Winter 12-15-2021

Weedy rice as a model system for the study of microevolutionary interactions in agricultural contexts

Marshall Jon Wedger
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Ecology, Evolution, and Population Biology

Dissertation Examination Committee:

Kenneth Olsen, Chair
Carlos Botero
Rachel Penczykowski
David Queller
Christopher Topp

Weedy Rice as a Model System for the Study of Microevolutionary Interactions in Agricultural Contexts

by
Marshall Jon Thomas Wedger

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

May 2022
St. Louis, Missouri

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Chapter 2

Appendix I

Chapter 3

Appendix II

Chapter 4

Appendix III

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank the people, without which this degree just would not have been possible. To Ken, thank you for everything you've done for me – from the *extensive* text edits, to the chit-chats in the hallways. Thank you for always having your door open and always making time for me and the other students. You've created a fun and engaging lab culture that I have been privileged to be a part of for the last number of years. I look forward to our conversations each week and revel in our discussions. I hope we maintain a relationship as colleagues, and I can't wait to see where your research program takes you and your lab in the years to come. Sometimes I catch myself looking back to the scientist I was six years ago, and I know I have you to thank for the growth I've achieved.

It would be dishonest of me to not also thank my undergraduate advisor, Briana Gross. To Briana, I cannot begin to thank you enough for taking me into your lab and starting me on this path. I knocked on your door before you even had the chance to set up your lab. You gave me a few of your research papers and, kindly, told me to go away and come back later. The seeds of this dissertation were sown in those papers and the work we did. My passion for plant population genetics and my interest in domestication as a model for evolution can be traced to those silly little apple leaves in the Lee lab freezers. Thank you, for everything.

Next, I thank my committee members, Carlos Botero, Rachel Penczykowski, Christopher Topp, and David Queller for their guidance and insight across this thesis. Every meeting we had was useful and furthered my goals as a student researcher. The kindness and thoughtfulness of these members has been a boon to my time here at Wash-U. A special thank you to Chris Topp, for lending me his expertise and lab equipment / space for one chapter of this work.

I must acknowledge generous funding from the National Science Foundation. The work presented here was partially funded by the Graduate Research Fellowship Program (GRFP)

which allowed me to pursue my interests in the field beyond what was otherwise possible (Fellow ID: 2017221153). Weedy rice research in the Olsen lab has also been supported through the NSF Plant Genome Research Program (IOS-1032023 and IOS-1947609).

To the members of the Olsen Lab, old and new, thank you for all the time we spent together. I'd like to specifically thank Sara, Dave, Jordan, and Rachel. Sara, you taught me more about life than you know. Believe-it-or-not, you were the first friend I had who spoke *so* candidly about everything in life. I enjoyed our conversations, and they always gave me things to think about for days afterword. Dave, when I first arrived, I poked fun at you for having two whole monitors. Here I sit writing this document surrounded by four computer screens, one of which is dedicated to Linux. Thank you for starting me on my coding adventure and only sometimes laughing at my spaghetti code. Jordan, we arrived the same year and from day one I knew you were an interesting person. I will always look back fondly on the things we got into – from tea bag slinging, to pouring drinks into plants while your back was turned. I'll be rooting for you while you chase the R1 dragon. Rachel, we became fast-friends, and you were always looking out for me. You were someone I could reach out to for help, and I'll be eternally grateful for our interactions. Thank you to Linda and Mike for all your support during my experiments.

To the broader collection of grad-students I've had the privilege of knowing, thank you for just being awesome. When I was selecting grad schools, I had four offers and I chose Wash-U because of the grad students. More than the academic opportunities, or the people I'd get to work with, it was the grad students who drew me here. A special thank you to Kim, who took it upon herself to plan social events that I, begrudgingly, attended but always had fun. I'd also like to say a huge thank you to Cassie, Tony, Vince, Tyler, James S., Kim, Brendon, Erika, Rachel, Dave, Jordan, and all the rest for truly unforgettable Dungeons and Dragons, Betrayal at the House on the Hill, and other game nights. Truly, they kept me sane during my time here.

I'd like to specifically thank three fellow grad students for taking me under their tutelage. I learned everything I know about being a well-rounded grad student and person from all of you. Sara, I thanked you earlier, but here again I must thank you for being as open and truthful as you've always been. Kim, we joke about your effectiveness as an assigned mentor, but I honestly could not have asked for someone better. You and I are complete social opposites, but I've learned from you how to take a step out and enjoy the company of the people around me. You've also taught me how to take a lashing in board games, but that's another story. Bo, your passion for STEM outreach, specifically YSP, has instilled in me a love for the same. You were the one to suggest Gretchen and I take on the challenge of YSP director and it was only through your confidence in me that I took the position. I hope to continue bringing my love of science to those around me and I will always strive to put in the same effort I've seen from you.

The next set of people I'd like to thank is somewhat unorthodox, but they have been instrumental in maintaining my mental health. Everyone needs a hobby, and mine is online video games. Through this hobby I've met and stayed in contact with so many interesting people. Firstly, I've been able to stay connected with friends I've made throughout me life. Thank you, Matt, Zack, and Jordan, for all the times in college and beyond. All three of you were groomsman in my wedding and I still love you as much now as I did then. Next, I've met friends of real-life friends that have become friends of my own. Thank you, Lyle, Hank, and Josh, for the near daily meetups online. I cannot begin to explain how therapeutic it is to hunt some ghosts, or tame some dinosaurs, or shoot some aliens after a rough week. Lastly, I've met complete strangers from around the world who have become people I talk to regularly. I've met a student in Florida about to receive her degree in aerospace engineering. I've met a long-haul trucker from Virginia. I've met a pilot who moved from the US to France and listened as she tried to re-earn her license. I've met a janitor at a police station in a far northern village of Canada. I've met a nurse from Las Vegas and commiserated with him about the pandemic. I've met a doctor from

a rural town in Minnesota and listened to an over-the-phone diagnosis when he forgot to mute himself. Thank you to, Rose, Jimmy, Lei, Eli, Gray, Lalush, Fireball, Twisted, Kerry, Icho, Guinness, Dakotakid, and so many more. To all these people, and their stories, thank you for all the great memories. I love talking with people who have a different life experience from me, and I've been lucky to meet so many of you.

Finally, to my family, who have been my support for these long years. To the Drehers, I could not have asked to marry into a more welcoming, kind, and interesting family. You've welcomed me into your family and treated me like I've always been there. Margie and Jim, I see the passion with which you support your kids, and it inspires me to hopefully one day do the same. Annie, Nick, Bella, and Max, the walls of my apartment are plastered with art from your family, and I can't help but smile every time I see them. The joy and light these simple drawings bring to my life is invaluable. Bear, Skyler, and Ruby, I think the future is bright for the three of you. I can't wait to see what the next ten years will bring and am excited to be a part of it.

My brothers have ever-been simultaneous sources of tremendous joy, and incredible suffering. Still, I'm not sure I would want anything different. Caleb, you and I are as oil and water. I remember wandering around our childhood neighborhood and random older kids would stop me and ask if I "was Caleb's brother". I would always be proud to answer yes! At the same time, teachers would ask "aren't you Caleb's brother" and I would have to explain how "yes, but…". Despite this I've always been proud to call you my brother and you've taught me so much about life, and what's worth lingering on. So, thank you. To Hannah, the joy and kindness you bring to every one of our conversations makes me so happy to have you in my life. The compassion and excitement you have for your students shines from you for everyone to see, and that is why I'm so excited to work with you, hopefully, soon. Will and Jo, heck you're both cute. I'm so excited to watch you both grow up and see who you become.

Russell, we've been apart for quite some time, but I still think about you often. You and I have been though a lot and you've come out so strong. I still remember sitting in our room trading Pokémon cards even though neither of us knew how to play. I remember chasing you around the house with a music stand only for you to turn around and chase me through the house with something bigger. I'm proud to call you my brother and I know life has conspired to separate us for a while, but things will get better, I know it. So, Thank you for all the memories – past and future. Gabrielle, Vivian, and Gordon, I'm so excited to get to know all of you much better as we continue on in life.

My mom and her partner have been the guiding star to which I've set my compass for so many years. Sue, like a wizard, you arrived in my life not late, nor early, but precisely when you were meant to. Since that day, we've been like peas-in-a-pod and I'm eternally grateful to have you in my life. Thank you for getting me my first job, and teaching me how to do an oil change, and bringing me snowboarding, and for all the love and support you've shown me in the last decade.

Mom, from the bottom of my heart to the tips of my toes, thank you for the unending support you've given me. You were at every baseball game, you were next to me every time I ended up in the hospital, and you were there when I moved out to college and then grad school. When I was a kid, one of my chores was to dust a truly obscene amount of picture frames. One that I thought was particularly dumb had a saying something like "Family gives you roots to stay grounded, and wings so you can fly". Now, nearly 20 years wiser, I can confidently say that you have succeeded on all accounts. I know where my roots are, and I've traveled the country learning everything I can. Thank you.

Finally, to my partner, Allie. We met in junior high, when we were both dating each other's best friends. It took until our junior year of high school, sharing a blanket on the shores

of First Lake watching the 4th of July fireworks, to finally start dating and we've been together ever since. Nearly 12 years at this point - and what 12 years it's been. We went to college together and made memories I'll always cherish in Duluth. Then we moved to St. Louis and got two of the goodest cats ever. You battled your way through nursing school and earned your nursing degree, which I'm immensely proud of you for. Somewhere in there we got married on the bank of the Mississippi and we got to laugh as you tried to recite your vows. In high school you stole my phone and changed your name to "The Greatest Bestest Thing In My Life" with a heart emoji. I think you were joking back then, but I haven't changed it back and now I can thank you, truly, for being the single greatest bestest thing in my life.

