate of adaptive fixation, ω_a (Gossmann 2010). Because adaptive fixation can be strongly biased by effective population size (N_e), the alternative parameterizaion of $\alpha - \omega_a$ – allows us to account for any effect of N_e on fixation of adaptive alleles. In other words, α tells us about the accumulation of nonsynonymous substitutions relative to nonsynonymous ones, whereas ω_a tells us whether it could be attributed to effects of N_e. We provide both estimates in this study.

We estimated pairwise divergence between *thaliana-lyrata* or *lyrata-halleri* as described in Paper I. In brief, we identified whole-genome orthologs using a custom version of InParanoid v.4 (Remm *et al.* 2001) using the *A. thaliana* (TAIR10, https://www.arabidopsis.org), *A. lyrata* (*v.* 1.0, Hu *et al.* 2011), and *A. halleri* (*v.* 1.0, Briskine *et al.* 2017) genomes. We aligned protein sequence of the 1:1 orthologs using MUSCLE (Edgar *et al.* 2004), which we back-translated with PAL2NAL v. 14 (Suyama *et al.* 2006) and trimmed with trimAl v. 1.2 (Capella-Gutierrez *et al.* 2009). We estimated *Dn* and *Ds* counts with the Nei and Gojobori method (1986) in the codeml package of PAML v. 4.0 (Yang 2007 (runmode = -2, CodonFreq = 2). All genes with

saturated divergence were excluded (dS ≥ 1, *thaliana-lyrata* 61 out of 21,292 genes, *lyrata-halleri* 37 out of 14,843 genes).

We estimated polymorphism from *A. thaliana* and *A. lyrata* as described in detail in Paper I and summarized below. We wanted to reduce the odds of spurious signals of adaptation due to recent population history, thus we estimated polymorphism for five populations of *A. thaliana* and two populations of *A. lyrata.* For *A. thaliana*, we estimated polymorphism counts from SNP data from re-sequenced genomes (1,001 Genomes Project, http://www.1001genomes.org). The five populations contain geographically clustered accessions to reduce any effects of population structure: Germany (N=43), Czech (N=17), Russia (N=21), E. Spain (N=18), and W. Spain (N=23). For A. lyrata, we estimated polymorphism counts from SNP data from two populations from Eria, PA (N=14) and Jamesville, NY (N=25) (Fracassetti *et al.* 2015). These SNP data came from both pooled resquenced genomes and genotype-by-sequencing (GBS) data.

To get the estimates of *Pn* and *Ps* for each population, we performed the following for each population. We converted the variant call format files for each accession to a FASTA of that accession's pseudogenome based on the TAIR10 reference genome (downloaded 7 May 2012). We did this using a pipeline of custom Perl scripts and BEDTools (Quinlan *et al.* 2010). For details of the pipeline and code, please see Paper I. After aligning, back-translating, and trimming any misaligned regions using the same methods described for the divergence counts, we used a custom version of PolyMORPHOrama (Bachtrog and Andolfatto 2006; Andalfatto 2007; Haddrill *et al.* 2008) we generated estimates of *Pn* and *Ps* without any minor allele frequency cutoff.

Statistics

To obtain confidence intervals around our estimates, for any given set of N imprinted genes we drew with replacement an N-sized set of loci and estimated α on that set 1,000 times to obtain 95% confidence intervals around the estimate. For a given background gene set, we drew without replacement an N-sized set of loci and estimated α on that set of loci for 1,000 repetitions to obtain a median and 95% confidence intervals.

To test whether a given set of N imprinted genes (which we will call X) differed from another gene set Y, we employed a permutation test. When the comparison gene set Y was larger than N (e.g., background genes), we drew an N-sized set Y_N without replacement from it and pooled those genes with our focal imprinted set to create $X+Y_N$. We shuffled the combined set $X+Y_N$, and divided it in half to create two new sets X' and Y_N', and estimated our metrics for them. We then compared the difference in metric (e.g., α) between the newly estimated $X'–Y_N'$ and compared that to the difference between our estimate for Y_N' and our original estimate for X. Over 1,000 iterations, we counted how often $X'–Y_N' > X–Y_N'$ and calculated our P-value as the this number divided by 1,000.

When the comparison gene set Y was similarly sized or smaller than N (e.g., comparing endosperm imprinted genes vs. embryo imprinted genes), we employed a slightly different version of the permutation test. Instead of generating an N-sized set Y_N' through resampling, we simply combined the N-sized gene set X and the M-sized gene set Y into X_N+Y_M. Then, as before, we shuffled and divided the pooled set X_N+Y_M into an N-sized X_N' and an M-sized Y_M'. We compared the difference in the metric (*e.g.,* α) between $X_N'–Y_M'$ to the original difference in the metric, $X_N–Y_M$. We repeated this 1,000, and calculated our P-value as the number of times $X_N'–Y_M' > X_N–Y_M$ divided by 1,000.

Because we calculate *P* for seven populations for every metric and comparison tested, we also applied Fisher's (1925) Method for Combined Probabilities generate a single *P*-value.

Software

All gene set resampling was done with a custom wrapper I wrote in Perl for the DFE-α program (Eyre-Walker and Keightley 2009). The wrapper also creates all necessary run files for DFE-α and launches the program. Results from DFE-α were concatenated, and *P*-values and data summaries calculated in R v. 3.3.1.

2.4. RESULTS

Imprinted genes have more adaptive substitutions than genomic or seed backgrounds

We used two metrics to assess whether adaptive evolution has occurred between two species: α, the proportion of adaptive substitutions between species, and ω_a, the rate of adaptive substitution fixation. If imprinted genes (N=889) in *Arabidopsis* have evolved under a regime of adaptive evolution, we expect them to have an elevated α and ω_a relative to the rest of the genome. We tested whether values differed with a resampling method that draws at random an equal-sized number of genes from the genomic background. We compared the resampled α or ω_a to the point estimates generated for each population and species to obtain a *P*-value (see Methods). We also estimated α and ω_a using two different species pairs: *A. thaliana* polymorphism with an *A. lyrata* outgroup, and *A. lyrata* polymorphism with an *A. halleri* outgroup. *A. thaliana* has evolved a smaller seed size than its congeners (*A. thaliana:* 0.3-0.5 mm vs *A. lyrata* or *A. halleri*: 0.8-1.2 mm) (Al-Shehbaz and O'Kane 2002). Thus, we include both

comparisons to control for any selection on seed size that could alternatively explain selection on imprinted loci because imprinting affects seed size (Bai and Settles 2015). We use *A. thaliana*-*A. lyrata* as our primary comparison because imprinting status is largely conserved between *A. thaliana* and *A. lyrata* (Klosinka *et al.* 2016).

For all populations of both species, we found α and ω_a to be significantly elevated relative to the genomic background (Fisher's Combined $P \ll 0.001$) (Figure 2.2). We also tested whether imprinted genes have an elevated α or ω_a relative to genes preferentially expressed in the seed because the seed is evolving rapidly compared to other vegetative and floral tissues (Paper I). We again found imprinted loci significantly elevated compared against the seed for all populations (Fisher's Combined $P \ll 0.001$) (Figure 2.2). We found consistent patterns for the rate of adaptive fixation, ω_a, for imprinted loci vs. genomic and seed backgrounds.

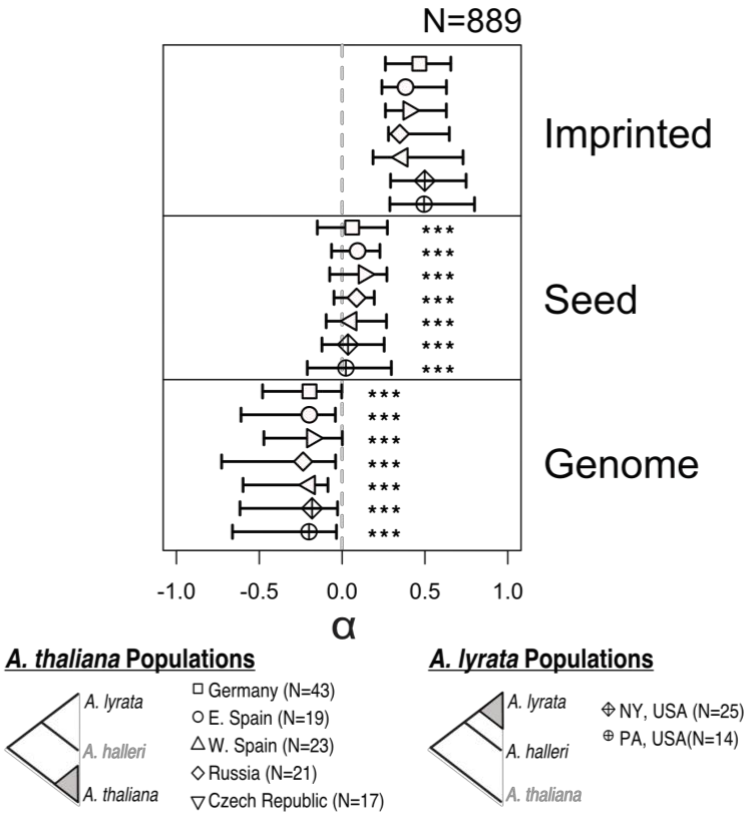


FIGURE 2.2. ELEVATED PROPORTION OF ADAPTIVE SUBSTITUTIONS IN IMPRINTED GENES. Compared to both the seed and genomic backgrounds, α is significantly elevated in imprinted genes (N=889). Shown are 95% confidence intervals about the median. Asterisks indicate level of significance from permutation tests of Imprinted vs. Seed and Imprinted vs. Genome (* $P<0.05$, ** $P<0.01$, *** $P<0.001$). A Test of Combined Probability (Fisher 1925) showed an overall $P < 0.001$ for both Imprinted vs. Seed and Imprinted vs. Background.

Imprinted genes show elevated adaptive evolution regardless of identification method used

Methods to identify imprinted genes range from conventional forward and reverse genetic approaches to modern high-throughput screens. The first generally identifies a single gene whereas the second captures dozens of putative genes in a single study. Here, "confirmed" imprinted genes (N=102) refers to those identified from forward genetic studies, many with known phenotypic effects. "Confirmed" also includes those genes with imprinting status validated after identification with high-throughput methods. "Candidate" imprinted genes

(N=787) are those identified through high-throughput screens and have no known phenotypic effect. Although confirmed imprinted genes are more reliable, our power to detect a signal could be lower. Thus, we also assess candidate imprinted loci to boost power despite some possible error introduced from misidentification of imprinted status. On the whole, if candidate genes are imprinted, we expect a signal similar to that for confirmed genes. Further, if this prediction holds, a gene's signal of selection may useful in screening for future imprinted genes.

We found very similar estimates of α and ω_a between confirmed (Figure 2.3 A) and candidate (Figure 2.3 B) imprinted genes. They were not statistically distinguishable for any population of *A. thaliana* or *A. lyrata* tested. We also tested each of these imprinted gene categories against both the genomic and seed backgrounds (Figure 2.3). Consistent with our above findings for the pooled set of imprinted loci, we find that both confirmed and candidate imprinted genes have significantly elevated α and ω_a relative to both the seed and genomic backgrounds.

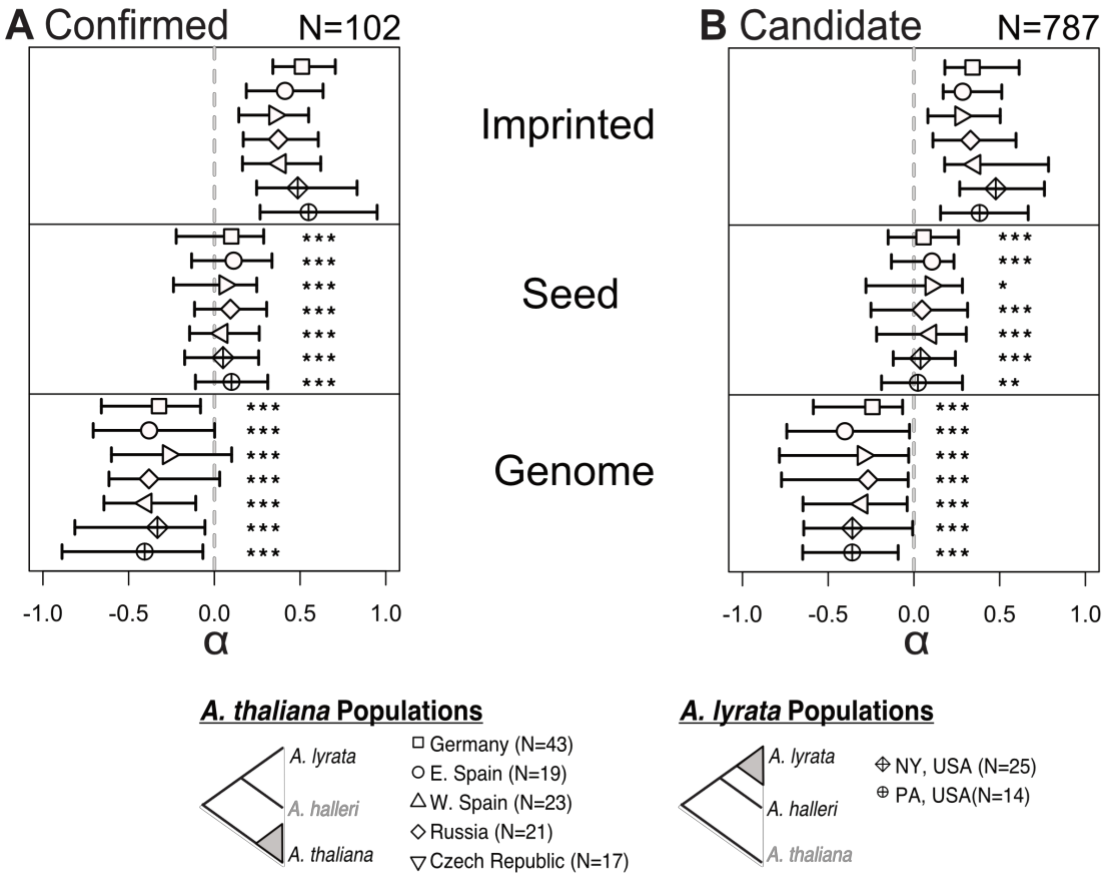


FIGURE 2.3. REGARDLESS OF METHOD OF IMPRINTING IDENTIFICATION, THE PROPORTION OF ADAPTIVE SUBSTITUTIONS IS ELEVATED IN IMPRINTED GENES. (A) *Confirmed Imprinted Genes*. 'Confirmed' imprinted genes (N=102) have been empirically validated, many of which have known phenotypic effects. They show a significantly elevated α relative to both the Seed and Genomic backgrounds (Fisher (1925) Combined P < 0.001) for both Confirmed vs. Seed and Confirmed vs. Background. (B) *Candidate Imprinted Genes*. 'Candidate' imprinted genes (N=787) are those identified through high-throughput screens. They also show a significantly elevated α relative to both the Seed and Genomic backgrounds (Fisher (1925) Combined P < 0.001 for both Confirmed vs. Seed and Confirmed vs. Background. For all gene sets, we show 95% confidence intervals about the median. Asterisks indicate level of significance (permutation test, * P<0.05, ** P<0.01, *** P<0.001).

Schon and Nodine (2017) argued that many of the loci identified as maternally imprinted through high-throughput screens may result from contamination with maternal tissue. To control for any maternal contamination in our candidate gene sets, we also tested the 'stringently filtered' imprinted sets from Schon and Nodine (2017). If we find no difference between the sets from Schon and Nodine (2017) and the much larger number of candidate imprinted genes, it suggests that high-throughput screens are getting it right often enough to address the hypotheses of interest.

We found no difference between the 'stringently' identified imprinted loci of Schon and Nodine (2017) (N=123, Figure 2.4) and the other candidate imprinted loci (N=787, Figure 2.3 B) identified by high-throughput methods. Further, we find α (Figure 2.4) and ω_a similarly elevated between this set of imprinted loci (N=123) and both the seed background (Fisher's Combined $P < 0.001$) and genome background (Fisher's Combined $P < 0.001$). The similar results for confirmed, candidate, and stringently identified candidate genes suggests we may use an elevated α to distinguish between imprinted and non-imprinted loci despite any maternal contamination that may occur in high-throughput studies.

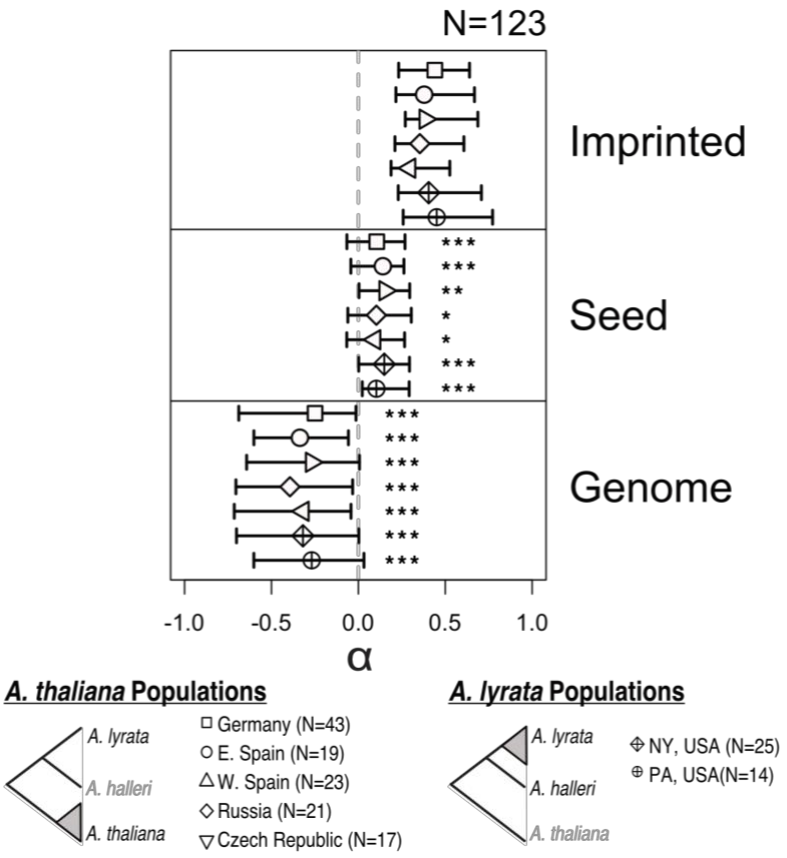


FIGURE 2.4. AFTER CONTROLLING FOR POSSIBLE MATERNAL CONTAMINATION, HIGH-THROUGHPUT IMPRINTED LOCI SHOW AN ELEVATED PROPORTION OF ADAPTIVE SUBSTITUTIONS. We compared the stringently-identified imprinted loci of Schon and Nodine (2017) (N=124) to seed and genomic backgrounds. We found α significantly elevated compared to both backgrounds for all five of the *A. thaliana* populations and both of the *A. lyrata* populations. They show a significantly elevated α relative to both the Seed and Genomic backgrounds (Fisher (1925) Combined P < 0.001) for both Imprinted vs. Seed and Imprinted vs. Background. Shown are 95% confidence intervals about the median. We tested whether Imprinted loci differed from either Seed or Genomic background with permutation tests where asterisks indicate level of significance (* P<0.05, ** P<0.01, *** P<0.001).

Endosperm- and embryo-imprinted genes evolve differently from their respective backgrounds but not each other

We have previously shown a difference in the adaptive evolution between genes expressed in the endosperm and embryo of *Arabidopsis* seeds (Paper I). This suggests the endosperm is more involved in conflict with the mother than the embryo. We therefore tested for a difference in α or ω_a of genes imprinted in the endosperm versus embryo. We subdivided the total set of imprinted genes into those identified as imprinted in the endosperm (N=625) or embryo (N=155), mutually exclusive. We then compared α and ω_a of endosperm- or embryo-imprinted genes to their respective seed compartment backgrounds and to each other. We found a significantly elevated α (Figure 2.5 A, Fisher's Combined $P < 0.001$) and ω_a (Fisher's Combined $P < 0.001$) in endosperm-imprinted genes relative to their endosperm background. For embryo-imprinted genes, we also found a significantly elevated α (Figure 2.5 B, Fisher's Combined $P < 0.001$) and ω_a (Fisher's Combined $P < 0.001$) relative to embryo background genes. With a permutation test, we tested whether the α or ω_a differed between endosperm- and embryo-imprinted genes. We found no difference between them. This suggests that, although the embryo and endosperm evolve with different rates of adaptive evolution (Paper I), imprinted genes in both tissues experience a similar and stronger selection pressure.

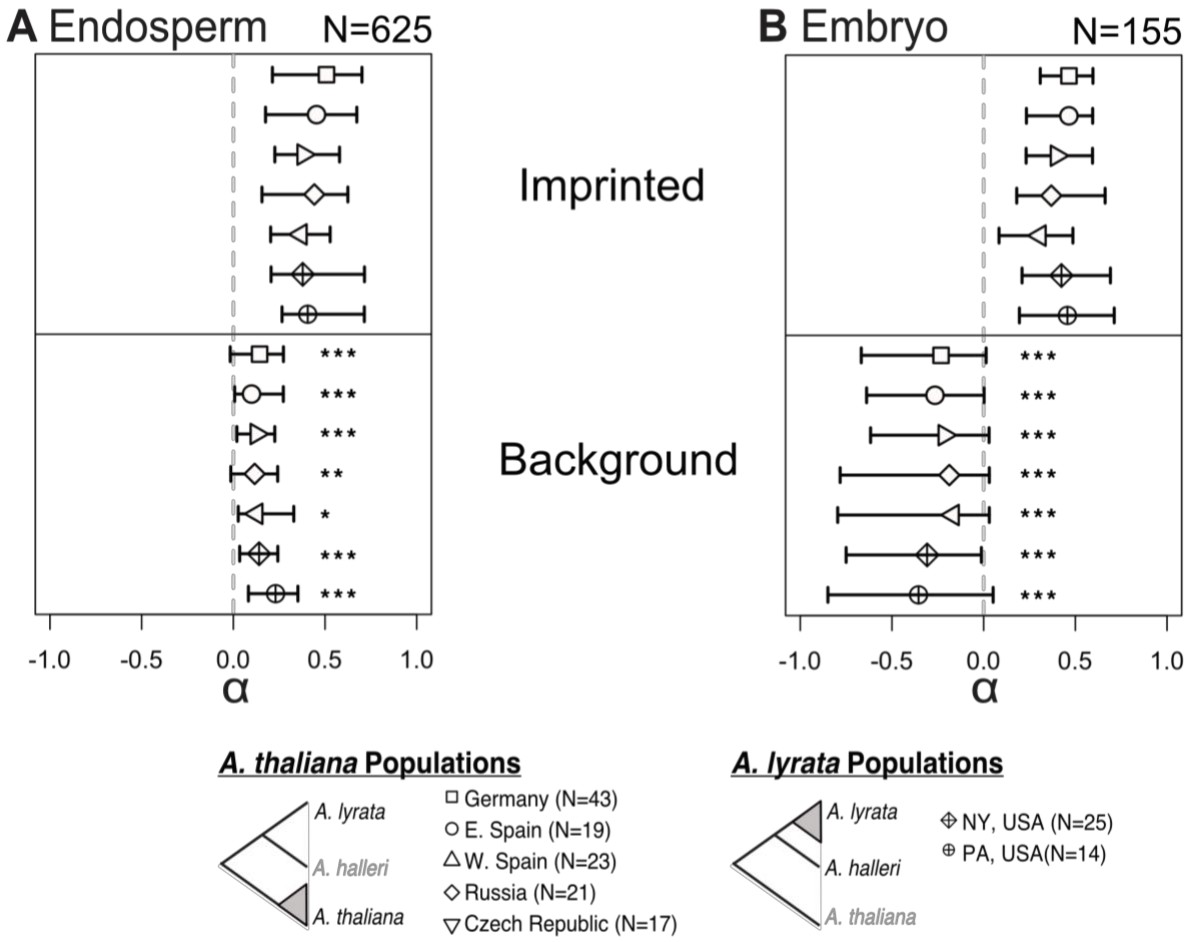


FIGURE 2.5. THE PROPORTION OF ADAPTIVE SUBSTITUTIONS IS SIMILARLY ELEVATED IN ENDOSPERM- AND EMBRYO-IMPRINTED GENES. The rate of adaptive evolution, α, is significantly elevated in imprinted genes of the (A) *Endosperm* (N=625) and (B) *Embryo* (N=155) compared to background sets of genes preferentially expressed in the two tissues. Shown are 95% confidence intervals about the median. Asterisks indicate level of significance from permutation tests of Imprinted vs. their respective backgrounds (* $P<0.05$, ** $P<0.01$, *** $P<0.001$). A Test of Combined Probability (Fisher 1925) showed an overall $P < 0.001$ for both Imprinted vs. Endosperm and Imprinted vs. Embryo. However, there is no difference between Endosperm- and Embryo-Imprinted (Fisher Combined $P > 0.05$).

Patrigenes evolve with more rapidly fixed adaptative substitutions than matrigenes

We also tested whether the proportion and rate of adaptive evolution differed between matrigenes (N=664) and patrigenes (N=179), mutually exclusive. We found patrigenes have a consistently higher α than matrigenes (Figure 2.6). Although this was not the case across all populations of *A. thaliana*, patrigenes had a higher α in both *A. lyrata* populations. The Fisher's Combined $P < 0.001$ suggests that, on the whole, patrigenes are evolving with more adaptive substitutions than matrigenes. We found patrigenes' ω_a similarly elevated. We again used the Schon and Nodine (2017) dataset to control for maternal tissue contamination given that so many of these imprinted genes had been identified by high-throughput assay. We again found an elevated patrigenic α (Figure 2.7, Fisher's Combined $P < 0.001$) and ω_a (Fisher's Combined $P = 0.001$). Both results support the conclusion that patrigenes are evolving with more rapid adaptation than matrigenes.

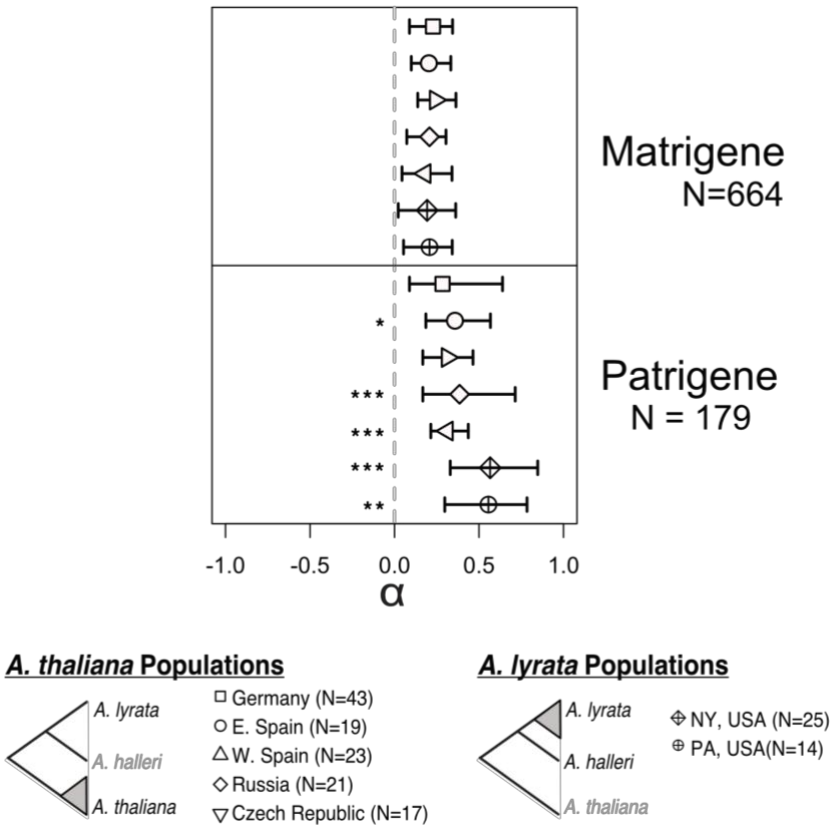


FIGURE 2.6. THE PROPORTION OF ADAPTIVE SUBSTITUTIONS IS ELEVATED IN PATRIGENES OVER MATRIGENES. With permutation tests, we compared α of all matrigenes (N=664) and patrigenes (N=179) identified to date. The point estimate of α is always higher in patrigenes, and it is significantly higher in three populations of *A. thaliana* and both populations of *A. lyrata.* Shown are 95% confidence intervals about the median. Asterisks indicate level of significance (permutation test, * $P<0.05$, ** $P<0.01$, *** $P<0.001$). A Test of Combined Probability (Fisher 1925) showed an overall $P < 0.001$ for higher patrigenic α.

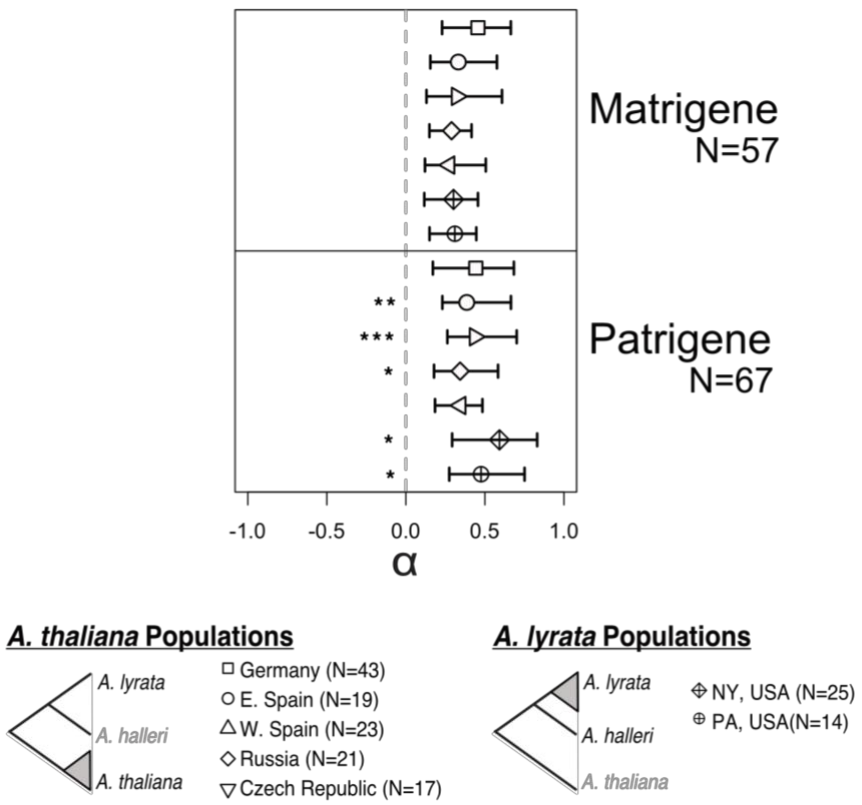


FIGURE 2.7. AFTER CONTROLLING FOR MATERNAL CONTAMINATION, HIGH-THROUGHPUT PATRIGENES SHOW AN ELEVATED PROPORTION OF ADAPTIVE SUBSTITUTIONS. Using the set of high-throughput endosperm-imprinted genes 'stringently' filtered for maternal tissue contamination by Schon and Nodine (2017), we compared the α of matrigenes (N=57) to patrigenes (N=67) with permutation tests. The point estimate of α is always higher in patrigenes, and it is significantly higher in three populations of *A. thaliana* and both populations of *A. lyrata.* Shown are 95% confidence intervals about the median. Asterisks indicate level of significance (permutation test, * $P<0.05$, ** $P<0.01$, *** $P<0.001$). A Test of Combined Probability (Fisher 1925) showed an overall $P < 0.001$ for higher patrigenic α.

We further subdivided matrigenes and patrigenes on the tissue in which they are imprinted, endosperm (matrigenic N=545, patrigenic N=113) or embryo (matrigenic N=138, patrigenic N=43). We compared the endosperm- and embryo-imprinted genes within matrigenes and patrigenes, respectively, to test whether there has been differential selection upon them. We find there has not (Figure 2.8). These results suggest that while patrigenes tend to have higher

rates of adaptive evolution than matrigenes (Figure 2.6), this is not attributable to the seed tissue in which they are expressed.

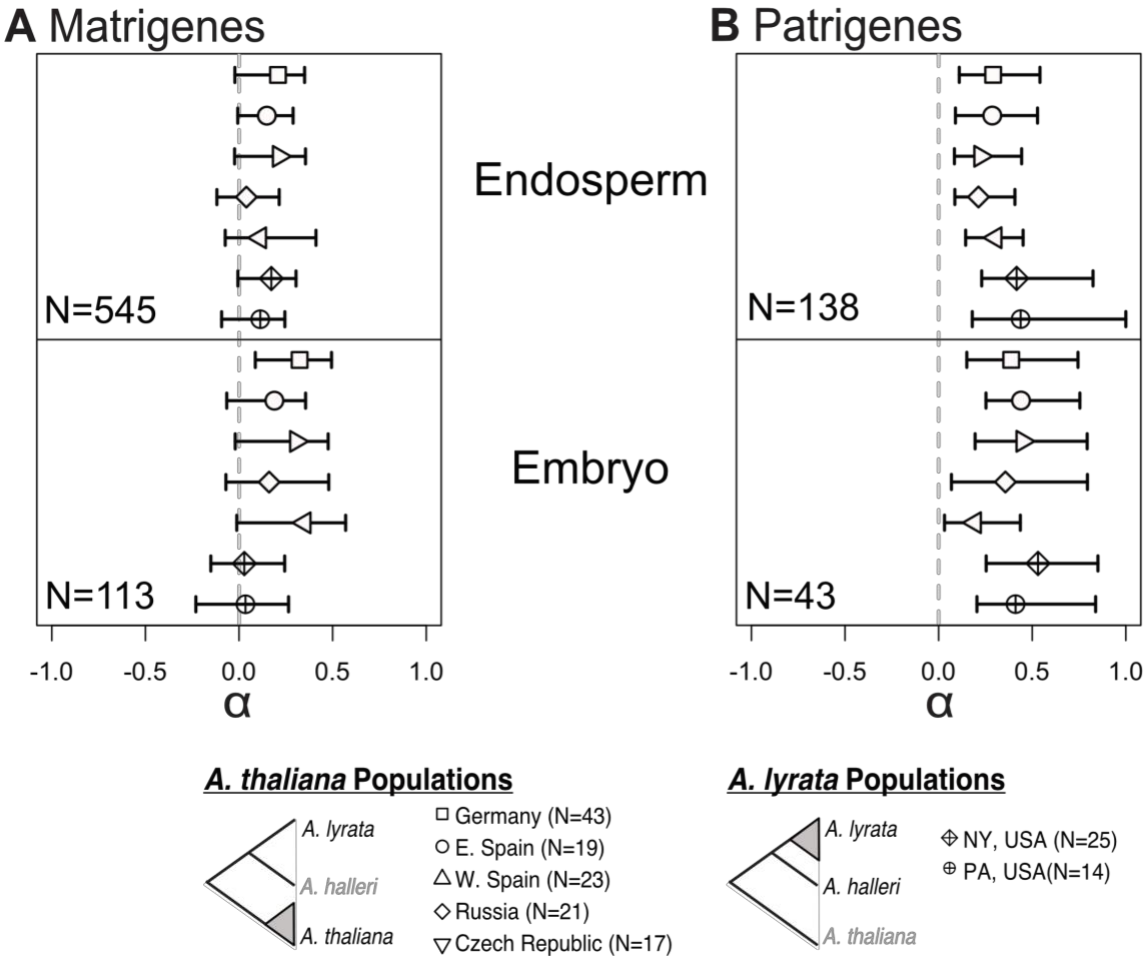


FIGURE 2.8. ENDOSPERM- AND EMBRYO-IMPRINTED GENES SHARE A SIMILAR SELECTION HISTORY WITHIN THEIR RESPECTIVE MATRIGENIC AND PATRIGENIC BACKDROPS. We found no difference between either the (A) matrigenic endosperm- and embryo-imprinted genes nor between the (B) patrigenic endosperm- and embryo-imprinted genes with a Fisher (1925) Combined P > 0.05 in both cases. Shown are 95% confidence intervals about the median. Asterisks indicate level of significance (permutation test, * P<0.05, ** P<0.01, *** P<0.001).