Marshall J. T. Wedger
Washington University in St. Louis
May 2022

ABSTRACT OF THE DISSERTATION

Weedy rice as a model system for the study of microevolutionary interactions in agricultural contexts

By

Marshall J.T. Wedger

Doctor of Philosophy in Biology and Biomedical Sciences
Ecology, Evolution, and Population Biology
Washington University in St. Louis, 2022
Professor Kenneth M. Olsen, Chair

Just under one-half of the global population relies on cultivated rice (*Oryza sativa* as their primary source of calories, making the optimization of rice agriculture immensely important. One of the primary constraints to rice agriculture is the de-domesticated (feral form of rice known as 'weedy rice' that aggressively competes for space, soil nutrients, and light. Heavy infestation can reduce crop yields by as much as 80%. As a closely-related weedy descendant of cultivated rice, chemical control is difficult in rice fields, and physical weeding is labor intensive, time consuming, and largely ineffective due to early life-stage mimicry of the crop.

Weedy rice occurs in almost every world region that cultivates rice and in most cases has evolved from local cultivated varieties. This thesis focuses on two regions. The first is the southern United States, where two strains, black-hull awned (BHA and straw-hull awnless (SH have coexisted for >150 years at relatively equal frequencies. Historically, there has been minimal gene flow between these two independently-evolved strains or with cultivated varieties of rice. The other region is Thailand, where, unlike the US, the reproductively compatible wild rice ancestor (*Oryza rufipogon* is present at the margins of fields. Both US and Thai weedy rice populations have been rapidly adapting to changing agricultural practices in recent decades. Using these two sampled regions, my dissertation research focused on three questions: 1 Do independently evolved strains of weedy rice use similar or different genetic mechanisms when evolving competitive growth traits — specifically, with respect to root system architecture? 2

How have weedy rice genomes adapted to rapidly changing 21st Century agricultural practices? And 3) How does the presence of an obligately outcrossing wild ancestor alter the gene flow dynamics of the predominantly selfing cultivated-weedy rice system?

Chapter One provides an overview of weedy rice and the evolutionary forces that have shaped it worldwide.

In **Chapter Two** I used two large recombinant inbred line (RIL) populations to perform a QTL mapping study using root system architecture as my trait of interest. Both mapping populations were generated from a cultivated × weedy rice cross using the same cultivated plant as the paternal line (SH × cultivated and BHA × cultivated). Seedlings were grown for 13 days in an agarose gel, after which they were placed on a turntable and imaged at 5° increments in rotation. All 72 images were integrated to produce a 3D model of root system architecture. Over 100 root traits were measured on more than 600 seedlings. Mapping identified 10 traits that mapped to different genomic locations in each mapping population. Only a single trait, convex hull volume, mapped to the same genomic location in both populations. The results of this chapter suggest that different populations, experiencing the same selection pressures, rely on very different genetic mechanisms to evolve similar weedy phenotypes.

Chapter Three focuses on the effects of changing agricultural practices on the genomes of southern US weedy rice strains (SH and BHA). In 2002, herbicide resistant rice was commercialized. At the same time, hybrid rice technology was introduced as a high-yield alternative to traditional inbred cultivars. Due to the propensity of hybrid rice to drop its seed onto the soil before harvest, leading to crop volunteers, this technology became a bridge for gene flow between cultivated and weedy rice. Weedy rice populations became herbicide resistant by the mid 2000s, and experimentation showed this was due to adaptive introgression of the crop resistance allele. We collected 48 samples of weedy rice in 2018 and performed

whole genome sequencing to determine the long-term genomic consequences of this event. We discovered that, while descendants of both the SH and BHA populations are represented, their genomes have been irrevocably altered from the strains that had existed prior to 2000. First, we find that most weedy rice plants (44/48) are of crop-weed hybrid origin and that BHA is the dominant weedy ancestor in these hybrid derivates. Moreover, a local ancestry analysis reveals that within the genomes of these contemporary weeds, most (~70%) of the genome is derived from the weedy ancestors, suggesting a selective advantage of the weed genome in contemporary hybrid-derived weeds. Lastly, we find that while herbicide resistance is primarily derived from herbicide resistant crop cultivars, the four weeds of non-hybrid origin also contain resistance alleles, suggesting that convergent evolution has played a part in the persistence of weedy rice.

Chapter Four seeks to identify the impact of local wild rice populations on gene flow dynamics of weedy rice in Thailand. We utilized two complementary data sets, twelve neutral microsatellite markers, and three domestication-associated genes, to track gene flow from wild and cultivated rice into weedy rice populations. Interestingly, while both data sets identified gene flow, there was little overlap between them in the accessions showing admixed ancestry. This suggests a temporal discordance (allowing for multiple generations of recombination between the gene flow events), with these historic introgression events occurring separately and quite long ago.

The **Conclusion Chapter** provides synthesis and conclusions.

INTRODUCTION

Genetics of Adaptation - *Domesticated plant species*

The importance of studying domesticated species for insights on adaptation and natural selection has been noted since the inception of the field of evolutionary biology (Darwin, 1859). Darwin discusses, in the very first chapter of *On the Origin of Species*, the curious phenotypic differences between domesticated and wild species, including the weight of duck bones, the development of cow udders, and the ear shape of dogs. In each of these scenarios he posits a purely adaptive explanation: domestic ducks fly less often, so their leg bones weigh more; cows which are routinely milked have evolved udders adapted to their regular use; domestic dogs have droopy ears because the ear muscles are no longer needed to detect danger as often.

The examples above focus on animals, but domesticated plants are particularly useful to the study of adaptation due to the amount of available information. Nearly all domesticated plants have a known evolutionary age bounded by the end of the last ice age and the origin of agriculture (10-12,000 years ago) (Mannion, 1999). Additionally, their cultural importance as staples of food security, for religious ceremonies, and in barter economies has resulted in the extensive preservation of seeds, cultural artifacts, and other archaeological remains (Gross & Zhao, 2014; Swarts *et al.*, 2017; Guo *et al.*, 2019; Milla & Osborne, 2021). The availability of this historical information has allowed researchers to develop a time-calibrated understanding of the genomic consequences of strong selection over a relatively short period of time on a plant genome (Flint-Garcia, 2013).

Modern domesticated plant species are primarily under the selective influence of ‘artificial selection’, a process by which humans take the role of primary selector. This process often results in the emergence of traits that are beneficial for yield, taste, or food preservation, but detrimental for the species in other ways. This tradeoff has been dubbed the “cost of

domestication" and has been documented in the accumulation of deleterious mutations (Moyers *et al.*, 2018), the loss of population-wide genomic diversity (Raduski *et al.*, 2021), and the loss of interspecific competitive traits (Ottis *et al.*, 2005; Burgos *et al.*, 2006).

Using standard phenotype-to-genotype approaches, such as quantitative trait locus (QTL) mapping and genome wide association studies (GWAS), and genotype-to-phenotype approaches such as outlier analyses, some of the genes responsible for traits behind this cost of domestication have been identified (Ross-Ibarra *et al.*, 2007). Single genes responsible for adaptation to increased yield in crowded fields have been mapped in many domesticated species, such as semi-dwarfism caused by a loss-of-function mutation at *SD1* in rice (Ashikari *et al.*, 2002) and *Rht* in wheat (Peng *et al.*, 1999), and reduced leaf angle of maize caused by a mutation at *LG1* (Tian *et al.*, 2019). Both of these phenotypes increase yield by reducing intraspecific competition, allowing for more dense plantings. This reduced intraspecific competitive ability, however, comes at the cost of reduced ability to compete effectively against the agricultural weeds that invade and exploit resource-rich crop fields.

Although much of the research effort in the field of domesticated plant adaptation has gone to identifying single gene-to-phenotype connections, more complex systems of adaptation have been identified as well. Sunflower (*Helianthus annuus*) has been shown to utilize transposable element (TE) variation as the basis for local adaptation (Todesco *et al.*, 2020), while the wild progenitor of domesticated *Camelina sativa* appears to have escaped a relatively restricted native area via polyploidization (Brock *et al.*, 2018).

Despite the insights gleaned on the genetics of adaptation from domesticated plant species, the inherently artificial nature of the system makes it difficult to test hypotheses related to other microevolutionary forces. In addition, generalizations about natural selection developed by studying artificial selection, or 'methodological selection' as Darwin names it, must be

tempered due to the persistently strong, unidirectional selection under domestication, which has a much larger effect on a crop species than most instances of natural selection. In this respect, studies that instead focus on the *non-domesticated* relatives of crop species can offer attractive alternative systems for studying the genomic impacts of natural selection and other microevolutionary forces. Such species can offer the genomic toolkit available for crop species while also providing a study system with more natural microevolutionary processes at play.

Weedy crop relatives

In many cases, plant domestication was accompanied by the evolution of weedy crop relatives (Ellstrand *et al.*, 2010). These strains are often either the same biological species as the cultivated plant (e.g., weedy rice, weedy rye, and shattercane sorghum) or crop-wild hybrid derivatives (e.g., weedy beet, weedy sunflower, and johnsongrass). Due to their close phylogenetic relationship to crops, and their persistence in agricultural fields between years, weedy crop relatives are useful systems to study all five basic processes of microevolution (natural selection, gene flow, mutation, mating systems, and genetic drift), as described below.