2.5. DISCUSSION

Imprinted genes in *Arabidopsis* show a distinctly elevated proportion (α) and rate (ω_a) of adaptive substitutions relative to other genes in the genome (Figure 2.2). This is consistent with previous findings of positive selection in some imprinted loci (Wolff *et al.* 2011; Tuteja *et al.* 2019), although early evidence in mice found no evidence of positive selection on imprinted genes (McVean and Hurst 1997). However, we go further to show that this elevated α is not solely because the seed itself is under positive selection. We find that despite the elevated α of genes preferentially expressed in the seed (Paper I), α of imprinted genes is higher yet (Figure 2.2).

These results are consistent with the Kinship Theory (Haig 1997; Haig 2000) of imprinting. This theory predicts imprinted genes would have an elevated α because of an evolutionary conflict, and thereby possible arms race, between mothers and fathers over maternal resources. Some of the earliest evidence cited in support of kinship theory included evidence of positive selection at the rodent imprinted *Igf2r* locus in rodents (Smith and Hurst 1998) and the human imprinted *KLF14* locus (Parker-Katiraee *et al.* 2007). Both of these loci are imprinted in the placenta and influence maternal-resource transfer to the embryo.

But why is there more adaptive evolution in imprinted genes than in other seed genes involved in conflict? This is because not all seed genes are involved in conflict, whereas all imprinted genes might be. If Haig's theory is correct, genes become imprinted by parents because they can affect this conflict over resource investment. This would define a narrow set of genes involved in conflict. In contrast, the higher adaptive evolution of genes expressed preferentially in seeds (or in particular parts of seeds) (Paper I) likely reflects an average. That

signal of adaptation presumably averages over conflict genes plus other genes involved in other seed functions, resulting in a lower average rate of adaptative evolution than seen here.

We find evidence of rapid adaptation and conflict for a broader set of imprinted genes than previous studies have found because we include imprinted genes identified through high-throughput screens. Our results are not affected by inclusion of these high-throughput gene sets (Figure 2.3 A vs. 2.3 B). There was no difference in α or ω_a between genes with confirmed imprinting status and those candidates identified through high-throughput screens. This finding held when we used imprinted genes 'stringently' filtered for maternal contamination (Schon and Nodine 2017). Even on this reduced set, we find imprinted genes have higher α and ω_a relative to the genomic and seed backgrounds (Figure 2.4). Because we find no difference in adaptive evolution between this drastically reduced set and the much larger candidate set, metrics like α or ω_a could provide additional evidence for a gene's imprinted status.

We also find that endosperm- and embryo-imprinted genes have a greater signature of rapid adaptation relative to genes enriched in those tissues (Figure 2.5). This shows that that imprinted genes experience selection distinct from that of the endosperm (Paper I). We previously showed that genes enriched in the endosperm have a higher α than those in the embryo. In this study, we found no difference between endosperm- (Figure 2.5 A) and embryo-imprinted genes (Figure 2.5 B) with regard to rapid adaptation. We previously argued that the elevated α of the endosperm reflects its position in the conflict between embryo and mother over resource allocation. The endosperm, serving as intermediary for the embryo, takes over the embryo's role in the conflict with the mother plant (Paper I). However, our results here suggest that imprinted embryo genes are fully involved in conflict.

As summarized in Table 2.2, the only hypothesis fully consistent with the elevated α we find for endosperm- and embryo-imprinted genes is the kinship theory (Haig 1997; Haig 2000, Table 2.1). Although there are numerous hypotheses that have been proposed to explain the origin and maintenance of imprinting, the kinship theory is the only one clearly based on evolutionary conflict of interests (Haig 2014, Spencer and Clark 2014, Table 2.1). These non-conflict hypotheses do not predict arms races between imprinted genes and the mother plant; thus, they do not predict rapid adaptation and a high α or ω_a. Thus, they cannot explain the elevated α and ω_a found in imprinted genes by this study.

There is one proximate hypothesis – increased selection exposure on imprinted genes in heterozygous endosperms – that could explain rapid adaptation in imprinted genes. Full silencing of matrigenes or patrigenes in endosperms would result in functionally haploid alleles in heterozygotes. This would expose recessive alleles to stronger selection, both purifying and positive. There are two possible outcomes for α based on the selection-exposure hypothesis, but neither of them predicts the high α we saw in patrigenes (Figure 2.8). Instead, the selection-exposure hypothesis predicts a high α for matrigenes but not patrigenes because partial silencing could expose matrigenes to stronger selection. In contrast, imprinted matrigenes and patrigenes in the embryo will be equally exposed to selection (because silencing one allele will fully expose the other). This means α should be the same between matrigenes and patrigenes in the embryo.

Instead, we find patrigenes have evidence of more rapid adaptation than matrigenes, which supports the kinship theory over either of the two non-conflict hypotheses (Figure 2.6, Table 2.2). As others have observed in endosperm-imprinted genes *A. thaliana* (Tuteja *et al.* 2019) and *Capsella rubella*, another member of the Brassicaceae (Hatorangan *et al.* 2016), patrigenes show elevated adaptive evolution relative to matrigenes. Patrigenes had a significantly

higher α (Figure 2.6) and ω_a than matrigenes. We found a consistent result when we looked at matrigenes and patrigenes after Schon and Nodine (2017) accounted for possible maternal contamination (Figure 2.7).

	Prediction		
	Higher imprinted α compared to genome, seed, endosperm, embryo	**Patrigene α higher than matrigene**	**Supported?**
Kin Conflict	Yes	Yes	Yes
Non-conflict ultimate theories	No	No	No
Selection-exposure proximate theory	Yes	No	No
Our Result	**Yes**	**Yes**	

Table 2.2. The only hypothesis supported by our results is the kinship hypothesis. The two ultimate hypotheses we tested, the kin conflict (Haig 1997; 2000) and various non-conflict (Table S1) hypotheses tested make differential predictions about arms races and thus α. Because we find an elevated α of imprinted genes relative to the genome, seed, endosperm, and embryo, our results support the kinship theory. We could also differentiate between the kin conflict theory and a proximate (mechanistic) theory, one of selection-exposure on recessive alleles. The selection-exposure hypothesis could explain a high α on imprinted genes but not a higher patrigenic α because it makes the opposite prediction.

Further, the higher patrigenic α and ω_a is not explained by its tissue of enrichment (Figure 2.8 A vs. B). Our results suggest that there is asymmetrical selection on patrigenes vs. matrigenes even though selection acts equally on endosperm- and embryo-imprinted genes (Figure 2.5). This favors the kin conflict theory over the selection-exposure hypothesis because the latter predicted higher α in endosperm matrigenes over embryo matrigenes. The kinship theory, on the other hand, makes no differential prediction on matrigenes or patrigenes in either

tissue, only that patrigenes will be in greater conflict with the other seed parties relative to matrigenes (Figure 2.1). Patrigenes are not related to half-sibling embryos on the same plant but matrigenes and genes of the mother plant are.

What becomes clear from this and previous studies (Wolff *et al.* 2011, Tuteja *et al.* 2019) is that rapid adaptation has acted on imprinted genes in *Arabidopsis* in a way that can only be best explained by the kinship theory (Haig 1997; Haig 2000). Matrigenes and patrigenes are evolving under a greater arms race over maternal resources than the endosperm or embryo in which they are imprinted. Patrigenes experience faster adaptation in most populations of *Arabidopsis* regardless of mating system. Even though *A. thaliana* evolved self-compatibility about one million years ago (Tang *et al.* 2007) and is still predominantly selfing (Bomblies *et al.* 2010), our results may reflect the shared outbred history between *A. thaliana* and *A. lyrata.* This means the imprints would have been made in the outbreeding ancestor of *A. thaliana* and *A. lyrata*. This seems likely because the majority of imprinted genes in the endosperm of *A. lyrata* are orthologous to *A. thaliana* imprinted genes (Klosinska *et al.* 2016). Interestingly, our results show that even with facultative inbreeding in *A. thaliana*, the adaptive signature of imprinting conflict over maternal resources is still evident. Although this may seem contrary to the kinship theory, it is not. The kinship theory predicts that outbreeding increases conflict over maternal resource allocation (Haig 1997; Haig 2013) because it results in a large zone of potential conflict for patrigenes relative to the mother plant (Figure 2.1). Although this zone will get smaller with increased inbreeding, even a low level of outcrossing could be sufficient to generate conflict-driven selection on imprinted genes.

Our results point to the increasing importance of kin conflicts in seed evolution (Paper I) and ultimately seed size. Because of the effects of imprinting on seed size – both in terms of a

proximate and ultimate cause – imprinted loci are likely to have evolved as we have selected for increased seed size and yield. Thus, imprinting evolution could have substantial implications for these loci during domestication: how has domestication impacted these loci as humans have selected not only for larger, more nutritious seeds, but more seed abundance? Fruitful gains will likely come from an improved understanding of how kin conflicts affect maternal resource transfer to the developing seed.

2.6. LITERATURE CITED

Al-Shehbaz, I.A. and O'Kane, S.L. 2002. Taxonomy and phylogeny of *Arabidopsis* (Brassicaceae). *Arabidopsis Book* **1**.

Andolfatto, P. 2007. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res* **17**: 1755–1762.

Bachtrog, D. and Andolfatto, P. 2006. Selection, recombination and demographic history in *Drosophila miranda*. *Genetics* **174**: 2045–2059.

Bai, F. and Settles, A.M. 2015. Imprinting in plants as a mechanism to generate seed phenotypic diversity. *Front. Plant Sci.* **5**.

Barlow, D.P. 1993. Methylation and imprinting: from host defense to gene regulation? *Science* **260**: 309–310.

Bierne, N. and Eyre-Walker, A. 2004. The genomic rate of adaptive amino acid substitution in *Drosophila*. *Mol Biol Evol* **21**: 1350–1360.

Bomblies, K., Yant, L., Laitinen, R.A., Kim, S.-T., Hollister, J.D., Warthmann, N., *et al.* 2010. Local-scale patterns of genetic variability, outcrossing, and spatial structure in natural stands of *Arabidopsis thaliana*. *PLoS Genet* **6**: e1000890.

Briskine, R.V., Paape, T., Shimizu-Inatsugi, R., Nishiyama, T., Akama, S., Sese, J., *et al.* 2017. Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol Ecol Resour* **17**: 1025–1036.

Day, T. 2004. Intralocus Sexual conflict can drive the evolution of genomic imprinting. *Genetics* **167**: 1537–1546.

Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Coan, P.M., Burton, G.J. and Ferguson-Smith, A.C. 2005. Imprinted genes in the placenta – A review. *Placenta* **26**: S10–S20.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Eyre-Walker, A. and Keightley, P.D. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* **26**: 2097–2108.

Fay, J.C. and Wu, C.-I. 2001. The neutral theory in the genomic era. *Current Opinion in Genetics and Development* **11**: 642–646.

Fisher, R.A. 1925. *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Fracassetti, M., Griffin, P.C. and Willi, Y. 2015. Validation of pooled whole-genome re-sequencing in *Arabidopsis lyrata*. *PLOS ONE* **10**: e0140462.

Gehring, M., Missirian, V. and Henikoff, S. 2011. Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* **6**: e23687.

Gossmann, T.I., Song, B.-H., Windsor, A.J., Mitchell-Olds, T., Dixon, C.J., Kapralov, M.V., *et al.* 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant dpecies. *Mol Biol Evol* **27**: 1822–1832.

Haddrill, P.R., Bachtrog, D. and Andolfatto, P. 2008. Positive and negative selection on noncoding DNA in *Drosophila simulans*. *Mol Biol Evol* **25**: 1825–1834.

Haig, D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences* **264**: 1657–1662.

Haig, D. 2000. The kinship theory of genomic imprinting. *Annual Review of Ecology and Systematics* **31**: 9–32.

Haig, D. 2013. Kin conflict in seed development: an interdependent but fractious collective. *Annual Review of Cell and Developmental Biology* **29**: 189–211.

Hamilton, W.D. 1964a. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* **7**: 1–16.

Hamilton, W.D. 1964b. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* **7**: 17–52.

Hatorangan, M.R., Laenen, B., Steige, K.A., Slotte, T. and Köhler, C. 2016. Rapid evolution of genomic imprinting in two species of the Brassicaceae. *The Plant Cell* **28**: 1815–1827.

Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., *et al.* 2011. Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proceedings of the National Academy of Sciences* **108**: 1755–1762.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.-F., Clark, R.M., *et al.* 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* **43**: 476–481.

Immler, S. and Otto, S.P. 2018. The evolutionary consequences of selection at the haploid gametic stage. *Am. Nat.* **192**: 241–249.

Iwasa, Y. and Pomiankowski, A. 1999. Sex specific X chromosome expression caused by genomic imprinting. *J. Theor. Biol.* **197**: 487–495.

Jahnke, S. and Scholten, S. 2009. Epigenetic resetting of a gene imprinted in plant embryos. *Current Biology* **19**: 1677–1681.

Kaneko-Ishino, T., Kohda, T. and Ishino, F. 2003. The regulation and biological significance of genomic imprinting in mammals. *J. Biochem.* **133**: 699–711.

Kawabe, A., Fujimoto, R. and Charlesworth, D. 2007. High diversity due to balancing selection in the promoter region of the *Medea* gene in *Arabidopsis lyrata*. *Current Biology* **17**: 1885–1889.

Keverne, E.B. and Curley, J.P. 2008. Epigenetics, brain evolution and behaviour. *Front Neuroendocrinol* **29**: 398–412.

Klosinska, M., Picard, C.L. and Gehring, M. 2016. Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat Plants* **2**: 16145.

Köhler, C., Wolff, P. and Spillane, C. 2012. Epigenetic mechanisms underlying genomic imprinting in plants. *Annual Review of Plant Biology* **63**: 331–352.

McDonald, J.F. 1999. Genomic imprinting as a coopted evolutionary character. *Trends in Ecology and Evolution* **14**: 359.

McDonald, J.F., Matzke, M.A. and Matzke, A.J. 2005. Host defenses to transposable elements and the evolution of genomic imprinting. *Cytogenet. Genome Res.* **110**: 242–249.

McDonald, J.H. and Kreitman, M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.

McVean, G.T. and Hurst, L.D. 1997. Molecular evolution of imprinted genes: no evidence for antagonistic coevolution. *Proceedings of the Royal Society B: Biological Sciences* **264**: 739–746.

Meisel, R.P. and Connallon, T. 2013. The faster-X effect: integrating theory and data. *Trends Genet* **29**: 537–544.

Miri, K. and Varmuza, S. 2009. Imprinting and extraembryonic tissues – mom takes control. *Int Rev Cell Mol Biol* **276**: 215–262.

Moore, T. and Mills, W. 2008. Evolutionary theories of imprinting – enough already! In: *Genomic Imprinting* (J. F. Wilkins, ed), pp. 116–122. Springer New York.

Nei, M. and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.

Nodine, M.D. and Bartel, D.P. 2012. Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**: 94–97.

Pál, C., Papp, B. and Lercher, M.J. 2006. An integrated view of protein evolution. *Nature Reviews Genetics* **7**: 337.

Parker-Katiraee, L., Carson, A.R., Yamada, T., Arnaud, P., Feil, R., Abu-Amero, S.N., *et al.*

2007. Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLOS Genetics* **3**: e65.

Patten, M.M., Ross, L., Curley, J.P., Queller, D.C., Bonduriansky, R. and Wolf, J.B. 2014. The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**: 119–128.

Pignatta, D., Erdmann, R.M., Scheer, E., Picard, C.L., Bell, G.W. and Gehring, M. 2014. Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife* **3**.

Pires, N.D. and Grossniklaus, U. 2014. Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Rep* **6**.

Prickett, A.R. and Oakey, R.J. 2012. A survey of tissue-specific genomic imprinting in mammals. *Mol Genet Genomics* **287**: 621–630.

Queller, D.C. 1983. Kin selection and conflict in seed maturation. *Journal of Theoretical Biology* **100**: 153–172.

Queller, D.C. 1984. Models of kin selection on seed provisioning. *Heredity* **53**: 151–165.

Queller, D.C. 2003. Theory of genomic imprinting conflict in social insects. *BMC Evolutionary Biology* 23.

Queller, D.C. and Strassmann, J.E. 2018. Evolutionary conflict. *Annual Review of Ecology, Evolution, and Systematics* **49**: 73–93.

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

Raissig, M.T., Bemer, M., Baroux, C. and Grossniklaus, U. 2013. Genomic imprinting in the *Arabidopsis* embryo is partly regulated by PRC2. *PLoS Genetics* **9**: e1003862.

Raunsgard, A., Opedal, Ø.H., Ekrem, R.K., Wright, J., Bolstad, G.H., Armbruster, W.S., *et al.* 2018. Intersexual conflict over seed size is stronger in more outcrossed populations of a mixed-mating plant. *Proceedings of the National Academy of Sciences* **115**: 11561–11566.

Remm, M., Storm, C.E.V. and Sonnhammer, E.L.L. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology* **314**: 1041–1052.

Sapienza, C. 1989. Genome imprinting and dominance modification. *Ann. N. Y. Acad. Sci.* **564**: 24–38.

Schon, M.A. and Nodine, M. 2017. Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome datasets. *The Plant Cell* tpc.00845.2016.

Smith, N.G. and Hurst, L.D. 1998. Molecular evolution of an imprinted gene: repeatability of patterns of evolution within the mammalian insulin-like growth factor type II receptor. *Genetics* **150**: 823–833.

Smith, N.G.C. and Eyre-Walker, A. 2002. Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.

Solter, D. 1988. Differential imprinting and expression of maternal and paternal genomes. *Annu. Rev. Genet.* **22**: 127–146.

Spencer, H.G. and Clark, A.G. 2006. A chip off the old block: a model for the evolution of genomic imprinting via selection for parental similarity. *Genetics* **174**: 931–935.

Spencer, H.G. and Clark, A.G. 2014. Non-conflict theories for the evolution of genomic imprinting. *Heredity* **113**: 112–118.

Spillane, C., Schmid, K.J., Laoueillé-Duprat, S., Pien, S., Escobar-Restrepo, J.-M., Baroux, C., *et al.* 2007. Positive Darwinian selection at the imprinted MEDEA locus in plants. *Nature* **448**: 349–352.

Suyama, M., Torrents, D. and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl. Acids Res.* **34**: W609–W612.

Tang, C., Toomajian, C., Sherman-Broyles, S., Plagnol, V., Guo, Y.-L., Hu, T.T., *et al.* 2007. The evolution of selfing in *Arabidopsis thaliana*. *Science* **317**: 1070–1072.

Tuteja, R., McKeown, P.C., Ryan, P., Morgan, C.C., Donoghue, M.T.A., Downing, T., *et al.* 2019. Paternally expressed imprinted genes under positive Darwinian selection in *Arabidopsis thaliana*. *Mol Biol Evol* **36**: 1239–1253.

Úbeda, F. and Haig, D. 2004. Sex-specific meiotic drive and selection at an imprinted locus. *Genetics* **167**: 2083–2095.

Varmuza, S. and Mann, M. 1994. Genomic imprinting – defusing the ovarian time bomb. *Trends Genet.* **10**: 118–123.

Wilkins, J.F. and Haig, D. 2003. What good is genomic imprinting: the function of parent-specific gene expression. *Nature Reviews Genetics* **4**: 359–368.

Willi, Y. 2013. The battle of the sexes over seed size: support for both kinship genomic imprinting and interlocus contest evolution. *The American Naturalist* **181**: 787–798.

Wolf, J.B. and Hager, R. 2006. A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biology* **4**: e380.

Wolf, J.B. and Hager, R. 2009. Selective abortion and the evolution of genomic imprinting. *Journal of Evolutionary Biology* **22**: 2519–2523.

Wolf, J.B. 2013. Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proceedings of the National Academy of Sciences* **110**: 5085–5090.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., *et al.* 2011. High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genetics* **7**: e1002126.

Yang, Z. and Bielawski, J.P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol. (Amst.)* **15**: 496–503.

Yang, Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

APPENDIX I

Family quarrels in seeds and rapid adaptive evolution in *Arabidopsis*

Katherine S. Geist[a], Joan E. Strassmann[a,b], and David C. Queller[a,b,1]

[a]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; and [b]Wissenschaftskolleg zu Berlin–Institute for Advanced Study, 14193 Berlin, Germany

Edited by Raghavendra Gadagkar, Indian Institute of Science, Bangalore, India, and approved March 29, 2019 (received for review October 15, 2018)

Evolutionary conflict can drive rapid adaptive evolution, sometimes called an arms race, because each party needs to respond continually to the adaptations of the other. Evidence for such arms races can sometimes be seen in morphology, in behavior, or in the genes underlying sexual interactions of host–pathogen interactions, but is rarely predicted a priori. Kin selection theory predicts that conflicts of interest should usually be reduced but not eliminated among genetic relatives, but there is little evidence as to whether conflict within families can drive rapid adaptation. Here we test multiple predictions about how conflict over the amount of resources an offspring receives from its parent would drive rapid molecular evolution in seed tissues of the flowering plant *Arabidopsis*. As predicted, there is more adaptive evolution in genes expressed in *Arabidopsis* seeds than in other specialized organs, more in endosperms and maternal tissues than in embryos, and more in the specific subtissues involved in nutrient transfer. In the absence of credible alternative hypotheses, these results suggest that kin selection and conflict are important in plants, that the conflict includes not just the mother and offspring but also the triploid endosperm, and that, despite the conflict-reducing role of kinship, family members can engage in slow but steady tortoise-like arms races.

kin selection | arms race | molecular evolution | parent–offspring conflict | endosperm

Evolutionary arms races (1, 2) have been documented for strong conflicts between hosts and pathogens (3, 4) and between males and females (5). The mother–offspring relationship is largely amicable, with the mother ensuring the success of her own genes by helping her offspring. However, some conflict is predicted (6, 7), although the conflict is reduced by kinship, so it might be weaker and harder to detect. Mothers are equally related to all their offspring and should help one of them only when the benefit to it exceeds the cost to other offspring. However, each offspring is more related to itself than to its siblings, so it should therefore try to acquire resources in excess of the maternal optimum. There is some evidence that genes expressed in mammalian placentas, which function to provision embryos and are genetically identical to them, evolve rapidly (8). Seeds offer a special opportunity to test within-family conflict theory (9–14). In flowering plants, seeds contain the embryo, a covering of maternal tissue, and the endosperm (Fig. 1). The endosperm does most of the acquisition of resources from the mother and sometimes also stores the resources (15, 16), presumably allowing the embryo to specialize more in developing properly. In most angiosperms, the endosperm is triploid, identical to the embryo but with an extra dose of the maternal alleles, which gives it its own peculiar relatedness patterns (9–14).

Fig. 2 shows how kin selection is predicted to operate on mothers, endosperms, and embryo with respect to transfer of resources to this embryo instead of to other embryos on the same maternal plant (9, 10). There is a large zone of potential conflict where an embryo and its endosperm favor this transfer but the mother does better to provision her other embryos. There is a much smaller zone of conflict where the endosperm is predicted to side with the mother against the embryo. Note also that, if two nonrelatives were selected with respect to providing a benefit to one at a cost to the other, they would be in conflict over the entire positive benefit–cost space in Fig. 2, so relatedness is a moderating factor that reduces conflicts within families.

However, evidence for seed conflict has been indirect. For example, there are possible morphological features consistent with conflict, such as invasive haustoria of endosperms and maternal integumental barriers (9, 11). There is also evidence for paternal effects on seed size (17). Here we present strong tests of the prediction that conflict among these tissues in *Arabidopsis* will lead to high rates of adaptive evolution.

Conflict-based arms races are most likely to be found in genes that are specialized for in tissues engaged in conflict. We therefore first identified sets of *Arabidopsis* genes specialized for a focal organ or tissue as those genes that show significantly higher expression in that organ or tissue compared with other organs or tissues, using published microarray expression datasets (18, 19). We then compared rates of adaptive evolution (α) of these gene sets using $\alpha = 1 - \left(\frac{Pn}{Ps}/\frac{Dn}{Ds}\right)$, where Dn/Ds is the ratio of nonsynonymous to synonymous substitutions between species and Pn/Ps is the corresponding ratio for polymorphisms within species (20–22). This statistic is based on the logic of the McDonald–Kreitman test (23): The two ratios should be the same under a combination of neutrality and purifying selection, but positive selection will elevate Dn/Ds, and therefore α; α provides a

Significance

Evolutionary conflict, such as between pathogens and hosts, can lead to arms races in which each party evolves rapidly in response to the harm inflicted by the other. Kin selection makes relatives much more cooperative, but some conflict is usually still expected. We show that even this reduced conflict appears to drive arms races in seeds of the plant *Arabidopsis*, which contain three genetic relatives: maternal tissue, the embryo, and the triploid endosperm. As expected from potential conflict over how much nutrition the embryo should receive from the mother, genes expressed in seed tissues evolve rapidly, particularly the parts directly involved in nutrient transfer. Moreover, the endosperm appears to have largely taken over the embryo's role in this parent–offspring conflict.

Author contributions: K.S.G., J.E.S., and D.C.Q. designed research; K.S.G. performed research; K.S.G. analyzed data; and K.S.G., J.E.S., and D.C.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: Data and authored programs, including gene lists for each category, a measure of each gene's degree of adaptive evolution, statistics computed for each gene set, and the authored Perl wrapper that performs bootstrapping and permutation tests on gene sets using DFE-α, are available at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds.

[1]To whom correspondence should be addressed. Email: queller@wustl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1817733116/-/DCSupplemental.

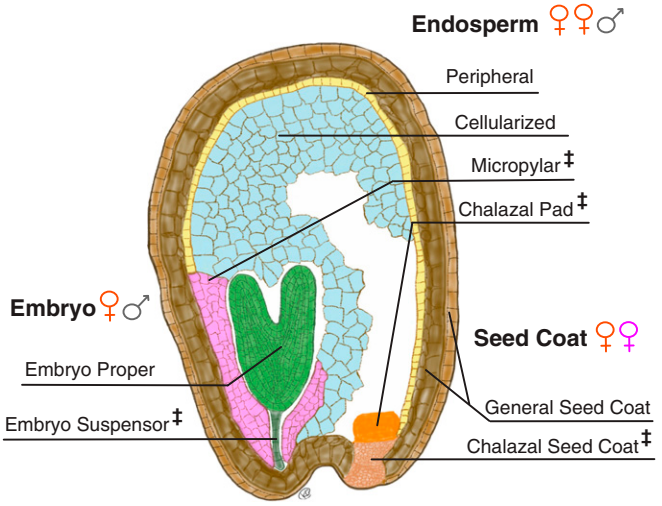


Fig. 1. Components of the angiosperm seed. The developing angiosperm seed, here *Arabidopsis*, has three distinct genetic parties. The mother contributes the seed coat, with the chalazal region being a portal for nutrients. The offspring consists of the embryo proper and a temporary suspensor that may be involved in nutrient acquisition. The endosperm results from a second fertilization and is triploid, genetically identical to the embryo except for an extra set of the maternally derived chromosomes. Nutrients flow to the endosperm through its chalazal region. ‡, Subtissues most involved in transfer between genetically distinct parties.

powerful summary of adaptive evolution averaged over a gene set, and, when two gene sets share the same population history, a higher value of α indicates greater adaptive evolution.

We estimated α in two species pairs, (*i*) *Arabidopsis thaliana* populations with an *Arabidopsis lyrata* outgroup and (*ii*) *A. lyrata* populations with an *Arabidopsis halleri* outgroup (Fig. 3). The first pairing is natural because the expression data come from *A. thaliana*, but there are two complications which we can remove with the second test. First, *A. thaliana* seeds are smaller than those of *A. lyrata* [0.3 mm to 0.5 mm vs. 0.8 mm to 1.2 mm (24)], meaning any excess adaptation observed in seeds could result not just from conflict but from any factor selecting for seed size. However, this is not an issue for *A. lyrata* and *A. halleri*, which have very similar seed sizes [both 0.8 mm to 1.2 mm (24)]. Second, *A. thaliana* is inbred, which changes the relatedness patterns and should reduce conflict (9). This should not be a serious problem, because *A. thaliana* has been outbred for more than 90% of the time since it diverged from *A. lyrata* (25), but the *lyrata–halleri* pairing provides a check with an outbred pair. There are polymorphism data from multiple populations, from which we selected five *A. thaliana* populations and two *A. lyrata* populations for analysis, allowing seven partially independent tests of each prediction (independent polymorphism data but shared outgroup for divergence).

Results

We test conflict predictions at three levels using genes preferentially expressed in organs, in seed tissues, and in seed subtissues (*SI Appendix*, Table S1). First, if there is sufficient family conflict in seeds, then genes with specialized expression in seeds should show more adaptive evolution (higher α) than genes specialized for other organs that are genetically uniform and therefore not subject to conflict. This prediction is successful, with seed genes showing higher α than genes in floral buds, leaf rosettes, stems, and roots in both species pairs; 25 of the 28 comparisons are statistically significant (Fig. 3; all adaptive evolution tests in this paper are permutation tests; see *Methods*).

Second, because the endosperm has taken over the primary nutrient acquisition role for the embryo, we test whether the primary conflict is now between mother and endosperm rather than mother and embryo. This shift would make sense (9) because the endosperm is far less constrained than the embryo, which has to develop in a precise way, and the two tissues differ little in their interests [Fig. 2; even this difference disappears for endosperm genes that are strictly dominant or recessive (10)]. Again, the data support the prediction. Compared with genes with specialized expression in the embryo, we see elevated rates of adaptive evolution in genes specialized for maternal seed coat (four of seven comparisons are significant) and especially for the endosperm (seven of seven comparisons are significant) (Fig. 4, *Left*).

Third, even more informative predictions can be tested within each tissue—maternal, endosperm, and embryo. Genes with specialized expression in subtissues that are most involved in nutrient transfers are predicted to be more engaged in conflict and to have higher α than other subtissues. Most of the actual conflict between mother and endosperm should occur in their chalazal regions where more nutrients are transferred (15, 16) so the conflict hypothesis predicts that these will evolve more rapidly than other subtissues in their respective tissues. This prediction is confirmed both for maternal chalazal seed coat versus the maternal general seed coat (six of seven tests; Fig. 4, *Right*) and for endosperm chalazal pad versus cellularized/peripheral endosperm (seven of seven tests, Fig. 4, *Right*). Two other subtissues are predicted to evolve rapidly only if the embryo still participates in some conflict: the embryo suspensor that is terminally differentiated and participates in nutrient transfer (26) and the micropylar endosperm that surrounds the embryo (15, 16). These predictions are also confirmed, though less strongly, with somewhat higher adaptive evolution in genes with specialized expression in the embryo suspensor versus those in the embryo proper (four of seven tests) and in genes in the micropylar endosperm versus those in the cellularized/peripheral endosperm (five of seven tests) (Fig. 4, *Right*).

Gene ontology (GO) analyses show that extracellular and intracellular communication genes figure prominently, as one might predict under conflict, but a number of other categories are also significantly enriched (*SI Appendix*, Tables S2–S6).

Discussion

The evidence strongly supports multiple predictions of greater adaptive evolution expected from conflict in seeds. Alternative hypotheses cannot account for all of the results. First, the higher α in seeds could reflect an arms race between the maternal seed coat on the outside of the seed against evolving seed predators or soil pathogens and fungi. However, this is not supported by the similarly high α in the endosperm and especially not by the lower α in the general seed coat surrounding the seed versus the chalazal seed coat supplying the nutrients (Fig. 4, *Right*). Second, imprinted genes, which affect seed nutrition, might add still another dimension of kin-selected conflict, that between mother and father (27). We are conducting a separate analysis of imprinted genes, but they constitute small fractions of our gene sets and cannot explain all of the patterns observed. Endosperm tissues show a pattern opposite to an imprinting arms race: Genes preferentially expressed in the slowly evolving cellularized/peripheral region are more often imprinted (8.1%) than those expressed in the rapidly evolving micropylar (0.5%) or chalazal (2.2%) regions [based on 124 stringent, imprinted genes (28)]. Moreover, in maternal tissues, there should be no imprinting conflict, so this hypothesis cannot explain either the high α in the seed coat genes or the higher α in genes in the chalazal region of the seed coat. Finally, the *lyrata–halleri* comparison removes a seed-size selection explanation. However, conflict is always over something—here provisioning—that might be selected for nonconflict reasons, so our results are also consistent with a post hoc hypothesis of nonconflict selection on provisioning. However, some such post hoc explanation could be posited for any