Natural Selection - Agricultural fields are, by design, incredibly resource rich for plants. Nutrients, water, and pest protection are in abundance, making these environments perfect for any weedy species that can exploit them. However, under the eyes of a watchful farmer, these environments are also incredibly hostile to unwanted plants. The combination of these two factors makes the phenotypic and genomic evolution of weedy crop relatives interestingly dynamic. In many cases, this leads to an evolutionary arms race between agricultural weeds and farmers in which each new agro-biotech advance is slowly made less effective by the inevitable creep of natural selection (Vigueira *et al.*, 2013b).

Weedy species have adapted to survive and reproduce in human-mediated agricultural environments and have thus evolved traits to facilitate this. Some traits, such as a selfing mating system (Arnaud *et al.*, 2010) and disease resistance (Goad *et al.*, 2020), can be seen as 'crop-like' in that they are also favored under domestication, while other traits such as seed dispersal (Paterson *et al.*, 1995), and seed dormancy (Gu *et al.*, 2011) are more 'wild-like'. Still more traits are unique to agricultural weed species, including crop mimicry and highly competitive growth in crop fields (Olsen *et al.*, 2007). Together, this suite of traits, collectively called the "agricultural weed syndrome" (Vigueira *et al.*, 2013b), produces weedy species uniquely adapted to agricultural fields, capable of surviving and thriving there while devastating crop productivity.

Gene Flow - One of the most environmentally important questions raised by modern agriculture is: how will the escape of crop alleles (transgenic or otherwise) through gene flow affect wild populations of interfertile species? Crop species with weedy relatives, especially those growing near their native wild ancestor, are particularly prone to this escape concern. Mediating these risks is of paramount concern in both an economic and academic sense. Escape of herbicide resistance alleles into wild or weedy populations could render expensive technologies essentially useless while the ecological consequences of rapidly altering relative fitness of natural populations (or populations within a broader community) could be enormous. Evidence for crop-wild gene flow mediated by weedy crop relatives is abundant (e.g., Sagnard *et al.*, 2011; Wongtamee *et al.*, 2017). The continued study of gene flow dynamics in these systems can also shine a light on processes related to adaptive or maladaptive introgression.

Mutation - As mentioned above, many weedy crop relatives are direct descendants of cultivated species. In these weeds, the double bottleneck associated with domestication followed by feralization would be expected to produce weedy strains with very little genetic diversity – even when compared to that of the crop. Despite this glaring lack of genetic diversity, we often see very rapid adaptation in weedy crop relatives (Vigueira *et al.*, 2013b). At least

some of these adaptations have emerged through novel mutations. For example, in some populations of Europe and Northern China, novel private weedy alleles at *sh4* and the related *qsh1* genes have, while not functionally confirmed as causal variants, been linked to the reemergence of the shattering phenotype (Zhu *et al.*, 2021). In another example, novel mutations at the herbicide resistance gene *ALS* have pre-adapted a small sub-population of weedy rice in the US to contemporary applications of the IMI-class of herbicides (Sales *et al.*, 2008).

Mating systems – Cultivated annual plants are often bred to be mostly selfing in order to make breeding programs more efficient. Studies on mating system shifts in weedy crop relatives have shown that this transition has diverse effects on fitness measures. Weedy rye has been shown to shift mating system, from outcrossing to fully selfing in colonizing situations (Sun & Corke, 1992). Outside of colonizing situations, when weedy species are already present at high frequency, both selfing and outcrossing have been show as dominant mating strategies. In weedy rice, a largely selfing species, low levels of outcrossing have been maintained and have facilitated adaptive introgression of certain beneficial cultivated alleles (discussed below in Chapter Three and Chapter Four). One study in weedy beet showed that weedy populations segregating for the self-incompatibility gene *Sf*, had no significant relationship between outcrossing rate and weed density (Arnaud *et al.*, 2010).

Genetic Drift – The stochastic processes underlying genetic drift as an evolutionary force have been well studied using weedy crop relatives. Drift is often invoked in the context of interactions with other evolutionary forces. Perhaps the best-studied interaction is the interplay between drift (a force that leads to loss of variation within populations and increased differentiation among populations) and gene flow (a homogenizing force with the opposite effects). The results of these studies seem to be species specific. Southern US weedy rice shows no evidence of geographic structure throughout its range due to the mobile nature of

seeds that get caught in shared farm equipment (Reagon *et al.*, 2010; Li *et al.*, 2017). Weedy beets, formed by crop-wild hybridization, show strong evidence of isolation by distance as a result of introgression from local wild populations (Arnaud *et al.*, 2010). Besides gene flow-drift interactions, genetic drift has also consistently been invoked to explain the rapid adaptation seen in weedy crop relatives, as phenotypically and genetically divergent populations have more possible responses to a novel selection pressure (Sun & Corke, 1992; Burger *et al.*, 2006; Fogliatto *et al.*, 2020).

The understanding of microevolutionary theory, including each of these underlying forces, has been substantially improved by the use of weedy crop relatives as model systems. The short and time-stamped evolutionary history of these species has offered invaluable insights into the field. Beyond their academic utility for heightened evolutionary insights, weedy crop relatives are also major pests in agricultural fields, so understanding their evolutionary history and contemporary interactions with cultivated plants is paramount to food security around the world.

Thesis Study System

Asian cultivated rice (*Oryza sativa* L.) is one of the most important cereal crops, with over one-half of humans from around the world relying on it as their primary source of calories (Muthayya *et al.*, 2014; Wang *et al.*, 2018). Thus, stable rice production is essential for global food security. Rice production and consumption is highest in China, India, and large parts of sub-Saharan Africa, where human population sizes continue to grow at a rapid pace. Meeting this rise in population sizes with an equivalent response in agricultural production is a challenge, as much of the world's high-quality farming land is either claimed or rapidly degrading (Potapov

et al., 2021). The solution to this problem is not to find more land, but to be more efficient with the land already in use.

Asian cultivated rice was first domesticated from the wild ancestor *Oryza rufipogon* in the Yangtze River basin in China about 8-14,000 years ago, leading to the emergence of domesticated *Oryza sativa* subspecies *japonica* (Yasuda, 2008). A second, largely independent center of cultivation arose around the same time in the Ganges River plains of India, producing *Oryza sativa* subspecies *indica* (Fuller *et al.*, 2010). After a period of time that may be considered 'proto-domestication', when *indica* and *japonica* rice were grown independently, emerging trade routes led to the exchange of seed stock and the introgression of desirable *japonica* traits into *indica* genomic backgrounds. Today, Asian cultivated rice is composed of five major recognized varietal groups. The subspecies *indica* is composed of two genetically distinct variety groups, *indica* and *aus*, while the subspecies *japonica* is comprised of three genetically distinct variety groups, *tropical japonica, temperate japonica,* and *aromatic* (Garris *et al.*, 2005)

One of the primary constraints to rice agriculture is the de-domesticated derivative of rice known as 'weedy rice' (also called 'red rice', due to its characteristic reddish-brown pericarp). Weedy rice has independently evolved in almost every world region where rice is grown (Qiu *et al.*, 2020), including China (Guo *et al.*, 2018), Japan (Imaizumi *et al.*, 2021) Malaysia (Song *et al.*, 2014), Korea (He *et al.*, 2017), Italy (Grimm *et al.*, 2013), France (Bourdineaud, 2020), and Columbia (Hoyos *et al.*, 2020). In each case, at least one of the predominant weedy rice strains are descendants of the locally grown cultivated rice variety, suggesting that it is remarkably easy to evolve weedy rice from different cultivated rice varieties.

Weedy rice is devastating in agricultural fields: just one plant per square meter can lead to a >200 kilogram per hectare loss of yield (Burgos *et al.*, 2006), while heavy infestations can

lead to complete yield loss and abandonment of fields. These yield losses are due to competition for soil nutrients (Burgos *et al.*, 2006), sunlight (E. Schaedler *et al.*, 2020), and space (Cao *et al.*, 2007; Yang *et al.*, 2018a). Annual economic costs of weedy rice are massive as well, with losses in the hundreds of millions each year due to lower yield and quality (Chauhan, 2020).

Regardless of the varietal source or country of origin, nearly all weedy rice shares a suite of phenotypes, including rapid growth, seed dormancy, and seed shattering, while remaining incredibly diverse for other traits such as hull color, height, and presence / absence of secondary dispersal mechanisms such as awns and barbs (Zhu *et al.*, 2012; Hua *et al.*, 2015; Qi *et al.*, 2015; Roma-Burgos *et al.*, 2021a). The repeated emergence of weed-adaptive phenotypes in independently-evolved strains has spurred researchers to examine the evolutionary and genetic mechanisms by which this convergence occurs. The results of this line of inquiry are discussed below in Chapter One.

Weedy rice in the southern US – One exception to the rule of weedy rice de-domestication from local cultivated rice is in the southern US rice growing region (including parts of Arkansas, Mississippi, Missouri, Louisiana, and Texas). Weedy rice in this region is *indica* or *aus* derived, while the region grows exclusively *tropical japonica* varieties with no recorded history of *indica* or *aus* cultivation. Population genetic studies have traced the likely origin of these two weed strains to contaminated imports from South or Southeast Asia (Londo & Schaal, 2007; Reagon *et al.*, 2010).

There are two distinct strains of weedy rice that have historically co-occurred in southern US rice fields. The first is derived from *indica* varieties of rice and has been given the name “strawhull awnless” (SH) due to its grain phenotype (Londo & Schaal, 2007; Reagon *et al.*, 2010). The SH strain is known for its crop mimicry characteristics. Plants are short in stature,

with straw-colored hulls, a lack of awns and barbs, an open panicle exertion, and a closed panicle branching architecture (Roma-Burgos *et al.*, 2021a). Despite the superficial mimicry of cultivated rice, SH weedy rice is highly competitive against the crop in the field. In one study, SH strains took up nearly twice the amount of soil-nitrogen during seed filling when compared to cultivated competitors (Burgos *et al.*, 2006).

The second strain of weedy rice historically found in the southern US is “black-hull awned” (BHA). BHA strains are derived from *aus* varieties of cultivated rice (Londo & Schaal, 2007; Reagon *et al.*, 2010). As the name suggests, BHA strains have a dark outer hull with a long and barbed awn, which are both characteristic of wild rice (and some traditional, unimproved crop landraces). BHA strains are readily identifiable due to their tall stature and rapid growth, purple culms and wild-like grain characteristics. While the SH strain escapes detection, the BHA strain competes more overtly with cultivated rice by conspicuously outgrowing it. The tall stature of BHA plants means they capture more sunlight, shading out their cultivated counterparts (Estorninos *et al.*, 2005).

Both SH and BHA strains have been the predominant weeds of southern US rice fields for the last two centuries. Despite their long history of sympatry, there has been very little hybridization recorded between SH and BHA or between either strain and the local *tropical japonica* cultivated varieties. Outcrossing rates are well below one percent (Shivrain *et al.*, 2009).

Southern US weedy rice is one the best studied weedy crop relatives, and as such has been used as a model to study multiple facets of microevolutionary theory. The resources for these strains of weedy rice are extensive. Firstly, cultivated rice was the first crop plant to have a fully annotated reference genome sequence (Yu *et al.*, 2002), and it is continuously updated (Kawahara *et al.*, 2013). This has led to the possibility of large whole genome sequencing

projects in weedy rice (Li *et al.*, 2017, Chapter Three of this Dissertation) and interesting insights into the de-domestication process. Next, crop × weed mapping populations have been developed using SH and BHA strains (Qi *et al.*, 2015; Goad *et al.*, 2020), allowing for investigations into the genetics of weed adaption (see Chapter Two of this Dissertation). Finally, the economic costs of controlling ag weeds have given impetus to developing genetic tools for tracing weed origins; this has opened possibilities to study gene flow and genetic drift in environments with and without the wild ancestor present (see Chapter Four of this Dissertation).

US weedy rice and 21st Century changes in US rice agriculture - Finally, we find ourselves in times of rapid technological advances. These shifts have not bypassed the agricultural sector and have, in fact, exposed weedy populations to novel selective pressures. Rice agricultural technologies, specifically, have yielded major advances in the last two decades. Prior to the year 2003, US rice was grown entirely as elite inbred varieties. Advances in heterosis technologies spurred the production of high-yield hybrid rice varieties which have been slowly adapted by farmers and now constitute ~50% of US rice acreage (Moldenhauer *et al.*, 2020). One issue with hybrid rice is the partial reemergence of the wild-like seed shattering phenotype. This leads to hybrid seed falling into the soil where they overwinter and grow as so-called 'weedy volunteers' the next year (Singh *et al.*, 2017b). Segregating alleles in F_2 hybrids and their descendants lead to a wide variation in flowering time, which increases outcrossing rates with weedy rice. Thus, hybrid rice technologies have been found to act as a bridge for gene flow between cultivated and weedy rice (Singh *et al.*, 2017d).

Concurrent with the introduction of hybrid rice was the commercialization of Clearfield™ rice, a non-transgenic herbicide resistant (HR) cultivar. HR rice is resistant to the imidazolinone class of herbicides due to nucleotide substitutions in the acetolactate synthase (*ALS*) gene that result in individual amino acid replacements in ALS enzyme (Sudianto *et al.*, 2013). As ALS is required for aromatic amino acid synthesis in plants, inhibition of this enzyme is lethal. The

Clearfield™ technology was introduced in 2002, and by 2004 farmers were reporting herbicide resistant strains of weedy rice. Upon further investigation it was found that resistance in weedy rice was conferred by adaptive introgression of the cultivated allele into weedy populations using hybrid rice as a gene flow bridge (Burgos *et al.*, 2014). The combination of these technological changes and their impacts have thus opened pressing new questions into the mechanisms by which contemporary weedy rice is evolving and adapting.

Chapters of the Dissertation

Chapter One of this dissertation reviews the history of the study of global weedy rice and the insights gleaned from each technological step forward in molecular genetic techniques. Firstly, we discuss how SSRs helped develop our understanding of the de-domestication origin. Next, we review how candidate genes uncovered the genetic basis of the 'agricultural weed syndrome' in weedy rice populations worldwide. Lastly, we come to the era of whole genome sequencing and lay out the next steps in the study of weedy rice evolution. This chapter was published in *Ecological Genetics and Genomics* (Wedger & Olsen, 2018).

Chapter Two is the first of three data chapters. In this chapter we utilized two previously generated QTL mapping populations to map root system architecture traits in cultivated and weedy rice. We use 2- and 3-dimentional imaging techniques in an agarose gel medium to measure 13-day old root traits. We use a random forest machine learning model to develop an understanding of root traits specific to weedy rice and discuss how these traits might contribute to the competitive nature of weedy rice. Next, we compare where in the genome each root trait maps and discuss how the lack of overlap suggests that each population has utilized entirely different genetic mechanisms to reach similar root phenotypes. This chapter was published in *New Phytologist* (Wedger *et al.*, 2019b).

In **Chapter Three** we performed whole genome sequencing on 48 weedy rice plants collected in 2018 from 5 fields in Greene County, Arkansas. We demonstrate how the genomes of these plants, which have experienced nearly two decades of herbicide application in the presence of hybrid rice cultivars, differ greatly from their ancestors from the 1990's. We document the escape of HR alleles into weedy rice and investigate the genomic consequences of hybridization as a response to strong selection. First, we report the contemporary fates of historic SH and BHA populations. Next, we explore the genomes of contemporary crop-weed hybrids and estimate the degree to which one ancestral genome has become over-represented. Lastly, we perform herbicide resistance trials and map the known herbicide resistance haplotypes at *ALS*. We uncover one allele that was thought to be lost in weedy populations and two instances of potential convergent evolution. This Chapter is in preparation for submission at *Nature Ecology and Evolution*.

The final data chapter, **Chapter Four**, moves away from the United States and into Thailand to investigate the impact of gene flow in regions of the world where the wild ancestor of cultivated and weedy rice, *Oryza rufipogon*, is abundant. We examine three known domestication genes and 12 SSRs in 124 cultivated, 166 weedy, and 98 wild rice accessions from three rice growing regions of Thailand. We find that both datasets identify gene flow in weedy rice but differ on the exact accessions. We discuss the apparent discordance of these results, the role of genetic drift, and suggest potential evolutionary histories which might result in these patterns. This chapter was published in the *Journal of Heredity* (Wedger *et al.*, 2019a).

Chapter Four is followed by a **Conclusions chapter** that provides a summary and synthesis of the dissertation chapters.

Works Cited

Arnaud J-F, Fénart S, Cordellier M, Cuguen J. **2010**. Populations of weedy crop-wild hybrid beets show contrasting variation in mating system and population genetic structure. *Evolutionary Applications* **3**: 305–318.

Ashikari M, Sasaki A, Ueguchi-Tanaka M, Itoh H, Nishimura A, Datta S, Ishiyama K, Saito T, Kobayashi M, Khush GS, *et al.* 2002. Loss-of-function of a Rice Gibberellin Biosynthetic Gene, GA20 oxidase (*GA20ox-2*), Led to the Rice 'Green Revolution'. *Breeding Science* **52**: 143–150.

Bourdineaud JP. **2020**. Toxicity of the herbicides used on herbicide-tolerant crops, and societal consequences of their use in France. *Drug and Chemical Toxicology* **0545**.

Brock JR, Dönmez AA, Beilstein MA, Olsen KM. **2018**. Phylogenetics of Camelina Crantz. (Brassicaceae) and insights on the origin of gold-of-pleasure (*Camelina sativa*). *Molecular Phylogenetics and Evolution* **127**: 834–842.

Burger JC, Lee S, Ellstrand NC. **2006**. Origin and genetic structure of feral rye in the western United States. *Molecular Ecology* **15**: 2527–2539.

Burgos NR, Norman RJ, Gealy DR, Black H. **2006**. Competitive N uptake between rice and weedy rice. *Field Crops Research* **99**: 96–105.

Burgos NR, Singh V, Tseng TM, Black H, Young ND, Huang Z, Hyma KE, Gealy DR, Caicedo AL. **2014**. The Impact of Herbicide-Resistant Rice Technology on Phenotypic Diversity and Population Structure of United States Weedy Rice. *Plant Physiology* **166**: 1208–1220.

Cao QJ, Li B, Song ZP, Cai XX, Lu BR. **2007**. Impact of weedy rice populations on the growth and yield of direct-seeded and transplanted rice: Research paper. *Weed Biology and Management* **7**: 97–104.

Chauhan BS. **2020**. Grand Challenges in Weed Management. *Frontiers in Agronomy* **1**: 1–4.

Darwin C. **1859**. *On the origin of Species by Means of natural Selection, or Preservatoin of Favoured Races in the Struggle for Life*. London: John Murray.