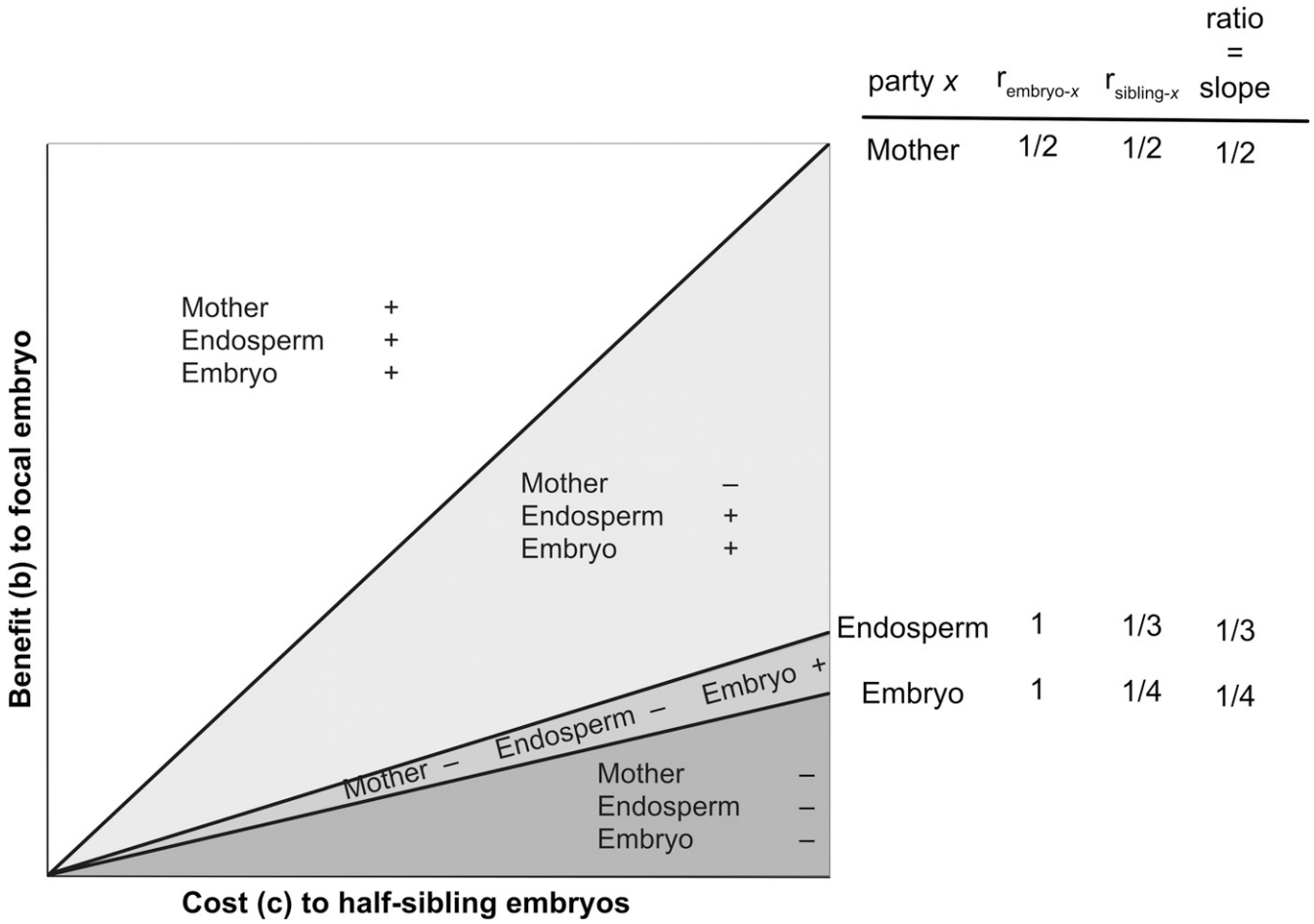


Fig. 2. Conflict in seeds. Suppose an allele is expressed in tissue *x* (*x* = embryo, endosperm, mother) of a focal seed. It increases nutrient flow to the embryo of this seed, increasing its fitness by *b* (*y* axis) and decreasing the total fitness of its current or future maternal half siblings by *c* (*x* axis). Hamilton's kin selection rule (29) states that this allele will be favored when $r_{embryo\text{-}x}b - r_{sibling\text{-}x}c > 0$, where the two *r*s are the relatednesses of tissue *x* to the focal embryo and to half-sibling embryos (9). Therefore, each party should favor this transfer (+) when $b > c^{*}r_{sibling\text{-}x}/r_{embryo\text{-}x}$, and disfavor it (–) when this inequality is reversed. If *b*/*c* is high enough (white), all parties favor the focal embryo, and, if it is low enough, none do (dark gray). In between, there are zones of potential conflict where some tissues would gain from the transfer and others would lose from it.

of the over 600 possible patterns of significance rankings (*SI Appendix*, Table S7) of the five plant organs tested (Fig. 3), and more if we included the patterns of Fig. 4. In contrast, the conflict hypothesis predicted a priori the single pattern actually observed (seeds show more adaptive evolution than the other four), passing a severe attempt at falsification.

In the absence of viable alternative explanations, these results suggest a number of implications. They add support to Hamilton's (29) assertion that kin selection is important far beyond its canonical applications to the evolution of altruism in animals such as social insects, in this case, to plants (8–14, 30). They also add support to the idea that kin selection is relevant not just to driving altruism but also to limiting selfish behavior and conflicts. Our results—together with parallel ones on rodents (8)—also provide support for parent–offspring conflict in general and against the idea (31) that parents should completely win because they have initial control of the contested resources. If that were true, there would be no ongoing conflict and elevated rates of adaptation. Our results suggesting pronounced endosperm conflict with the mother and weaker conflict with the embryo provide support for the idea that the peculiar triploid endosperm, which never lives independently or reproduces directly, evolves according to its own relatedness-based interests. They therefore lend credence to kin-selection theories of the origin and evolution of the endosperm (9–13, 27, 32). Finally, the idea that some parts of the seed have evolved to increase seed size and others have evolved to moderate seed size is likely relevant to strategies for artificially selecting seeds, such as in the cereals that constitute a large part of the human diet.

Relatedness is expected to decrease conflict, so it is interesting that kin interactions nevertheless seem to drive rapid evolution, consistent with an evolutionary arms race. One reason may be the constancy of the conflict. In Aesop's fable, a tortoise raced against a hare, but arms races can pair two hares or two tortoises. Hosts and pathogens may be hares, with selection that is strong but irregular because not every host encounters a pathogen and also because host–pathogen species pairings shift (2). In contrast, family quarrels may resemble races among tortoises. The pace may be slower, but it never wanes, because every offspring has a mother and every evolutionary successful mother has offspring.

Methods

Genes Specialized for Particular Organs and Tissues. From two published *A. thaliana* microarray expression datasets, we identified genes specialized for seed tissues (19) (series GSE12404; *SI Appendix*, Table S1), as well as for seeds and the following nonseed organs: floral bud, leaf rosette, stem, and root (18) (series GSE680; *SI Appendix*, Table S1). The seed expression dataset (19) (series GSE12404) includes time series Affymetrix ATH1 microarray data across six developmental stages from microdissected seed tissues and from

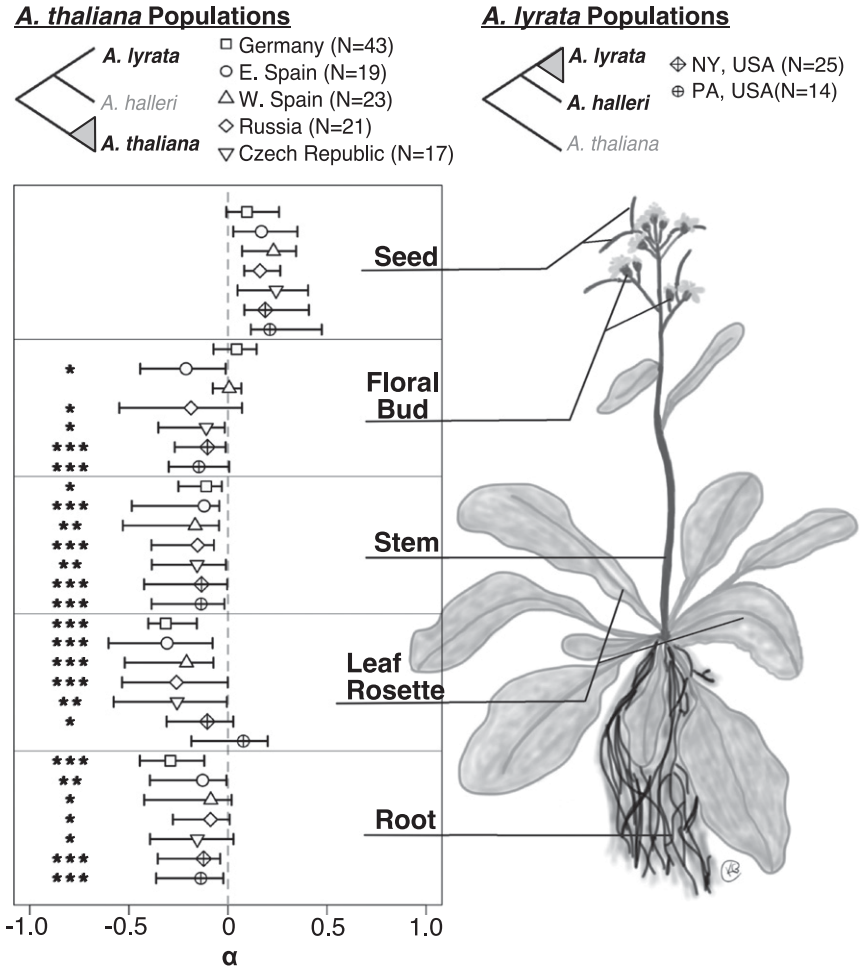


Fig. 3. Adaptive evolution α is higher for genes upregulated in seeds compared with other organs. Each panel shows seven estimates of the rate of adaptive evolution, α, for five *A. thaliana* populations with an *A. lyrata* outgroup, as well as two *A. lyrata* populations with an *A. halleri* outgroup. For floral buds, stems, leaf rosettes, and roots, asterisks show significant differences from the corresponding seed (permutation tests; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

two to three biological replicates. The life stages expression dataset [GSE680 (18)] includes time series ATH1 microarray data from seeds across the same six developmental stages, as well as mature plant organs at single time points. For each microarray experiment, we extracted normalized $\log_2$-transformed expression values for each mRNA sequence with the Robust Multiarray Average preprocessing approach (33) using the Affy package (34) in Bioconductor v.2.12 (35) implemented in R v.2.15 (36).

We used the limma package (37) to identify which RNA sequences were significantly enriched in each organ or tissue that we used in subsequent analyses. We performed pairwise contrasts of a focal tissue against other tissue(s), as specified in the next three paragraphs, with *t* tests moderated by an empirical Bayes function, because there were few replicates available for each microarray experiment. The mRNA transcripts with a Benjamini–Hochberg (38) adjusted *P* value less than 0.01 were considered enriched for a focal tissue. We assigned gene identities to mRNA transcripts using a reannotated array based on the TAIR10 *A. thaliana* genome release (https://www.arabidopsis.org). Transcripts that mapped to more than one sequence were excluded from further analyses. When multiple transcripts mapped to the same gene, we required all of them to be significantly enriched in the focal tissue.

We applied these procedures to perform contrasts to identify gene sets specialized for particular organs or tissues at three different levels. First, to identify genes specialized for particular organs, we performed pairwise contrasts to identify genes encoding mRNAs enriched in each of the following organs: seed, floral bud, leaf rosette, stem, and root [data from series GSE680 (18)]. For the seed genes in this analysis, expression data were averaged across ontogeny.

Second, to identify genes specialized for particular seed tissues, we identified genes significantly enriched in mRNA expression, in at least one developmental stage, in each of the three seed tissues (maternal seed coat, endosperm, and embryo) relative to the other two tissues [data from series GSE12404 (19)]. From these gene sets, we deleted any genes previously found to have significantly enriched expression in floral bud, leaf, stem, or root organs to limit any effects of selection on genes during those stages of the plant life cycle.

Third, to identify genes specialized for particular subtissues, we identified genes with significantly enriched mRNA expression in each of the seed subtissues within the seed coat, embryo, and endosperm tissues. We compared subtissues within a given tissue only. For example, genes specialized for chalazal endosperm were those with significantly enriched expression in the chalazal region of the endosperm, relative to the other three endosperm regions. Again, genes were chosen when their expression was enriched in at least one developmental stage of the focal subtissue but not in any other subtissue. We again deleted any genes in the set previously found to have enriched expression in floral bud, leaf, stem, or root organs. Because of small sample sizes, limited presence in stages, shared ontogenetic origins (16), and shared predictions, we combined the gene sets for "cellularized" and "peripheral" endosperms.

Tests for Molecular Signatures of Positive Selection in Tissue-Specific Genes. We used both interspecific and intraspecific sequence comparisons to test for positive selection in plant organs, seed tissues, and subtissues. For the gene sets specialized in each, we estimated the proportion of adaptive substitutions as $\alpha = 1 - \left(\frac{Pn}{Ps}/\frac{Dn}{Ds}\right)$, where *Pn* and *Ps* are the numbers of nonsynonymous and synonymous polymorphisms within species, respectively, and *Dn* and *Ds* are the numbers of nonsynonymous and synonymous differences between species (21, 39). The metric α is an extension of the McDonald–Kreitman test (23), which assumes that, if an adaptive mutation arises, it is swept to fixation quickly, contributing to between-species divergence but not within-species polymorphism. Thus, if $Dn/Ds > Pn/Ps$, the sequence is thought to be under strong positive selection. The McDonald–Kreitman test typically looks at a single gene, limiting sample size and power, but α is calculated cumulatively across a class of genes. Thus, no single gene in the set need be significant under the McDonald–Kreitman criteria to detect adaptation. We first estimated α with polymorphism counts from *A. thaliana* and with divergences

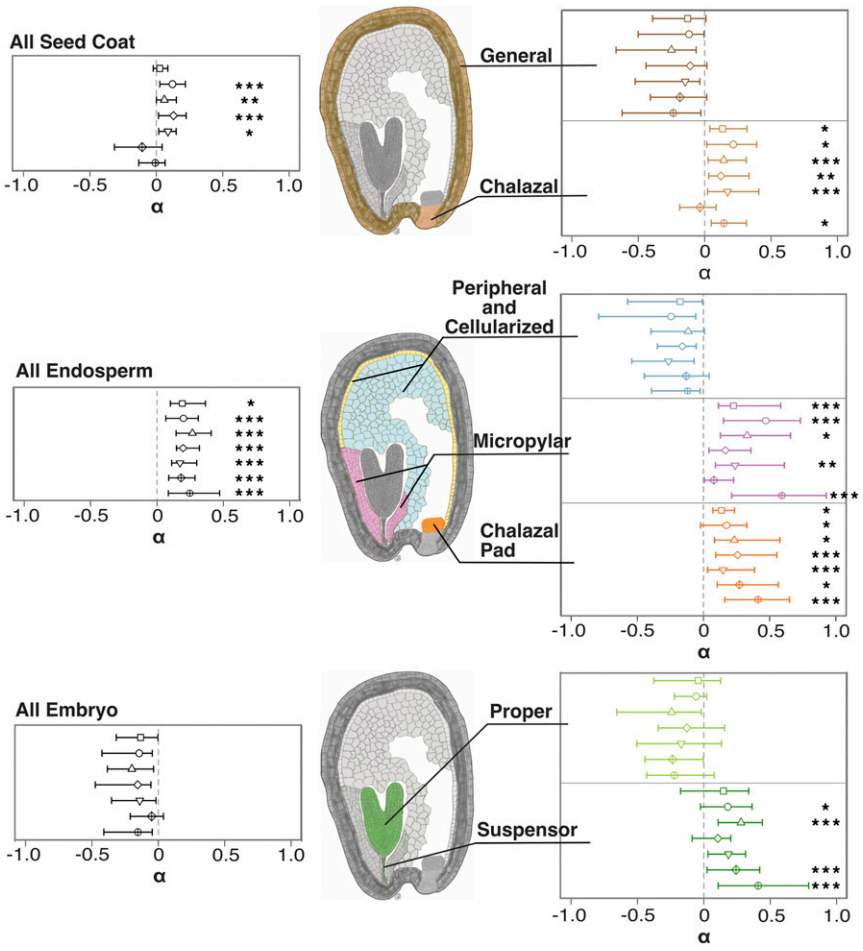


Fig. 4. Adaptive evolution α in genes specialized for different seed tissues and subtissues. (*Left*) Adaptive evolution in the genes specialized for maternal seed coat and endosperm is higher than in those specialized for embryo, supporting the hypothesis that most of the conflict is between mother and endosperm. Asterisks indicate significance relative to the same-population embryo α. (*Right*) Within each of the maternal, endosperm, and embryo tissues, adaptive evolution is higher in genes specialized for subtissues more directly involved in nutrient transfers. Asterisks indicate significant differences relative to same-population α of the subtissue(s) less involved in nutrient transfer (maternal general seed coat, peripheral and cellularized endosperm, embryo proper). The seven populations in each panel are as in Fig. 3 (permutation tests; $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$).

from its sister species, *A. lyrata*. We also estimated α for a pair of two closely related outbred species, *A. lyrata* and *A. halleri*, for which we obtained publicly available polymorphism data for *A. lyrata* (40).

Pairwise divergence estimates. To estimate divergence between *A. thaliana* and *A. lyrata*, or between *A. lyrata* and *A. halleri*, we first identified whole-genome orthologs using our version of standalone InParanoid v.4 (41), which we updated to work with BLAST+ (42). This gave us reciprocal BLAST comparisons of *A. thaliana*, *A. lyrata* (*v*. 1.0) (43), and *A. halleri* (*v*. 1.0) (44) protein sequences. With the 1:1 orthologs, we performed pairwise global alignments of *A. thaliana*, *A. lyrata*, and *A. halleri* proteins using MUSCLE (45), which were again back-translated and trimmed using PAL2NAL v. 14 (46) and trimAl v. 1.2 (47), respectively. We then used the Nei and Gojobori method (48) implemented in the codeml package of PAML 4.0 (runmode = −2, CodonFreq = 2) (49) to estimate the numbers of nonsynonymous and synonymous sites and substitutions per gene between each species pair. We excluded all genes with saturated divergence ($dS \geq 1$; *A. thaliana*–*A. lyrata*, 61 genes; *A. lyrata*–*A. halleri*, 37 genes) from future tests.

***Polymorphism estimates of* A. thaliana *and* A. lyrata.** To reduce the chance of unusual results owing to an unusual recent population history, we estimated within-species polymorphism for five populations of *A. thaliana* and two populations of *A. lyrata*. To obtain *Pn* and *Ps* for *A. thaliana*, we used SNP data from resequenced *A. thaliana* genomes as part of the 1,001 Genomes Project (www.1001genomes.org, *SI Appendix*, Table S2). We chose five populations, each consisting of geographically clustered accessions, to minimize any effects of population structure: Germany ($n = 43$), Czech ($n = 17$), Russia ($n = 21$), E. Spain ($n = 18$), and W. Spain ($n = 23$). We converted the variant call format (VCF) files for each *A. thaliana* individual in each population into a variant FASTA sequence file of the *A. thaliana* reference genome (TAIR10, downloaded May 7, 2012) with a custom Perl script. We used BEDTools (50) to extract the coding sequences for each gene and translated these to amino acids with a custom Perl script. We aligned, back-translated, and trimmed misaligned regions using the same methods described for estimating divergence. We analyzed the coding alignments with PolyMORPHOrama without a minor allele frequency cutoff.

For *A. lyrata*, we obtained pooled resequenced genome data and genotype-by-sequencing (GBS) data for the two available populations, one collected from Erie, PA ($n = 14$), and the other from Jamesville, NY ($n = 25$) (40). To obtain our polymorphism counts for the two available populations of *A. lyrata*, we began with all GBS and pooled-sequencing FASTQ files (European Nucleotide Archive: https://www.ebi.ac.uk/ena, accession PRJEB8335). These FASTQ were deinterleaved and had been demultiplexed (GBS) and trimmed to a minimum PHRED quality score of 20 before they were added to the repository. We merged the FASTQ sequence quality files from multiple lanes of pooled sequence for the New York population to increase coverage. We then mapped all paired end reads to version 1.0 of the *A. lyrata* reference genome (43) with *sampe* in the Burrows–Wheeler Aligner v. 0.7.15 (51). We sorted and indexed the alignment files with SAMTOOLS v. 1.3 (51–53), and then realigned insertions/deletions with the Genome Analysis ToolKit v.3.3.0 (54), removed low-quality reads (<20) and those that failed to map with SAMTOOLS, and removed duplicate reads with Picard (v. 1.128; broadinstitute.github.io/picard/). Because we were only interested in polymorphism in coding sequence, we called variants using the primary coding sequence of *A. lyrata* with Genome Analysis Toolkit HaplotypeCaller (54) with a quality score of >25 and filtered for SNPs only. To ensure optimal coverage across all coding sequence genome-wide, we merged all resulting VCF files by population with the vcf-merge tool of VCFTOOLS v. 1.14 (55). Our same custom Perl script converted the merged VCF files to a variant FASTA, which we then used to extract the coding sequences for each gene and calculate nonsynonymous and synonymous polymorphism counts as described for the *A. thaliana* populations.

***The proportion of adaptive substitutions as estimated by* α.** We estimated the proportion (α) and rate (ω_a) of adaptive substitutions and the proportion of nonsynonymous mutations that are deleterious ($1 - f$) for our different gene sets using the standalone version of the Distribution of Fitness Effects (DFE)-α program v. 2.15 (22). Using a custom Perl wrapper that sums and formats the nonsynonymous and synonymous polymorphism frequency spectra provided

by PolyMORPHOrama creates all necessary run files, incorporates divergence information, and performs either permutations or bootstrapping. In DFE-α, we used default parameters, except that we used a two-epoch model without folded site-frequency spectra and a Jukes–Cantor correction when calculating nucleotide divergence. Results are shown in *SI Appendix*, Table S1.

Statistics. For each focal gene set, we generated confidence intervals for α, ω_a, and $1 - f$ by bootstrapping across loci 1,000 times. For each parameter X, i genes in the focal set were randomly drawn with replacement 1,000 times, from which we reran DFE-α for each resample, recomputed X. These data were used for a 95% confidence interval.

To ask whether focal gene sets differed from each other, we employed permutations to test for differences between two samples. To test for a difference between the statistics of two gene sets, X_1 and X_2 with i and j numbers of genes, we randomly drew without replacement i and j genes from combined sets of genes, reran DFE-α, and recalculated $X_1 - X_2$. We calculated the P value as the proportion of times the permuted difference was greater than zero in the direction predicted. All P values reported are one-tailed.

GO. We performed a GO analysis to rudimentarily examine the functions of the genes in our focal tissues. We used the DAVID Functional Annotation Clustering Tool (56, 57) to obtain clusters of significantly enriched GO terms with a stringency setting of "High" for all gene sets, using the *A. thaliana* genome as background. Clusters for each gene set are given in *SI Appendix*, Tables S2–S6.

Data and Code Availability. Data and authored programs have been archived at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds (58). We include gene lists for each category along with a measure of each gene's degree of adaptive evolution, as well as summary statistics computed for each gene set. We also provide the Perl wrapper we authored that performs bootstrapping and permutation tests on gene sets using DFE-α (21).

ACKNOWLEDGMENTS. We thank Justin Fay, Kenneth Olsen, and Jonathan Myers for comments on this project. This material is based upon work supported by the National Science Foundation under Grants IOS-1656756, IOS-1753743, and DGE-1143954, and by John Templeton Foundation Grant 43667.

1. Dawkins R, Krebs JR (1979) Arms races between and within species. *Proc R Soc Lond B Biol Sci* 205:489–511.
2. Brockhurst MA, et al. (2014) Running with the red queen: The role of biotic conflicts in evolution. *Proc Biol Sci* 281:20141382.
3. Paterson S, et al. (2010) Antagonistic coevolution accelerates molecular evolution. *Nature* 464:275–278.
4. Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* 292:2281–2285.
5. Gossmann TI, Schmid MW, Grossniklaus U, Schmid KJ (2014) Selection-driven evolution of sex-biased genes is consistent with sexual selection in *Arabidopsis thaliana*. *Mol Biol Evol* 31:574–583.
6. Trivers RL (1974) Parent-offspring conflict. *Am Zool* 14:249–264.
7. Kilner RM, Hinde CA (2012) Parent-offspring conflict. *The Evolution of Parental Care*, eds Royle NJ, Smiseth PT, Kölliker M (Oxford Univ Press, Oxford), pp 119–132.
8. Chuong EB, Tong W, Hoekstra HE (2010) Maternal-fetal conflict: Rapidly evolving proteins in the rodent placenta. *Mol Biol Evol* 27:1221–1225.
9. Queller DC (1983) Kin selection and conflict in seed maturation. *J Theor Biol* 100: 153–172.
10. Queller D (1984) Models of kin selection on seed provisioning. *Heredity* 53:151–165.
11. Westoby M, Rice B (1982) Evolution of the seed plants and inclusive fitness of plant tissues. *Evolution* 36:713–724.
12. Willson MF, Burley N (1983) *Mate Choice in Plants* (Princeton Univ Press, Princeton).
13. Charnov EL (1979) Simultaneous hermaphroditism and sexual selection. *Proc Natl Acad Sci USA* 76:2480–2484.
14. Shaanker RU, Ganeshaiah KN, Bawa KS (1988) Parent-offspring conflict, sibling rivalry, and brood size patterns in plants. *Annu Rev Ecol Syst* 19:177–205.
15. Costa LM, Gutièrrez-Marcos JF, Dickinson HG (2004) More than a yolk: The short life and complex times of the plant endosperm. *Trends Plant Sci* 9:507–514.
16. Li J, Berger F (2012) Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol* 195:290–305.
17. Cailleau A, Grimanelli D, Blanchet E, Cheptou P-O, Lenormand T (2018) Dividing a maternal pie among half-sibs: Genetic conflicts and the control of resource allocation to seeds in maize. *Am Nat* 192:577–592.
18. Le BH, et al. (2010) Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proc Natl Acad Sci USA* 107:8063–8070.
19. Belmonte MF, et al. (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the *Arabidopsis* seed. *Proc Natl Acad Sci USA* 110:E435–E444.
20. Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155: 1405–1413.
21. Smith NGC, Eyre-Walker A (2002) Adaptive protein evolution in *Drosophila*. *Nature* 415:1022–1024.
22. Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* 26:2097–2108.
23. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
24. Al-Shehbaz IA, O'Kane SL, Jr (2002) Taxonomy and phylogeny of *Arabidopsis* (brassicaceae). *Arabidopsis Book* 1:e0001.
25. Tang C, et al. (2007) The evolution of selfing in *Arabidopsis thaliana*. *Science* 317: 1070–1072.
26. Kawashima T, Goldberg RB (2010) The suspensor: Not just suspending the embryo. *Trends Plant Sci* 15:23–30.
27. Haig D, Westoby M (1989) Parent-specific gene expression and the triploid endosperm. *Am Nat* 134:147–155.
28. Schon MA, Nodine MD (2017) Widespread contamination of *Arabidopsis* embryo and endosperm transcriptome data sets. *Plant Cell* 29:608–617.
29. Hamilton WD (1964) The genetical evolution of social behaviour. I. *J Theor Biol* 7:1–16.
30. Dudley SA (2015) Plant cooperation. *AoB Plants* 7:plv113.
31. Alexander RD (1974) The evolution of social behavior. *Annu Rev Ecol Syst* 4:325–383.
32. Friedman WE (1995) Organismal duplication, inclusive fitness theory, and altruism: Understanding the evolution of endosperm and the angiosperm reproductive syndrome. *Proc Natl Acad Sci USA* 92:3913–3917.
33. Irizarry RA, et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
34. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) Affy–Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–315.
35. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5:R80.
36. R Core Team (2014) R: A Language and Environment for Statistical Computing (R Found Stat Comput, Vienna).
37. Ritchie ME, et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47.
38. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300.
39. Fay JC, Wyckoff GJ, Wu C-I (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* 415:1024–1026.
40. Fracassetti M, Griffin PC, Willi Y (2015) Validation of pooled whole-genome resequencing in *Arabidopsis lyrata*. *PLoS One* 10:e0140462.
41. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and inparalogs from pairwise species comparisons. *J Mol Biol* 314:1041–1052.
42. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
43. Hu TT, et al. (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* 43:476–481.
44. Briskine RV, et al. (2017) Genome assembly and annotation of *Arabidopsis halleri*, a model for heavy metal hyperaccumulation and evolutionary ecology. *Mol Ecol Resour* 17:1025–1036.
45. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
46. Suyama M, Torrents D, Bork P (2006) PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34: W609-12.
47. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973.
48. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
49. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
50. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
51. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
52. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
53. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
54. McKenna A, et al. (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
55. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
56. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
57. Huang W, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1–13.
58. Geist KS (2019) Adaptation in *Arabidopsis* seeds. Github. Available at https://github.com/ksgeist/adaptation-in-arabidopsis-seeds. Deposited April 10, 2019.

APPENDIX II

Genetic signatures of microbial altruism and cheating in social amoebas in the wild

Suegene Noh[a,b], Katherine S. Geist[a], Xiangjun Tian[c,d], Joan E. Strassmann[a], and David C. Queller[a,1]

[a]Department of Biology, Washington University in St. Louis, St. Louis, MO 63130; [b]Department of Biology, Colby College, Waterville, ME 04901; [c]The Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX 77030; and [d]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030

Edited by Nancy A. Moran, University of Texas at Austin, Austin, TX, and approved February 11, 2018 (received for review November 21, 2017)

Many microbes engage in social interactions. Some of these have come to play an important role in the study of cooperation and conflict, largely because, unlike most animals, they can be genetically manipulated and experimentally evolved. However, whereas animal social behavior can be observed and assessed in natural environments, microbes usually cannot, so we know little about microbial social adaptations in nature. This has led to some difficult-to-resolve controversies about social adaptation even for well-studied traits such as bacterial quorum sensing, siderophore production, and biofilms. Here we use molecular signatures of population genetics and molecular evolution to address controversies over the existence of altruism and cheating in social amoebas. First, we find signatures of rapid adaptive molecular evolution that are consistent with social conflict being a significant force in nature. Second, we find population-genetic signatures of purifying selection to support the hypothesis that the cells that form the sterile stalk evolve primarily through altruistic kin selection rather than through selfish direct reproduction. Our results show how molecular signatures can provide insight into social adaptations that cannot be observed in their natural context, and they support the hypotheses that social amoebas in the wild are both altruists and cheaters.

altruism | cheating | kin selection | *Dictyostelium discoideum* | social amoeba

Cooperative behavior, once associated primarily with animals like social insects, is increasingly seen as widespread in nature. Some social microbes are now providing excellent model systems for the study of cooperation and conflict, because they can be genetically manipulated or because their short lifetimes facilitate experimental evolution over many generations. However, these systems have one major disadvantage. Unlike animals, which can be directly observed and assessed in their natural environments, microbes usually need to be taken out of their natural environments for observation. With a few exceptions (1, 2), we therefore know little about microbial social adaptations in nature (3), resulting in multiple controversies over the natural adaptive importance of even some of the best-studied phenomena such as bacterial quorum sensing (4–7), siderophore production (8, 9), and biofilms (10, 11).

The social amoeba *Dictyostelium discoideum* is a microbial model system for cooperation and conflict (12, 13). In this species, single-celled amoebas join together upon starvation to form multicellular fruiting bodies (14). About 20% of the cells die in the process of forming a stalk, which supports and promotes the dispersal (15) of the other 80%, which differentiate into spores. This appears to be an instance of kin-selected altruism (12, 16, 17). Laboratory studies also show the potential for extensive cheating. Here we use "cheating" as shorthand for any competition within the fruiting body, with the essential point being that when two or more clones aggregate together, they may be in conflict over who gets to produce the reproductive spores (18, 19). However, the relevance of both kin selection and cheating in the natural environment has been questioned (20–23).

It would clearly be useful to develop some alternative methods for understanding microbial social behavior in the wild. Here we deploy theories from population genetics and molecular evolution to search for, and find, molecular signatures that reflect both kin selection and cheating in wild *D. discoideum*.

Cheating. In the laboratory, different *D. discoideum* clones readily join the same fruiting body (18), despite some recognition and segregation (24). Often one clone will show apparent cheating in the sense of getting more than its proportional (fair) share of spores (12, 25). Laboratory evolution under conditions of low kin selection leads to an increase in the frequencies of cheating mutants and a decrease in cooperation, as predicted by theory (17, 26, 27). However, the importance of cheating in the wild is uncertain, partly because relatedness is known to be quite high (16) and partly because of two plausible alternative explanations invoking adaptive trade-offs that would be hard to assess in nature. First, there is a modest number of loner cells that do not join the aggregation (28). A clone that produces fewer loner cells would, other things being equal, contribute more spores in mixtures. It could therefore appear to cheat when selection was really just operating on the trade-off between loner cells and aggregators (20, 29). Second, a clone that makes more, smaller spores could appear to cheat against a clone that makes fewer, larger spores, without necessarily having gained any cheating advantage (21).

Significance

Microbes are surprisingly social organisms and are providing model systems for the study of the evolution of cooperation and conflict. Despite their many advantages in the laboratory, such as experimental evolution, it is rarely possible to study them in the field. We therefore know little about whether cooperation and conflict are adaptively important in nature. Here we use approaches from population genetics and molecular evolution to test the adaptive relevance of social behavior in a social amoeba. We find signatures of adaptation for both kin selection and social cheating. This provides evidence that these behaviors have been important in the natural evolution of this species and more generally shows a way to study microbial social adaptation in the wild.

Author contributions: S.N., K.S.G., X.T., J.E.S., and D.C.Q. designed research; S.N. performed research; S.N., K.S.G., X.T., J.E.S., and D.C.Q. analyzed data; and S.N., K.S.G., X.T., J.E.S., and D.C.Q. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All data, including genome alignments for the 16 published *Dictyostelium discoideum* genomes, along with the code that generated the statistics, are deposited in Dryad Digital Repository (doi: 10.5061/dryad.43cp320).

[1]To whom correspondence should be addressed. Email: queller@wustl.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1720324115/-/DCSupplemental.

If cheating is common in nature and also causes resistance to cheating to evolve, as it does in the laboratory (30–32), this may lead to evolutionary conflict and increased selection pressure. A set of *D. discoideum* genes whose knockouts cause cheating showed an unusual degree of balancing selection (in which rare alleles are favored), one possible outcome of cycling or stalemate conflict (33). However, this study did not show the more pronounced arms race outcome of rapid adaptation via directional selection. Here we use RNA-seq to screen specifically for genes that change expression in chimeric mixtures of two clones. This is the precise context in which cheating would be adaptive, so it may pinpoint the genes most likely to function specifically in cheating or resistance to cheating. Moreover, these would likely be facultative cheating or resistance genes, the kind that would most likely be favored under high relatedness. (Obligate ones are less likely to be favored because, when alone, they would still pay any cost of cheating without getting any benefits.) We then test for rapid adaptive evolution in these genes relative to other genes in the genome.

Kin Selection or Direct Selection. The importance of kin selection and altruism in the wild has also been questioned. Instead of getting kin-selected benefits by altruistically helping related spores to disperse, stalk cells might instead be making the best of a bad job, doing everything they can to reproduce directly (22, 23, 34). Some evidence consistent with this view comes from the fact that the stalk is made by cells with less glucose (35), that prestalk cells are suppressed and perhaps poisoned by a chlorinated molecule produced by prespore cells (22, 23), and that prestalk cells may actually reproduce on rare occasions (36). An acknowledged weak point of this hypothesis is how an effort to reproduce would lead to producing a complex stalk (23). The question could be settled by evidence on the relative importance of personal and kin effects in the field.

The two hypotheses differ in the proposed role of prestalk cells: are they being selected to reproduce directly or instead indirectly through giving aid to kin? This results in contrasting predictions about the strength of purifying selection in prestalk and prespore cells.

To test the hypothesis that indirect kin selection is irrelevant, and all selection on prespore cells is through direct reproduction (22, 23), we use theory about how selection operates on conditionally expressed genes. Other things being equal, a gene should be selected more weakly and be more variable in proportion to the fraction of individuals that express it (37). For example, genes preferentially expressed in rarer morphs of pea aphids show relaxed purifying selection (38). In *D. discoideum*, where 80% of the cells in an aggregate become prespore cells, and 20% become prestalk cells, purifying selection against mildly deleterious mutations will be four times less effective in prestalk cells than in prespore cells. Other things being equal, genes expressed primarily in prestalk cells should therefore be four times more polymorphic than genes expressed mainly in prespore cells (37), assuming similar initial distributions of mutant effects on fitness. [Actually, the difference may be more extreme than 4:1 because we have not accounted for the fact that, even in this direct selection hypothesis, prestalk is viewed as a best-of-a-bad-job strategy (22)]. Note that many other genes in the genome may also be conditionally expressed, to unknown degrees. That is why we do not use all genes in this test, but instead compare prestalk genes against prespore genes—these are expressed in the same circumstance (fruiting) but differ in their relative proportions.

The alternative kin-selection hypothesis is that all selection on prestalk cells is indirect selection operating through effects on related spores. Here, theory predicts that the effect of indirect selection, relative to direct selection, is diluted by a factor of the relatedness coefficient (39, 40). A probable empirical example is that honeybee worker genes show lower nonsynonymous variability than queen genes (41). For *D. discoideum* fruiting bodies, relatedness is high in nature, with two estimates based on molecular markers yielding 0.97 and 0.86 (16). At these levels, prestalk genes under pure kin selection should be only 1.03–1.17 times as variable as prespore genes under direct purifying selection.