E. Schaedler C, U.M. Taborda C, A.P. Goulart F, M. Chiapinotto D, Pinho PJ. **2020**. Rice root growth and development in competition with weedy rice. *Planta Daninha* **38**: 1–7.

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE. **2010**. Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications* **3**: 494–504.

Estorninos LE, Gealy DR, Gbur EE, Talbert RE, McClelland MR. **2005**. Rice and red rice interference. II. Rice response to population densities of three red rice (*Oryza sativa*) ecotypes. *Weed Science* **53**: 683–689.

Flint-Garcia SA. **2013**. Genetics and Consequences of Crop Domestication. *Journal of Agricultural and Food Chemistry* **61**: 8267–8276.

Fogliatto S, Ferrero A, Vidotto F. **2020**. How Can Weedy Rice Stand against Abiotic Stresses? A Review. *Agronomy* **10**: 1284.

Fuller DQ, Sato Y-I, Castillo C, Qin L, Weisskopf AR, Kingwell-Banham EJ, Song J, Ahn S-M, van Etten J. **2010**. Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* **2**: 115–131.

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S. **2005**. Genetic Structure and Diversity in *Oryza sativa* L. *Genetics* **169**: 1631–1638.

Goad DM, Jia Y, Gibbons A, Liu Y, Gealy D, Caicedo AL, Olsen KM. **2020**. Identification of Novel QTL Conferring Sheath Blight Resistance in Two Weedy Rice Mapping Populations. *Rice*

13: 21.

Grimm A, Fogliatto S, Nick P, Ferrero A, Vidotto F. **2013**. Microsatellite markers reveal multiple origins for Italian weedy rice. *Ecology and Evolution* **3**: 4786–4798.

Gross BL, Zhao Z. **2014**. Archaeological and genetic insights into the origins of domesticated rice. *Proceedings of the National Academy of Sciences* **111**: 6190–6197.

Gu XY, Foley ME, Horvath DP, Anderson J V., Feng J, Zhang L, Mowry CR, Ye H, Suttle JC, Kadowaki KI, *et al.* 2011. Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. *Genetics* **189**: 1515–1524.

Guo L, Qiu J, Li LF, Lu B, Olsen K, Fan L. **2018**. Genomic clues for crop–weed interactions and evolution. *Trends in Plant Science* **23**: 1102–1115.

Guo S, Zhao S, Sun H, Wang X, Wu S, Lin T, Ren Y, Gao L, Deng Y, Zhang J, *et al.* 2019. Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. *Nature Genetics* **51**: 1616–1623.

He Q, Kim KW, Park YJ. **2017**. Population genomics identifies the origin and signatures of selection of Korean weedy rice. *Plant Biotechnology Journal* **15**: 357–366.

Hoyos V, Plaza G, Li X, Caicedo AL. **2020**. Something old, something new: Evolution of Colombian weedy rice (*Oryza* spp.) through de novo de-domestication, exotic gene flow, and hybridization. *Evolutionary Applications* **13**: 1968–1983.

Hua L, Wang DR, Tan L, Fu Y, Liu F, Xiao L, Zhu Z, Fu Q, Sun X, Gu P, *et al.* 2015. *LABA1*, a domestication gene associated with long, barbed awns in wild rice. *Plant Cell* **27**: 1875–1888.

Imaizumi T, Ebana K, Kawahara Y, Muto C, Kobayashi H, Koarai A, Olsen KM. **2021**.

Genomic divergence during feralization reveals both conserved and distinct mechanisms of parallel weediness evolution. *Communications Biology* **4**: 952.

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, Mccombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, *et al.* 2013. Improvement of the *oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**: 3–10.

Li L-FF, Li Y-LL, Jia Y, Caicedo AL, Olsen KM. **2017**. Signatures of adaptation in the weedy rice genome. *Nature Genetics* **49**: 811–814.

Londo JP, Schaal BA. **2007**. Origins and population genetics of weedy red rice in the USA. *Molecular Ecology* **16**: 4523–4535.

Mannion AM. **1999**. Domestication and the origins of agriculture: an appraisal. *Progress in Physical Geography* **23**: 37–56.

Milla R, Osborne CP. **2021**. Crop origins explain variation in global agricultural relevance. *Nature Plants* **7**: 598–607.

Moldenhauer KK, Scott B, Hardke J. **2020**. *B.R. Wells Arkansas Rice Research Studies 2019*.

Moyers BT, Morrell PL, McKay JK. **2018**. Genetic costs of domestication and improvement. *Journal of Heredity* **109**: 103–116.

Muthayya S, Sugimoto JD, Montgomery S, Maberly GF. **2014**. An overview of global rice production, supply, trade, and consumption. *Annals of the New York Academy of Sciences* **1324**: 7–14.

Olsen KM, Caicedo AL, Jia Y. **2007**. Evolutionary genomics of weedy rice in the USA. *Journal of Integrative Plant Biology* **49**: 811–816.

Ottis B V., Smith KL, Scott RC, Talbert RE. **2005**. Rice yield and quality as affected by cultivar and red rice (*Oryza sativa*) density. *Weed Science* **53**: 499–504.

Paterson AH, Schertz KF, Lin YR, Liu SC, Chang YL. **1995**. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, *Sorghum halepense* (L.) Pers. *Proceedings of the National Academy of Sciences* **92**: 6127–6131.

Peng J, Richards DE, Hartley NM, Murphy GP, Devos KM, Flintham JE, Pelica F, Sudhakar D, Christou P, Centre JI, *et al.* 1999. 'Green revolution' genes encode mutant gibberellin response modulators. *Nature letters* **400**: 256–261.

Potapov P, Turubanova S, Hansen MC, Tyukavina A, Zalles V, Khan A, Song X-P, Pickens A, Shen Q, Cortez J. **2021**. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nature Food* **3**.

Qi X, Liu Y, Vigueira CC, Young ND, Caicedo AL, Jia Y, Gealy DR, Olsen KM. **2015**. More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Molecular Ecology* **24**: 3329–3344.

Qiu J, Jia L, Wu D, Weng X, Chen L, Sun J, Chen M, Mao L, Jiang B, Ye C, *et al.* 2020. Diverse genetic mechanisms underlie worldwide convergent rice feralization. *Genome Biology* **21**: 70.

Raduski AR, Herman A, Pogoda C, Dorn KM, Van Tassel DL, Kane N, Brandvain Y. **2021**. Patterns of genetic variation in a prairie wildflower, *Silphium integrifolium*, suggest a non-prairie origin and locally adaptive variation. *American Journal of Botany* **108**: 145–158.

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL. **2010**. Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evolutionary Biology* **10**.

Roma-Burgos N, Butts TR, Werle IS, Bottoms S, Mauromoustakos A. **2021**. Weedy rice

update in Arkansas, USA, and adjacent locales. *Weed Science* **69**: 514–525.

Ross-Ibarra J, Morrell PL, Gaut BS. **2007**. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *Proceedings of the National Academy of Sciences* **104**: 8641–8648.

Sagnard F, Deu M, Dembélé D, Leblois R, Touré L, Diakité M, Calatayud C, Vaksmann M, Bouchet S, Mallé Y, *et al.* 2011. Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild–weedy–crop complex in a western African region. *Theoretical and Applied Genetics* **123**: 1231–1246.

Sales MA, Shivrain VK, Burgos NR, Kuk YI. **2008**. Amino Acid Substitutions in the Acetolactate Synthase Gene of Red Rice (*Oryza sativa*) Confer Resistance to Imazethapyr . *Weed Science* **56**: 485–489.

Shivrain VK, Burgos NR, Gealy DR, Sales MA, Smith KL. **2009**. Gene flow from weedy red rice (*Oryza sativa* L.) to cultivated rice and fitness of hybrids. *Pest Management Science* **65**: 1124–1129.

Singh V, Burgos NR, Singh S, Gealy DR, Gbur EE, Caicedo AL. **2017a**. Impact of volunteer rice infestation on yield and grain quality of rice. *Pest Management Science* **73**: 604–615.

Singh V, Singh S, Black H, Boyett V, Basu S, Gealy D, Gbur E, Pereira A, Scott RC, Caicedo A, *et al.* 2017b. Introgression of Clearfield™ rice crop traits into weedy red rice outcrosses. *Field Crops Research* **207**: 13–23.

Song BK, Chuah TS, Tam SM, Olsen KM. **2014**. Malaysian weedy rice shows its true stripes: Wild *Oryza* and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular Ecology* **23**: 5003–5017.

Sudianto E, Beng-Kah S, Ting-Xiang N, Saldain NE, Scott RC, Burgos NR. **2013**.

Clearfield® rice: Its development, success, and key challenges on a global perspective. *Crop Protection* **49**: 40–51.

Sun M, Corke H. **1992**. Population genetics of colonizing success of weedy rye in Northern California. *Theoretical and Applied Genetics* **83**: 321–329.

Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peeples M, Lepak N, Prim L, Romay MC, *et al.* 2017. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**: 512–515.

Tian J, Wang C, Xia J, Wu L, Xu G, Wu W, Li D, Qin W, Han X, Chen Q, *et al.* 2019. Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* **365**: 658–664.

Todesco M, Owens GL, Bercovich N, Légaré J-S, Soudi S, Burge DO, Huang K, Ostevik KL, Drummond EBM, Imerovski I, *et al.* 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**: 602–607.

Vigueira CC, Olsen KM, Caicedo AL. **2013**. The red queen in the corn: Agricultural weeds as models of rapid adaptive evolution. *Heredity* **110**: 303–311.

Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, *et al.* 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**: 43–49.