Results

Cheating. Using four pairs of wild clones, we searched for genes changing expression in chimeric mixtures. For each clone pair, the chimeric treatment involved mixing the two clones in equal proportions under starving conditions so they would form fruiting bodies. The controls were identical except that each clone was allowed to form fruiting bodies on its own, starting from the same total number of cells. We harvested RNA at the tight aggregate stage, a key stage for stalk–spore differentiation, after which gene expression patterns switch abruptly (42, 43). Using a generalized linear model (GLM) that accounts for effects of clone pair, library, and sequencing batch, we identified 79 genes that consistently and significantly differed in expression between chimeras and controls at false discovery rate = 0.10 (20 up-regulated in chimeras, 59 down-regulated; Dataset S1). It is interesting to note that the change in expression between chimeras and controls was correlated (Pearson's $r = 0.311$, $P < 0.001$) with expression changes in a previous experiment where chimeras differed only at the *tgrB1* and *tgrC1* cell adhesion loci that control clonemate recognition (44), suggesting that our response is at least partly influenced by that recognition system.

We tested the hypothesis that these chimera-biased genes would show conflict-generated high rates of adaptive evolution of coding sequence using the program DFE-α (45). It estimates a modified McDonald-Kreitman (46) statistic, α, by measuring the proportion of nonsynonymous sites between species that have been fixed by selection, using within-species polymorphisms to provide an expectation if they were due only to neutral evolution and purifying selection. The program improves on typical McDonald-Kreitman tests by allowing α to be estimated over entire gene sets, yielding greater power, and by using the estimated frequency distribution of polymorphisms to better account for low-frequency deleterious alleles (47).

We used 15 *D. discoideum* genomes from Virginia and Texas to estimate nonsynonymous and synonymous polymorphism within species and the corresponding fixed differences relative to a diverged Costa Rican outgroup clone, S6B, which is probably a different species (48). As predicted by the conflict hypothesis, adaptive evolution is significantly higher in the genes that change expression in chimeras than in genes that do not (Fig. 1, $\alpha = 0.149$ versus -0.723, $P = 0.002$ *permutation test*). This is primarily due to genes up-regulated in chimeras, although they are not significantly different from down-regulated ones (Fig. 1). We found similar results using an alternative measure of adaptive evolution, ω_A (49) (Table S1). Balancing selection is another possible outcome of cheating (33), but we found no support for this in three measures of balancing selection. Chimera-biased genes and other genes were not different, using permutation tests, for either f_{st} ($P = 0.387$) or Tajima's D ($P = 0.514$). Fay and Wu's H ($P = 0.0172$) did show a difference, but one indicating directional selection, in agreement with our other results. In this case, the signature of selection was entirely due to genes up-regulated in chimeras ($P = 0.0096$), and not those that were down-regulated in chimeras ($P = 0.759$) (Table S1).

Kin Selection or Direct Selection. The results above support the hypothesis of cheater-driven molecular evolution, but what about altruism? To test the importance of direct selection versus indirect (kin) selection, we used previously identified (42) genes with significantly greater expression in prestalk cells than prespore cells, or vice versa. To eliminate effects of selection in other contexts, we also tested a second set ($n = 145, 113$; see Dataset S1), in which we removed genes with any expression during the vegetative (single-cell) stage. We estimated their

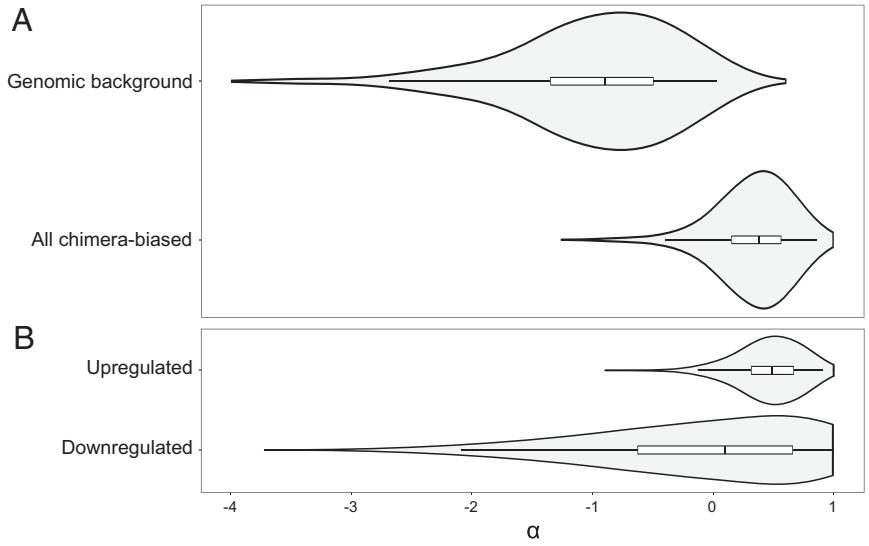


Fig. 1. Genes that change expression in chimeric mixtures show elevated rates of adaptive evolution **α**. Each violin plot shows a Gaussian kernel-density plot of 1,000 bootstrap replicates of α, the median, the interquartile range, and the 95% range or confidence interval (Table S1) (*A*) Bootstrap distributions for the 78 chimera-biased genes and for samples of 78 from genomic background genes. The chimera-biased genes show significantly higher α ($P < 0.002$ permutation test). (*B*) Bootstrap distributions for the chimera-biased genes separated into the 19 up-regulated genes and 59 down-regulated genes, both of which are significantly different from background genes (up-regulated $P = 0.006$, down-regulated $P = 0.034$, permutation tests).

variability (Table S2) using the same 15 *D. discoideum* genomes (33). Because the predictions apply only to nonneutral sites (neutral mutations are not subject to purifying selection), we focus on the nonsynonymous diversity, π_N. If a large fraction of nonsynonymous mutations were neutral, we would need to also exclude those, but that is not the case. (From the DFE-α program these fractions are low for prestalk and prespore genes combined: $f = 0.127$ and $f = 0.140$, with and without vegetative expression, respectively; Table S2.) For the two gene sets, the ratios of nonsynonymous diversity, π_N, for prestalk genes to prespore genes (prestalk π_N: prespore π_N) are 0.914 and 0.771. These are not significantly different from the two predicted values under kin selection (1.03, 1.17) but significantly differ from the value of 4 predicted by the direct selection hypothesis (Fig. 2 and Table S3).

The conclusion in favor of kin selection is not altered by two potential caveats. First, the expected ratio for two sets of random genes is a ratio of 1, close to our kin selection prediction, but the prestalk and prespore genes are not random sets. Compared with genes in the whole genome ($\pi_N = 0.00019$), both sets of prespore plus prestalk genes are significantly more variable (including vegetative expression, $\pi_N = 0.00021$, $P = 0.04$; without vegetative expression, $\pi_N = 0.00027$, $P = 0.03$, permutation tests). Moreover, the data very decisively reject the direct-selection prediction of a ratio of 4 or higher. The maximum values obtained in 10,000 bootstrap samples were only 2.24 and 1.48 for our two prestalk–prespore gene sets (Table S3).

Second, our direct-selection prediction of fourfold greater variation in prestalk genes may be too extreme, given that even our prestalk-enriched genes have some expression in prespore cells. Selection on these prestalk genes may therefore include a minority component of direct selection in prespore cells, which should tend to make selection in the two gene sets somewhat more similar. However, a far more conservative prediction is available concerning the correlation between diversity π_N and degree of prestalk versus prespore expression. The kin selection hypothesis predicts there should be little or no correlation because, with relatedness near 1, selection intensity would be roughly equal in the two tissues. In contrast, since the direct selection

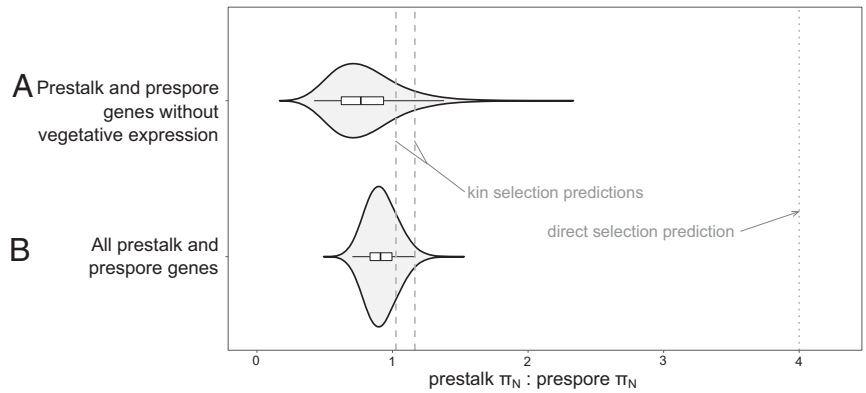


Fig. 2. Nonsynonymous diversity π_N supports kin selection, not direct selection, in prestalk cells. Violin plots (Fig. 1) for distributions from 10,000 resamples of the prestalk π_N: prespore π_N, the ratio of nonsynonymous nucleotide diversity π_N for genes expressed significantly more in prestalk to π_N for genes expressed significantly more in prespore. (*A*) All prestalk-biased genes ($n = 992$) and prespore-biased genes ($n = 879$). (*B*) Prestalk-biased genes ($n = 145$) and prespore-biased genes ($n = 113$) that are not expressed in the vegetative stage (Table S3).

hypothesis implies stronger purifying selection on prespore genes, prestalk genes should be more variable, and the correlation should be positive. In fact, the correlation is weak and negative (with vegetative expression $\tau = -0.062$, $P = 0.00025$, without $\tau = -0.096$, $P = 0.039$; Kendall's tau correlation between π_N and $\log_2$ of the prestalk: prespore expression ratio), again strongly rejecting the direct selection hypothesis.

Discussion

Testing adaptation is rarely simple because it requires understanding how the organism interacts with its natural environment. For social behavior, this is particularly difficult because it requires understanding the natural social context. For animals, we can at least observe their behavior in their natural environment. Microbes, however, are more difficult to observe, and they are typically studied in laboratory environments that may not accurately reflect their natural contexts.

For example, the social amoeba *Dictyostelium discoideum* has usually been studied in a uniclonal social context, with most work on the species being carried out on the clonal descendants of a single natural isolate, NC4. This obscured the possibility of interesting social behaviors, like cheating (18, 19) and kin recognition (24, 50, 51), that only revealed themselves when multiple clones were studied in mixtures. However, the studies are still carried out in the laboratory and might therefore miss important elements of the natural context. This has led to controversies about the adaptiveness of both cheating and altruism in *D. discoideum* (20–23).

Although one cannot usually observe microbial social adaptations operating in nature, molecular signatures of population genetics and molecular evolution can sometimes provide an alternative approach. Social conflict is expected to lead to rapid evolution of genes involved in the conflict. In *D. discoideum* the genes most likely to be specialized for cheating conflict are those that change expression in chimeric mixtures. We show that these genes do indeed show more rapid adaptive evolution, supporting the natural importance of cheating. Note that these have not been confirmed as cheating genes. However, that seems the only obvious reason why this particular set of genes should show more adaptive evolution. Similarly, the alternative hypotheses of kin-selected altruism versus direct reproduction are predicted to leave different signatures with respect to the amount of nonsynonymous variation in genes that are particularly expressed in prestalk cells. The results strongly reject the direct selection prediction and support the kin selection prediction. Thus, it appears that both cheating and kin selection are not just laboratory phenomena but are also important in the wild.

The main assumption underlying our analyses is that the observed differences in evolution are due primarily to the relative strengths of selection between the gene sets. Because we are comparing gene categories in the same population, it is reasonable to assume that other forces like drift and migration are equal. It is less certain that mutations must be equal, specifically, the distribution of selective coefficients of mutants. Although there is no specific reason to believe this assumption should fail for our gene sets, it is more questionable, and it is therefore good that kin selection and cheating are supported by other kinds of studies.

With respect to kin selection, there is evidence for all three components of kin selection. Stalk cells pay the large cost of sacrificing their lives to produce a stalk. Other cells have been assumed to benefit from the stalk by gaining more access to dispersers, an assumption supported using a model arthropod disperser in the laboratory (15). Finally, we know that relatedness within fruiting bodies is high in nature (16) in part due to kin recognition (50) but probably also due to passive population structure (25, 52, 53).

However, high relatedness makes our other finding—evidence for cheating in the wild—more surprising because most fruiting bodies are clonal. However, it should be remembered that selective forces that operate rarely, for example, certain pathogens, can still exert important selective forces. There is also some prior evidence with respect to cheating. We know that unequal contribution to spores is common among laboratory clones (18, 19), that mutants in many genes affect this (27), and that cheating mutants spread readily under conditions of low relatedness (17, 27). High relatedness in the field must prevent some cheating, especially from high-cost obligate cheaters that cannot fruit on their own (16). However, this high cost does not apply to facultative cheaters that cheat by changing expression only when a foreign partner is present, so selection might still favor some of these cheaters. Our results suggest that this is indeed the case. Additional supporting data come from several sources. A mutation accumulation experiment showed that random mutations tend to decrease cheating ability, which is the result expected if cheating is a fitness component (54). The presence of kin recognition and segregation seems best explained as a partial solution to the problem of foreign clones that might do harm (55). Finally, clones in chimeras show possible cheating adaptations. Chimeric slugs travel less far, consistent with cells trying to stay out of the front region that will form the stalk (56). Chimeras also produce more spores and higher spore-to-stalk ratios (19). Alternative explanations are possible for most of these phenomena. For example, chimeras could have reduced slug migration due to lower cell–cell adhesion, a side-effect of mismatches at their kin-recognition *tgrB1*/*tgrC1* loci (57). Collectively, however, all these phenomena build a consistent case for the importance of cheating in the wild.

The use of molecular signatures like these might also be employed in other controversies about microbial social evolution (4–11). To be useful, it is necessary to identify a target set of genes hypothesized to be subject to a particular kind of selection and then measure a reliable signature of that kind of selection. This might not always be feasible for some systems and questions, but this approach does add a valuable tool to other approaches such as making the laboratory setting more natural and conducting experiments in the field (3). When it is feasible, the method of using population-genetic or molecular-evolution signatures is superior in one important respect. It yields a more comprehensive record of selection, one that is automatically integrated over the full geographical range sampled and over very long periods of time.

Materials and Methods

Amoeba Samples. To detect genes changing expression in chimeras, we tested four pairs of *D. discoideum* strains or clones, originally isolated from soil from Mt. Lake Biological Station in Virginia: QS6 with QS160, QS4 with QS174, QS18 with QS154, and QS17 with QS157, a sufficient number to exclude effects that are idiosyncratic to particular clone pairs. For molecular evolution analyses we used the genomes of 16 strains. For polymorphism data, we used 15 *D. discoideum* strains, eight strains from Virginia and seven strains from Texas, all those that were available after excluding populations with only one strain (33). For divergence estimates, we compared these strains to a tropical outgroup clone S6B from Costa Rica, probably a separate species (48).

Chimera-Biased Genes: RNA Sequencing. We prepared samples from four strain pairs using the following procedures. We grew amoebas on SM/5 agar plates [2 g glucose, 2 g BactoPeptone (Oxoid), 2 g yeast extract (Oxoid), 0.2 g $MgCl_2$, 1.9 g KH_2PO_4, 1 g K2HPO4 and 15 g agar per liter] with $\sim 2 \times 10^5$ spores and a food bacterium *Klebsiella pneumoniae* (250 μL at 1.5 optical density). When amoebas were in log-phase growth, we used a sterile plastic spatula to scrape cells from the plates into KK2 buffer and washed three times to remove most of the food bacteria. For each replicate, we spread 10^8 cells in 1,000 μL KK2 onto 47-mm-diameter nitrocellulose filters (Millipore) for each of the two unmixed clonal strains and 10^8 total cells for the 50:50 chimeric mix of strains, resulting in a trio of samples (two clonal, one chimeric). When 90% of the cells were in the tight aggregate stage, we washed cells off of each filter with KK2 buffer into a 5× volume of RNAlater for storage at 4 °C . For each strain pair, we repeated this process three times

on different dates. We extracted RNA using a protocol for cytoplasmic RNA purification from animal cells with a Qiagen RNeasy Mini Kit, with modifications based on Kaul and Eichinger (58). From here, we prepared sequencing libraries using the standard Illumina protocol for the poly-A–tailed stranded mRNA library prep kit. We constructed three batches of libraries, each run in one sequencing lane, with each containing a full replicate of the experiment: two clonal and one chimeric sample for all four strain pairs. Sequencing was done on an Illumina Hiseq2500 for 50-bp single-end reads at the Washington University in St. Louis Genome Technology Access Center (GTAC).

Chimera-Biased Genes: Alignment and Differential Expression. After quality control of raw reads (removal of reads shorter than 12 bp and those with any N nucleotides), reads from each library were mapped onto the *D. discoideum* reference genome (downloaded Dec 2014 from Ensembl Protist v1.25). Before alignment, we masked the known duplicated region on chromosome 2 of the AX4 reference genome (2: 3016083–3768654) using bedtools v2.19.1 (59). We used GSNAP v2014-12–17 (60) using default alignment parameters, except for only allowing a single alignment path to be followed to avoid chimeric reads (npaths = 1). GSNAP uses an oligomer chaining method combined with dynamic programming to align transcript reads to genomic sequence and is splice junction aware. We derived splice junctions based on the *D. discoideum* GFF3 gene feature annotations (downloaded September 2015) from dictybase.org (61). We used Picard v1.128 (downloaded from broadinstitute.github.io/picard) to sort alignments and fix read groups. We used R v3.2.1 (62) and Bioconductor package ShortRead v1.26.0 (63) to assess sequence read quality statistics. We excluded one replicate of the strain pair QS6 and QS160 from our analyses because the bamQA report generated by ShortRead indicated it did not meet quality standards. We then used RSeQC v2.5 (64) to look at alignment statistics and read distributions across genomic features. We had aligned 5.7–28.1 million (median 10.3) read tags or reads split by indels per library, and 4.9–27.3 million of these were aligned to annotated coding genes.