Wedger MJ, Olsen KM. **2018**. Evolving insights on weedy rice. *Ecological Genetics and Genomics* **7**–**8**: 23–26.

Wedger MJ, Pusadee T, Wongtamee A, Olsen KM. **2019a**. Discordant Patterns of Introgression Suggest Historical Gene Flow into Thai Weedy Rice from Domesticated and Wild Relatives. *Journal of Heredity* **110**: 601–609.

Wedger MJ, Topp CN, Olsen KM. **2019b**. Convergent evolution of root system architecture in two independently evolved lineages of weedy rice. *New Phytologist* **223**: 1031–1042.

Wongtamee A, Maneechote C, Pusadee T, Rerkasem B, Jamjod S. **2017**. The dynamics of spatial and temporal population genetic structure of weedy rice (*Oryza sativa* f. *spontanea* Baker). *Genetic Resources and Crop Evolution* **64**: 23–39.

Yang XF, Li LL, Xu Y, Kong CH. **2018**. Kin recognition in rice (*Oryza sativa*) lines. *New Phytologist* **220**: 567–578.

Yasuda Y. **2008**. Climate Change and the Origin and Development of Rice Cultivation in the Yangtze River Basin, China. *AMBIO: A Journal of the Human Environment* **37**: 502–506.

Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**: 79–92.

Zhu Y, Ellstrand NC, Lu BR. **2012**. Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication. *Ecology and Evolution* **2**: 2106–2113.

Zhu Y-Q, Fang J, Wang Y, Pang L-H, Lu B-R. **2021**. Key Roles of De-Domestication and Novel Mutation in Origin and Diversification of Global Weedy Rice. *Biology* **10**: 828.

CHAPTER ONE

Evolving insights on weedy rice

Abstract

Agricultural weeds that have evolved from de-domesticated (feral) crop plants cause millions of dollars in annual yield losses worldwide and are one of the primary barriers to global crop productivity. Weedy rice (*Oryza sativa* f. *spontanea*) is a de-domesticated form of rice that has evolved multiple times independently from different cultivated rice varieties. This weedy crop relative has recently emerged as a genomic model system for studying the genetic basis of agricultural weed evolution and the mechanisms that govern the parallel evolution of independently-evolved weed strains. In this review we highlight findings from recent genetics and genomics studies that have led to our current understanding of weedy rice evolution.

1.1 Introduction

Crop fields account for more than one-tenth of the Earth's total surface area [1], making them one of the most widespread terrestrial habitats on the planet. Agricultural weeds have evolved to exploit these fertile habitats through a variety of different mechanisms. Some of the more striking of these include close mimicry of crop species by unrelated weeds [2] and weed evolution from feral crop varieties through de-domestication [3]. Regardless of the mechanism leading to their evolution, many cropland weed species share a suite of traits collectively called the "agricultural weed syndrome" [4]. These include some crop-like traits that allow them to thrive in agricultural habitats, such as the ability to grow upright in crowded, high-density crop fields and to reproduce within a narrow window of time. They also possess some wild-like traits, such as freely-dispersing seed and strong seed dormancy. This combination of crop-like and wild-like traits makes agricultural weeds particularly well suited for proliferation in croplands, while escaping human detection and eradication efforts. The repeated evolution of agricultural weed phenotypes is an active and important avenue of weed science research; it has also

recently gained attention a model system for studying the genetics of rapid convergent evolution [4-6].

On a genomic level, one of the best studied agricultural weeds is weedy rice (*Oryza sativa* f. *spontanea*), a de-domesticated form of cultivated Asian rice (*Oryza sativa* L.) (Figure 1). Weedy rice infests rice fields in almost every world region where rice is cultivated, including the United States [7], Europe [8], Latin America [9], East and South Asia [10-12], and Africa [13]. Because of its close phenotypic similarity to cultivated rice, particularly in the seedling stage, weedy rice is difficult to detect early in the growing season; if left unchecked, weedy rice infestations can reduce crop harvests by more than 80% [14]. Weedy rice has probably been present in the margins of rice fields since the inception of rice agriculture in southern Asia approximately 10,000 years ago. However, it has only become a major threat to global rice production in recent decades, due to shifts away from hand transplanting of rice seedlings (which, while labor-intensive, provides ample opportunities for hand-weeding of rice fields) toward direct-seeded mechanized farming. As described in section 2.1 below, weedy rice has evolved multiple times independently from different cultivated rice varieties. Although phenotypically diverse across its worldwide occurrences, it has convergently evolved traits associated with the agricultural weed syndrome, including highly shattering seed, strong seed dormancy, and competitive growth in agricultural fields. Identifying the genetic mechanisms underlying this convergent weediness evolution is an active area of research.

The cultivated Asian rice genome was the first reference genome published for a crop species, as well as the second angiosperm genome published after Arabidopsis [15]. Rice has become a genomic model system, particularly for cereal crops, due to its small genome size (~430 Mb) and ease of genetic modification. Since weedy rice is a direct descendant of cultivated rice, the wealth of genomic resources developed in cultivated rice can be easily transferred to the weedy rice system. We highlight some of the genetic and genomic studies

that have led up to our current understanding of weedy rice and potential avenues for future research directions.

2.0 Insights on weedy rice evolution

2.1 Insights on weedy rice origins from microsatellites and other neutral markers

One of the most basic questions about weedy rice evolution has been understanding the extent to which different weed strains around the world have evolved independently or have shared origins. Related to this question is whether weedy rice is descended directly from cultivated ancestors, or whether wild *Oryza* populations have contributed to its evolution. The earliest molecular studies of weedy rice evolution relied on neutral markers such as microsatellites to compare weed strains to cultivated and wild rice samples. A common theme to emerge from these studies is that populations of two or more genetically distinct weed strains often co-exist in a single geographical region, and that these have evolved independently from different cultivated rice varieties. This basic pattern has been detected, for example, in the United States [16], Italy [17], China [18], Korea [6], and South America [13].

Specifically in the United States, an analysis of 16 microsatellites and neutral sequence haplotype networks revealed that the two ecotypes found there SH (strawhull awnless) and BHA (blackhull awned), cluster with the genetically distinct cultivated varieties, *O. sativa indica* and *O. sativa aus*, respectively [16]. The authors noted that neither of these rice varieties was ever grown commercially in the US, suggesting that weedy rice was inadvertently imported from elsewhere. In contrast, a study of Italian weedy rice using 19 microsatellites showed that some weedy rice strains there cluster with the locally grown *O. sativa japonica* cultivars [17]. As with US weeds, however, two genetically distinct groups of weedy rice were identified. The authors of this study were able to use weed appearance records and the fact that no wild *Oryza* grows in Italy to rule out hybridization with the wild ancestor (*O. rufipogon*) as a potential cause of

origin. They were not, however, able to rule out crop-crop hybridization. From these and other studies world-wide (e.g., [19, 20]), weedy rice was shown to have evolved repeatedly and independently. Although many of these studies suggested de-domestication as the primary cause for the origin of weedy rice, none had strong evidence for that mechanism over another.

2.2 Insights from candidate genes

With the repeated independent evolution of weedy rice worldwide, many of the same phenotypic traits have convergently evolved, including highly shattering seed, dark-pigmented pericarps (and associated seed dormancy), and highly competitive growth against crop varieties. This phenotypic convergence of weedy traits raises questions on the extent to which similar underlying genetic mechanisms have been involved in this convergent phenotypic evolution. A wealth of previous work exists in cultivated rice that has characterized so-called domestication genes and causal mutations that underlie domestication traits (e.g. *sh4*, *Wx1*, *Rc*, etc.). For the wild-like traits that have emerged during weedy rice evolution, these domestication genes provide prime candidates to assess whether mutational reversions at the domestication loci underlie the phenotypic reversions occurring during de-domestication, or whether other genes or regulatory regions are responsible. In this section we compare inferences from two well studied candidate genes, *sh4* (controlling seed shattering) and *Rc* (controlling pericarp pigmentation).

The re-acquisition of seed dispersal mechanisms is one of the most important steps in escaping dependence on humans, and as such, seed-shattering is among the most ubiquitously evolved traits in weedy rice worldwide. Previous work in cultivated rice has identified several shattering-related genes, of which *sh4* is the major causative gene (reviewed in [21]). Sequencing this gene in weedy rice worldwide showed that most weedy rice strains carry the non-shattering domestication allele, suggesting the importance of other parts of the genome in

the reversion to shattering [19, 22-24]. Further quantitative trait locus (QTL) mapping of the shattering trait in two crop-weed hybrid mapping populations representing the two major US weedy rice ecotypes (SH, BHA) identified 7 QTL [25]. Interestingly, none of the QTLs identified in this weed × crop cross overlap with *sh4* or other well-characterized cultivated rice shattering loci. These findings also suggest that many different underlying genetic mechanisms can lead to convergent phenotypic evolution for quantitative traits such as shattering.

Like shattering, re-acquiring seed dormancy is an important step in the evolution of weedy rice. The gene *Rc* encodes a transcription factor that has been shown to pleiotropically control both pericarp color and seed dormancy [26]. The non-functional domestication *rc* allele results in white pericarps and a reduction in dormancy, while the functional *Rc* allele results in red pericarps and variable dormancy. Sequencing of this gene in US weedy rice revealed that these weed strains contain a functional *Rc* allele. Unlike shattering and the *sh4* domestication allele, the white pericarp *rc* allele was not universally under selection during rice domestication, and some rice landraces still have pigmented pericarps and functional *Rc* alleles [27]. Gross *et al*. [27] proposed that the presence of functional *Rc* alleles in US weeds is a legacy of these weeds having evolved from landraces that never underwent selection for white pericarps. Functional *Rc* alleles can also be found in some Asian weed strains. For those growing in Southeast Asia, the functional alleles have likely originated in part through introgression from local wild rice populations; the high frequency wild-derived *Rc* alleles in these weeds may reflect strong selection for seed dormancy [6, 24].

Candidate gene studies have furthered our understanding of weedy rice evolution by suggesting (as in the case of shattering) that many convergent phenotypic traits show no evidence of convergent molecular evolution. Conversely, genes like *Rc* have shown that similar underlying genetic mechanism can play a large role in convergent phenotypic evolution, but that the origins of the haplotypes should be investigated further.

2.3 Insights from whole genome sequencing

Recent advances in DNA sequencing technologies have made population-level genome-wide sequencing projects relatively cheap and easy to undertake. One recent weedy rice study that capitalized on this technology was Li *et al*. [5] in which 38 US weedy rice genomes (18 SH, 20 BHA) were compared to 145 previously published *Oryza* genomes including 89 cultivated rice accessions, 53 *O. rufipogon* accessions and three Chinese weedy rice accessions. Results from this study re-confirmed the origins of US and Chinese weedy rice as de-domesticated forms, provided relative divergence times, and identified regions of the weedy rice genomes that show signatures of selection (decreased π) and selective sweeps. The relative divergence times suggest that BHA weeds are older than SH and Chinese weeds, which suggests that BHA diverged from the very earliest ancient crops while SH and Chinese weeds diverged much later. This study identified 121 and 118 100-kb windows of low diversity in SH-*indica* and BHA-*aus* comparisons, respectively. Of these, only 12 windows were shared between the two comparisons. These 12 windows would be of particular interest for further study, as they are evidence of limited convergent molecular evolution. The broader implication of these data, however, is that the two US weed strains have convergently evolved phenotypes using largely different underlying genetic mechanisms. Although weedy rice accessions from more places around the world should be sequenced and analyzed in a similar manner, the results from this study provide a useful foundation for future comparative studies.

3.1 Avenues of future work

The next steps in the study of weedy rice evolution follow easily from the work described in section 2. Each weedy ecotype world-wide should be probed for independent origins and placed in a framework describing where and how many independent origins have occurred. Molecular and phenotypic evidence should be used to pin down relative divergence times

similar to Li *et al*. [5]. With origins and relative divergence times we can begin to answer important questions related to the circumstances leading to the evolution of weedy rice.

Advanced techniques can also be used to continue identifying the genetic basis of the agricultural weed syndrome traits. Connecting phenotype to genotype is not easy, but combining transcriptome, methylome and conventional QTL techniques should be used in a broad range of weedy rice ecotypes to begin to understand the genetic basis of important weedy traits.

Another obvious avenue of future work is more whole genome studies. Li *et al*. [5] focused largely on US weedy rice, but more accessions from world-wide occurrences can be collected and analyzed in a similar manner. Additionally, weedy rice is evolving in a rapidly changing agricultural environment. The introduction of both hybrid cultivated rice and herbicide resistant cultivars is changing how weedy rice interacts with its environment and thus, how its genome is evolving [28, 29]. Whole genomes of post-introduction weedy rice should be sequenced and used to evaluate how weedy rice is adapting to this new environment. Studies characterizing these rapidly evolving genomes could provide important insights not only for understanding the genetic underpinnings of weed adaptation, but also for devising more effective weed control strategies.

Acknowledgements

The authors thank members of the Olsen lab group for helpful comments on the manuscript. M.J.W. is supported through a National Science Foundation Graduate Research Fellowship (grant number 2017221153). Weedy rice research in the Olsen lab has been supported through the National Science Foundation Plant Genome Research Program (grant number IOS-1032023).

Works Cited

[1] N. Alexandratos, J. Bruinsma, World agriculture towards 2030/2050, Land Use Policy. 20 (2012) 375. doi:10.1016/S0264-8377(03)00047-4.

[2] S.H. Barrett, Crop Mimicry in Weeds. (1983) 37 (3):255-82.

[3] C.N. Stewart, Becoming weeds, Nat. Genet. 49 (2017) 654–655. doi:10.1038/ng.3851.

[4] C.C. Vigueira, W. Li, K.M. Olsen, The role of Bh4 in parallel evolution of hull colour in domesticated and weedy rice, J. Evol. Biol. 26 (2013) 1738–1749. doi:10.1111/jeb.12171.

[5] L.-F. Li, Y.-L. Li, Y. Jia, A.L. Caicedo, K.M. Olsen, Signatures of adaptation in the weedy rice genome, Nat. Genet. 49 (2017) 811–814. doi:10.1038/ng.3825.

[6] C.C. Vigueira, X. Qi, B.-K. Song, L.-F. Li, A.L. Caicedo, Y. Jia, K.M. Olsen, Call of the wild rice: *Oryza rufipogon* shapes weedy rice evolution in Southeast Asia, Evol. Appl. (2017) 1–12. doi:10.1111/eva.12581.

[7] R.F.W. Allston, The Rice Plant, DeBow's Rev (1846) 320-356.

[8] G. Biroli, Tipografia Giovanni Silvestri, Del Riso (1807) Milan, Italy.

[9] A. Merotto, I.C.G.R. Goulart, A.L. Nunes, A. Kalsing, C. Markus, V.G. Menezes, A.E. Wander, Evolutionary and social consequences of introgression of nontransgenic herbicide resistance from rice to weedy rice in Brazil, Evol. Appl. 9 (2016) 837–846. doi:10.1111/eva.12387.

[10] J. Sun, Q. Qian, D.R. Ma, Z.J. Xu, D. Liu, H.B. Du, W.F. Chen, Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China, New Phytol. 197 (2013) 290–299. doi:10.1111/nph.12012.

[11] J. Qiu, J. Zhu, F. Fu, C.Y. Ye, W. Wang, L. Mao, Z. Lin, L. Chen, H. Zhang, L. Guo, S. Qiang, Y. Lu, L. Fan, Genome re-sequencing suggested a weedy rice origin from domesticated indica-japonica hybridization: a case study from southern China, Planta. 240 (2014) 1353–1363. doi:10.1007/s00425-014-2159-2.

[12] Q. He, K.W. Kim, Y.J. Park, Population genomics identifies the origin and signatures of selection of Korean weedy rice, Plant Biotechnol. J. 15 (2017) 357–366. doi:10.1111/pbi.12630.

[13] M.T. Federici, D. Vaughan, N. Tomooka, A. Kaga, X.W. Wang, K. Doi, M. Francis, G. Zorrilla, N. Saldain, Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers, Electron. J. Biotechnol. 4 (2001) 42–57. doi:10.2225/vol4-issue3-fulltext-3.

[14] L.E. Estorninos, D.R. Gealy, E.E. Gbur, R.E. Talbert, M.R. McClelland, Rice and red rice interference. II. Rice response to population densities of three red rice (*Oryza sativa*) ecotypes, Weed Sci. 53 (2005) 683–689. doi:10.1614/WS-04-040R1.1.

[15] S.A. Goff, D. Ricke, T.-H. Lan, G. Presting, *et al*., A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), Science 296 (2002) 92–100. doi:10.1126/science.1068037.

[16] J.P. Londo, B.A. Schaal, Origins and population genetics of weedy red rice in the USA, Mol. Ecol. 16 (2007) 4523–4535. doi:10.1111/j.1365-294X.2007.03489.x.

[17] A. Grimm, S. Fogliatto, P. Nick, A. Ferrero, F. Vidotto, Microsatellite markers reveal multiple origins for Italian weedy rice, Ecol. Evol. 3 (2013) 4786–4798. doi:10.1002/ece3.848.

[18] Q. Cao, B.R. Lu, H. Xia, J. Rong, F. Sala, A. Spada, F. Grassi, Genetic diversity and origin of weedy rice (*Oryza sativa* f. *spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers, Ann. Bot. 98 (2006) 1241–1252. doi:10.1093/aob/mcl210.

[19] M. Akasaka, J. Ushiki, H. Iwata, R. Ishikawa, T. Ishii, Genetic relationships and diversity of weedy rice (*Oryza sativa* L.) and cultivated rice varieties in Okayama Prefecture, Japan, Breed. Sci. 59 (2009) 401–409. doi:10.1270/jsbbs.59.401.

[20] R. Ishikawa, N. Toki, K. Imai, Y.I. Sato, H. Yamagishi, Y. Shimamoto, K. Ueno, H. Morishima, T. Sato, Origin of weedy rice grown in Bhutan and the force of genetic diversity, Genet. Resour. Crop Evol. 52 (2005) 395–403. doi:10.1007/s10722-005-2257-x.

[21] L.F. Li, K.M. Olsen, To Have and to Hold: Selection for Seed and Fruit Retention During Crop Domestication, Curr. Top. Dev. Biol. 119 (2016) 63–109. doi:10.1016/bs.ctdb.2016.02.002.

[22] C.S. Thurber, M. Reagon, B.L. Gross, K.M. Olsen, Y. Jia, A.L. Caicedo, Molecular evolution of shattering loci in U.S. weedy rice, Mol. Ecol. 19 (2010) 3271–3284. doi:10.1111/j.1365-294X.2010.04708.x.

[23] Y. Zhu, N.C. Ellstrand, B.R. Lu, Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication, Ecol. Evol. 2 (2012) 2106–2113. doi:10.1002/ece3.318.