We extracted read counts from uniquely mapped reads using HTSeq v0.5.4p5 (65). Only reads with the correct strand orientation and mapping quality above 20 were counted. We imported these counts into R and examined the correlation between replicates within each strain pair across all expressed genes. The correlations across pairwise comparisons of replicates within strains were generally very high (mean $r = 0.94$), while the excluded sample showed a much lower correlation ($r = 0.61$), justifying our decision to omit it from further analysis.

We used DESeq2 v1.8.1 (66) to test for evidence of significant differential expression. We tested 9,089 genes, using a GLM model (count ~ batch + pair + condition), with sequencing lane and library preparation batch as the factor batch, strain pair identity as the factor pair, and the clonal vs. chimeric condition of aggregation as the factor condition. DESeq2 uses a negative binomial distribution to model read counts and correct for sequencing library size using median-of-ratios size factors and uses empirical Bayes shrinkage estimators that correct count variance in individual genes based on other genes with similar expression levels (66).

Prespore and Prestalk Genes. Our tests of the roles of direct and indirect (kin) selection in the evolution of stalk cells require identification of sets of genes with expression that is relatively specialized in prestalk and prespore cells. We used the candidate prespore and prestalk genes reported by Parikh et al. (42) with slightly reduced sample sizes after removing noncoding elements (see below). This study had separated prestalk and prespore cells and performed RNA-seq to determine which genes were significantly more expressed by each cell type (42). To reduce the influence of selection that occurs during the vegetative stage, we also tested a more restrictive set of 113 prespore and 145 prestalk genes that had no gene expression detected in vegetative cells.

Polymorphism and Divergence. We tested whether our candidate chimera-biased genes show high rates of adaptive evolution consistent with an arms race scenario driven by social conflict. We cleaned and clipped raw Illumina reads from the 16 *D. discoideum* strains using ngsShoRT v2.2 (https://research.bioinformatics.udel.edu/genomics/ngsShoRT/). We generated mpileup files that merged strains within each geographic location using samtools v0.1.19 (67, 68), with adjusted mapping quality (-C 50) and a minimum basecall quality of 30 (-Q 30). We used Varscan v2.3.9 (69) to call variants from these merged mpileup files. We specified a minimum coverage of 20 reads per SNP and filtered for strand bias at a *P* value of 0.01. We then resplit the VCF file by strain and reconstructed the sequences of over 12,000 genes using GATK FastaAlternateReferenceMaker. We used custom scripts to convert these genomic FASTA files into coding sequences by removing introns and reverse complementing as necessary. Because our downstream tests assume that genes are coding, we removed noncoding RNAs, pseudogenes, and transposable elements as annotated in the *D. discoideum* genome (61, 70). This eliminated one of our chimerism genes from further analyses.

We used PolyMORPHOrama (71) to estimate average pairwise nucleotide diversity (π) using a Jukes-Cantor correction (72) and counted the numbers of polymorphisms per site class, both nonsynonymous (Pn) and synonymous (Ps). PolyMORPHOrama also generated the allele frequency spectra that we used in estimates of Tajima's *D* (73), Fay and Wu's *H* (74), and other downstream analyses of molecular evolution (see below). Next, we created a consensus FASTA of the 15 wild clones for each gene for comparison with S6B as outgroup, using ancestral sequence reconstruction method implemented by codeml (runmode = 0, CodonFreq = 2) in PAML v4.8 (75) and a custom Perl script. From this, we used codeml (runmode = −2, CodonFreq = 2) to generate our pairwise estimates of Dn and Ds. We used vcftools v0.1.12a (76) to estimate Weir-Cockerham's f_{st} (77) directly from VCF files. We imported these data into R and identified the genes associated with each variant using ChIPpeakAnno v3.2.2 Anno (78).

Molecular Evolution Analyses. We assessed the relative strength of purifying selection on prestalk and prespore genes by taking the ratio of their nonsynonymous π's: prestalk π_N: prespore π_N. The mean π's are calculated for the numerator and denominator before dividing to reduce variance and eliminate zero denominators. This ratio was tested against predicted values of 4 for direct selection on prestalk genes, versus 1.025 and 1.165 for indirect selection on prestalk genes (the reciprocals of two relatedness estimates).

To test for adaptive selection on genes up-regulated in chimeras, we used tests of selection based on the McDonald-Kreitman test (46), originally instituted as a 2 × 2 Fisher's Exact test to compare nonsynonymous (Pn) to synonymous (Ps) polymorphism to nonsynonymous (Dn) to synonymous (Ds) divergence for a single gene. Related metrics have been developed to summarize the effects of numerous selective events over multiple genes. These include: α, the proportion of nonsynonymous substitutions driven to fixation by positive selection (45, 79, 80); ω_a, the rate of adaptive fixation relative to neutral fixation (80); and f, the proportion of nonsynonymous mutations that are effectively neutral. We generated these four parameters for our gene sets with the maximum likelihood method of Eyre-Walker and Keightley (45), implemented in the command-lined version of DFE-α v2.15 (www.homepages.ed.ac.uk/pkeightl/). We used a custom Perl wrapper to sum the allele frequency spectra generated by PolyMORPHOrama, incorporate divergence information, and perform either permutations or bootstrapping.

Statistics. Confidence intervals for all molecular evolution parameters are obtained by bootstrapping. From the *i* genes contributing to a statistic *X*, we repeatedly drew samples (either 1,000 or 10,000; see below) of *i* genes with replacement, recomputed *X* from each sample, and defined the 95% confidence interval as between the upper and lower 2.5% of the distribution.

Statistical tests for molecular evolution parameters were either bootstrap tests (tests against a predicted value *Y*) or permutation tests for differences between two samples. For a test of a difference in a statistic between two samples, X_1–X_2, based on *i* and *j* genes, we randomly drew, without replacement, samples (either 1,000 or 10,000; see below) of *i* and *j* genes from the total of $i + j$ genes and recalculated the difference X_1–X_2 for each. For comparisons against the genomic background, we randomly drew, without replacement, samples of *i* genes from the total of $i + j$ genes. *P* values were calculated as the proportion of times the permuted difference was more extreme than zero in the direction predicted. For two-tailed bootstrap tests of an estimate X_1 against a predicted value *Y*, we repeatedly drew with replacement samples of *i* genes from the *i* original genes and recalculated X_1. From this distribution, the percentage in the shorter tail cut off by *Y*, doubled, is the two-tailed *P* value.

For π, Tajima's *D*, Fay and Wu's *H*, and f_{st}, we drew 10,000 resamples. Because α and other site frequency spectrum metrics required extensive computation (rerunning the DFE-α program) for each replicate, we drew 1,000 resamples.

ACKNOWLEDGMENTS. We thank Tracy Douglas for the sequence of clone S6B as well as Tyler Larsen and two referees for comments on the manuscript. This material is based upon work supported by the National Science Foundation under Grants DEB1146375 and IOS-1656756.

1. Bozdag GO, Greig D (2014) The genetics of a putative social trait in natural populations of yeast. *Mol Ecol* 23:5061–5071.
2. Köhler T, Buckling A, van Delden C (2009) Cooperation and virulence of clinical Pseudomonas aeruginosa populations. *Proc Natl Acad Sci USA* 106:6339–6344.
3. Tarnita CE (2017) The ecology and evolution of social behavior in microbes. *J Exp Biol* 220:18–24.
4. Redfield RJ (2002) Is quorum sensing a side effect of diffusion sensing? *Trends Microbiol* 10:365–370.
5. Keller L, Surette MG (2006) Communication in bacteria: An ecological and evolutionary perspective. *Nat Rev Microbiol* 4:249–258.
6. Diggle SP, Gardner A, West SA, Griffin AS (2007) Evolutionary theory of bacterial quorum sensing: When is a signal not a signal? *Philos Trans R Soc Lond B Biol Sci* 362:1241–1249.
7. West SA, Winzer K, Gardner A, Diggle SP (2012) Quorum sensing and the confusion about diffusion. *Trends Microbiol* 20:586–594.
8. Zhang XX, Rainey PB (2013) Exploring the sociobiology of pyoverdin-producing Pseudomonas. *Evolution* 67:3161–3174.
9. Andersen SB, Marvig RL, Molin S, Krogh Johansen H, Griffin AS (2015) Long-term social dynamics drive loss of function in pathogenic bacteria. *Proc Natl Acad Sci USA* 112:10756–10761.
10. Webb JS, Givskov M, Kjelleberg S (2003) Bacterial biofilms: Prokaryotic adventures in multicellularity. *Curr Opin Microbiol* 6:578–585.
11. Xavier JB, Foster KR (2007) Cooperation and conflict in microbial biofilms. *Proc Natl Acad Sci USA* 104:876–881.
12. Strassmann JE, Queller DC (2011) Evolution of cooperation and control of cheating in a social microbe. *Proc Natl Acad Sci USA* 108:10855–10862.
13. Shaulsky G, Kessin RH (2007) The cold war of the social amoebae. *Curr Biol* 17:R684–R692.
14. Kessin RH (2001) *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity* (Cambridge Univ Press, Cambridge, UK), p 308.
15. Smith J, Queller DC, Strassmann JE (2014) Fruiting bodies of the social amoeba *Dictyostelium* discoideum increase spore transport by *Drosophila*. *BMC Evol Biol* 14:105.
16. Gilbert OM, Foster KR, Mehdiabadi NJ, Strassmann JE, Queller DC (2007) High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc Natl Acad Sci USA* 104:8913–8917.
17. Kuzdzal-Fick JJ, Fox SA, Strassmann JE, Queller DC (2011) High relatedness is necessary and sufficient to maintain multicellularity in *Dictyostelium*. *Science* 334:1548–1551.
18. Strassmann JE, Zhu Y, Queller DC (2000) Altruism and social cheating in the social amoeba Dictyostelium discoideum. *Nature* 408:965–967.
19. Buttery NJ, Rozen DE, Wolf JB, Thompson CRL (2009) Quantification of social behavior in D. discoideum reveals complex fixed and facultative strategies. *Curr Biol* 19:1373–1377.
20. Martínez-García R, Tarnita CE (2016) Lack of ecological and life history context can create the illusion of social interactions in Dictyostelium discoideum. *PLoS Comput Biol* 12:e1005246.
21. Wolf JB, et al. (2015) Fitness trade-offs result in the illusion of social success. *Curr Biol* 25:1086–1090.
22. Atzmony D, Zahavi A, Nanjundiah V (1997) Altruistic behaviour in *Dictyostelium discoideum* explained on the basis of individual selection. *Curr Sci* 72:142–145.
23. Zahavi A, Zahavi A (1997) *The Handicap Principle: A Missing Piece of Darwin's Puzzle* (Oxford Univ Press, Oxford).
24. Ostrowski EA, Katoh M, Shaulsky G, Queller DC, Strassmann JE (2008) Kin discrimination increases with genetic distance in a social amoeba. *PLoS Biol* 6:e287.
25. Gilbert OM, Strassmann JE, Queller DC (2012) High relatedness in a social amoeba: The role of kin-discriminatory segregation. *Proc Biol Sci* 279:2619–2624.
26. Ennis HL, Dao DN, Pukatzki SU, Kessin RH (2000) *Dictyostelium* amoebae lacking an F-box protein form spores rather than stalk in chimeras with wild type. *Proc Natl Acad Sci USA* 97:3292–3297.
27. Santorelli LA, et al. (2008) Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature* 451:1107–1110.
28. Dubravcic D, van Baalen M, Nizak C (2014) An evolutionarily significant unicellular strategy in response to starvation in Dictyostelium social amoebae. *F1000 Res* 3:133.
29. Tarnita CE, Washburne A, Martinez-Garcia R, Sgro AE, Levin SA (2015) Fitness tradeoffs between spores and nonaggregating cells can explain the coexistence of diverse genotypes in cellular slime molds. *Proc Natl Acad Sci USA* 112:2776–2781.
30. Khare A, et al. (2009) Cheater-resistance is not futile. *Nature* 461:980–982.
31. Levin SR, Brock DA, Queller DC, Strassmann JE (2015) Concurrent coevolution of intra-organismal cheaters and resisters. *J Evol Biol* 28:756–765.
32. Hollis B (2012) Rapid antagonistic coevolution between strains of the social amoeba *Dictyostelium discoideum*. *Proc Biol Sci* 279:3565–3571.
33. Ostrowski EA, et al. (2015) Genomic signatures of cooperation and conflict in the social amoeba. *Curr Biol* 25:1661–1665.
34. Zahavi A, Harris KD, Nanjundiah V (2018) An individual-level selection model for the apparent altruism exhibited by cellular slime moulds. *J Biosci* 43:49–58.
35. Thompson CRL, Kay RR (2000) Cell-fate choice in Dictyostelium: Intrinsic biases modulate sensitivity to DIF signaling. *Dev Biol* 227:56–64.
36. Kuzdzal-Fick J, Foster K, Queller D, Strassmann J (2007) Exploiting new terrain: An advantage to sociality in the slime mold *Dictyostelium* discoideum. *Behav Ecol* 18:433–437.
37. Van Dyken JD, Wade MJ (2010) The genetic signature of conditional expression. *Genetics* 184:557–570.
38. Purandare SR, Bickel RD, Jaquiery J, Rispe C, Brisson JA (2014) Accelerated evolution of morph-biased genes in pea aphids. *Mol Biol Evol* 31:2073–2083.
39. Linksvayer TA, Wade MJ (2009) Genes with social effects are expected to harbor more sequence variation within and between species. *Evolution* 63:1685–1696.
40. Hall DW, Goodisman MA (2012) The effects of kin selection on rates of molecular evolution in social insects. *Evolution* 66:2080–2093.
41. Warner MR, Mikheyev AS, Linksvayer TA (2017) Genomic signature of kin selection in an ant with obligately sterile workers. *Mol Biol Evol* 34:1780–1787.
42. Parikh A, et al. (2010) Conserved developmental transcriptomes in evolutionarily divergent species. *Genome Biol* 11:R35.
43. Rosengarten RD, et al. (2015) Leaps and lulls in the developmental transcriptome of *Dictyostelium discoideum*. *BMC Genomics* 16:294.
44. Hirose S, Benabentos R, Ho H-I, Kuspa A, Shaulsky G (2011) Self-recognition in social amoebae is mediated by allelic pairs of tiger genes. *Science* 333:467–470.
45. Eyre-Walker A, Keightley PD (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol Biol Evol* 26:2097–2108.
46. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
47. Keightley PD, Eyre-Walker A (2007) Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics* 177:2251–2261.
48. Douglas TE, Kronforst MR, Queller DC, Strassmann JE (2011) Genetic diversity in the social amoeba *Dictyostelium discoideum*: Population differentiation and cryptic species. *Mol Phylogenet Evol* 60:455–462.
49. Gossmann TI, et al. (2010) Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Mol Biol Evol* 27:1822–1832.
50. Benabentos R, et al. (2009) Polymorphic members of the lag gene family mediate kin discrimination in *Dictyostelium*. *Curr Biol* 19:567–572.
51. Ho H-I, Hirose S, Kuspa A, Shaulsky G (2013) Kin recognition protects cooperators against cheaters. *Curr Biol* 23:1590–1595.
52. smith j, Strassmann JE, Queller DC (2016) Fine-scale spatial ecology drives kin selection relatedness among cooperating amoebae. *Evolution* 70:848–859.
53. Buttery NJ, et al. (2012) Structured growth and genetic drift raise relatedness in the social amoeba Dictyostelium discoideum. *Biol Lett* 8:794–797.
54. Hall DW, Fox S, Kuzdzal-Fick JJ, Strassmann JE, Queller DC (2013) The rate and effects of spontaneous mutation on fitness traits in the social amoeba, Dictyostelium discoideum. *G3 (Bethesda)* 3:1115–1127.
55. Gilbert OM (2015) Histocompatibility as adaptive response to discriminatory within-organism conflict: A historical model. *Am Nat* 185:228–242.
56. Foster KR, Fortunato A, Strassmann JE, Queller DC (2002) The costs and benefits of being a chimera. *Proc Biol Sci* 269:2357–2362.
57. Gruenheit N, et al. (2017) A polychromatic 'greenbeard' locus determines patterns of cooperation in a social amoeba. *Nat Commun* 8:14171.
58. Kaul M, Eichinger L (2006) Analysis of gene expression using cDNA microarrays. *Methods in Molecular Biology: Dictyostelium Discoideum Protocols*, eds Eichinger L, Rivero F (Humana Press, Totowa, NJ), pp 75–93.
59. Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
60. Wu TD, Watanabe CK (2005) GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21:1859–1875.
61. Fey P, et al. (2009) dictyBase–A *Dictyostelium* bioinformatics resource update. *Nucleic Acids Res* 37:D515–D519.
62. R Core Team (2014) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
63. Morgan M, et al. (2009) ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics* 25:2607–2608.
64. Wang L, Wang S, Li W (2012) RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* 28:2184–2185.
65. Anders S, Pyl PT, Huber W (2015) HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169.
66. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550.
67. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993.
68. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
69. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22:568–576.
70. Chisholm RL, et al. (2006) dictyBase, the model organism database for Dictyostelium discoideum. *Nucleic Acids Res* 34:D423–D427.
71. Haddrill PR, Bachtrog D, Andolfatto P (2008) Positive and negative selection on noncoding DNA in Drosophila simulans. *Mol Biol Evol* 25:1825–1834.
72. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia Univ Press, New York).
73. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
74. Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.
75. Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
76. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
77. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
78. Zhu LJ, et al. (2010) ChIPpeakAnno: A bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* 11:237.
79. Fay JC, Wyckoff GJ, Wu C-I (2001) Positive and negative selection on the human genome. *Genetics* 158:1227–1234.
80. Gossmann TI, Keightley PD, Eyre-Walker A (2012) The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol* 4:658–667.