[24] B.K. Song, T.S. Chuah, S.M. Tam, K.M. Olsen, Malaysian weedy rice shows its true stripes: Wild *Oryza* and elite rice cultivars shape agricultural weed evolution in Southeast Asia, Mol. Ecol. 23 (2014) 5003–5017. doi:10.1111/mec.12922.

[25] X. Qi, Y. Liu, C.C. Vigueira, N.D. Young, A.L. Caicedo, Y. Jia, D.R. Gealy, K.M. Olsen, More than one way to evolve a weed: Parallel evolution of US weedy rice through independent genetic mechanisms, Mol. Ecol. 24 (2015) 3329–3344. doi:10.1111/mec.13256.

[26] X.Y. Gu, M.E. Foley, D.P. Horvath, J. V. Anderson, J. Feng, L. Zhang, C.R. Mowry, H. Ye, J.C. Suttle, K.I. Kadowaki, Z. Chen, Association between seed dormancy and pericarp color

is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice, Genetics. 189 (2011) 1515–1524. doi:10.1534/genetics.111.131169.

[27] B.L. Gross, M. Reagon, S.C. Hsu, A.L. Caicedo, Y. Jia, K.M. Olsen, Seeing red: The origin of grain pigmentation in US weedy rice, Mol. Ecol. 19 (2010) 3380–3393. doi:10.1111/j.1365-294X.2010.04707.x.

[28] V. Singh, N.R. Burgos, S. Singh, D.R. Gealy, E.E. Gbur, A.L. Caicedo, Impact of volunteer rice infestation on yield and grain quality of rice, Pest Manag. Sci. 73 (2017) 604–615. doi:10.1002/ps.4343.

[29] V. Singh, S. Singh, H. Black, V. Boyett, S. Basu, D. Gealy, E. Gbur, A. Pereira, R.C. Scott, A. Caicedo, N.R. Burgos, Introgression of Clearfield™ rice crop traits into weedy red rice outcrosses, F. Crop. Res. 207 (2017) 13–23. doi:10.1016/j.fcr.2017.03.004.

Figure

Figure 1. Seeds of weedy rice sampled from rice fields in the southern United States.

Photo credit: Kenneth M. Olsen

CHAPTER TWO

Convergent evolution of root system architecture in two independently evolved lineages of weedy rice

Abstract

Root system architecture (RSA) is a critical aspect of plant growth and competitive ability. Here we used two independently-evolved strains of weedy rice, a de-domesticated form of rice, to study the evolution of weed-associated RSA traits and the extent to which they evolve through shared or different genetic mechanisms. We characterized 98 two- and three-dimensional RSA traits in 671 plants representing parents and descendants of two recombinant inbred line populations derived from two weed × crop crosses. A random forest machine learning model was used to assess the degree to which root traits can predict genotype and the most diagnostic traits for doing so. We used QTL mapping to compare genetic architecture between the weed strains. The two weeds were distinguishable from the crop in similar and predictable ways, suggesting independent evolution of a 'weedy' RSA phenotype. Notably, comparative QTL mapping revealed little evidence for shared underlying genetic mechanisms. Our findings suggest that despite the double bottlenecks of domestication and de-domestication, weedy rice nonetheless shows genetic flexibility in the repeated evolution of weedy RSA traits. Whereas the root growth of cultivated rice may facilitate interactions among neighboring plants, the weedy rice phenotype may minimize below-ground contact as a competitive strategy.

Introduction

An active and important question in evolutionary biology is the extent to which the genetic basis of adaptation is flexible or constrained by phylogeny (Orr, 2005; Protas *et al.*, 2006; Weller *et al.*, 2012; Ng & Smith, 2016). One of the most common ways this question has been examined is through comparisons of separate populations or species that have independently evolved the same phenotype under similar environmental conditions and selective pressures. Evidence for this type of repeated phenotypic evolution is abundant across all kingdoms of life (e.g., Fong *et al.*, 2005; Losos, 2011; Pichersky & Lewinsohn, 2011), but

what is less well understood is the extent to which it results from selection acting on the same genes and/or developmental pathways (referred to herein as parallel evolution or parallelism) (Weller *et al.*, 2012) or different genetic and developmental processes (referred to herein as convergent evolution or convergence) (Larson, 2014; Ng & Smith, 2016). In general, parallelism is predicted to be most likely for lineages that are phylogenetically closely related and hence genetically similar (Losos, 2011), as well as for lineages that are constrained by low genetic diversity (limiting the pool of potentially advantageous alleles) (Orr, 2005), and traits that are controlled by genes with highly specialized functions (Pfenning *et al.*, 2014).

In plants, repeated evolution has been described for a diverse range of traits, including photoperiod response (reviewed in Lenser & Theißen, 2013), abiotic stress adaptation (Lyu *et al.*, 2018), chemical defense metabolites (Takos *et al.*, 2011), floral pigmentation and morphology (Ng & Smith, 2016), mating system (Fishman *et al.*, 2015), and other agronomic traits in crop species (Sang, 2009). Many of these studies have found evidence for evolutionary convergence. However, as most such studies have compared species across different families or higher taxonomic groups, their findings may be inherently biased towards observations of convergence over parallelism. Far fewer studies in plants have examined instances of repeated evolution within species. In addition, virtually all of these studies have focused on the above-ground half of the plant phenotype. As a result, little-to-nothing is known about the extent to which adaptive changes in root growth and development may be evolutionarily constrained to a greater or lesser extent than above-ground growth.

Weedy rice (*Oryza sativa* f. *spontanea*) offers an attractive system for overcoming these gaps in our understanding of repeated phenotypic evolution in plants. This agricultural weed is a feral (de-domesticated) descendant of the genomic model crop species rice (*O. sativa*). Weedy rice has evolved multiple times independently from different domesticated rice varieties around the world (Federici *et al.*, 2001; Cao *et al.*, 2006a; Londo & Schaal, 2007; Grimm *et al.*,

2013; Wedger & Olsen, 2018). The process of weedy rice evolution is associated with the repeated emergence of suites of adaptations that distinguish the weed from its domesticated ancestor and allow it to aggressively outcompete rice in the field. Weedy rice adaptations include highly shattering seeds (Qi *et al.*, 2015), strong seed dormancy (Gross *et al.*, 2010b), herbicide resistance (Singh *et al.*, 2017c), and the ability to outcompete cultivated rice for light and soil nutrients (Burgos *et al.*, 2006; Xu *et al.*, 2018). Infestations of as few as eight weedy rice plants/m^2 can reduce rice yields by almost two-thirds (Xu *et al.*, 2018). The weed's competitive ability has been proposed to be directly related to the pattern of root growth (Burgos *et al.*, 2006), although this hypothesis has not been directly tested. As independently derived, conspecific relatives of domesticated rice, weedy rice strains are highly amenable for directly comparing the genetic basis of repeated weediness evolution.

The genetic and genomic differences that distinguish different weedy rice varieties are very well characterized (Londo & Schaal, 2007; Vigueira *et al.*, 2013a; Li *et al.*, 2017), and this information has provided a basis for recent studies that are elucidating mechanisms of repeated evolution of weedy rice (Qi *et al.*, 2015; Li *et al.*, 2017). In the United States, two genetically distinct weed strains predominate in the major rice growing region of the southern Mississippi valley. These two strains are distinguishable based on grain characteristics and are referred to as black-hull awned (BHA) and strawhull awnless (SH) weedy rice. BHA weeds are feral descendants of cultivated *aus* rice varieties, while SH forms are descended from cultivated *indica* rice varieties. Since neither *aus* nor *indica* varieties of rice were ever commercially cultivated in the US, the BHA and SH weed strains are presumed to have originated in southern Asia, where *aus* and *indica* varieties are traditionally grown. Later introductions into the US likely occurred through weed-contaminated seed grain imports (Londo & Schaal, 2007). In the 150-year history of BHA and SH weed presence in the US, minimal amounts of hybridization have been detected between the weed strains due to high selfing rates (Singh *et al.*, 2017a).

Several factors support the prediction that parallelism rather than convergence might be expected to underlie the emergence of weediness traits in the BHA and SH weeds. First, like most annual crop species, *O. sativa* underwent a genome-wide loss of genetic diversity during the process of domestication. This domestication bottleneck would have left a more limited pool of genetic diversity as a starting point for weed evolution compared to evolution in a wild species. In addition, outcrossing rates are very low in both cultivated and weedy rice (typically <1%) (Cao *et al.*, 2006a; Gealy *et al.*, 2009), which would further limit opportunities to enhance the genetic diversity of evolving weed strains. Consistent with these factors, genetic diversity in both BHA and SH strains is exceedingly low compared to their direct domesticated ancestors and to wild rice (Reagon *et al.*, 2010; Li *et al.*, 2017). Taken together with the close phylogenetic relationships among all weedy and cultivated rice populations, parallelism would thus seem to be the most likely mechanism by which weedy traits would emerge. Interestingly, however, this is not the primary mechanism that has been observed in weedy rice studies to date (Mispan *et al.*, 2013; Thurber *et al.*, 2013; Qi *et al.*, 2015). These studies, all of which have examined above-ground traits, have revealed different genetic architectures for several weediness traits in the two US weed strains.

Below-ground root growth and spatial organization of root systems can be described in terms of root system architecture (RSA) (Topp *et al.*, 2013). Despite its critical role in determining efficiency of soil nutrient and water uptake, as well as neighbor-to-neighbor communication and levels of plant competition, RSA is far less characterized than above-ground aspects of plant growth (Casper & Jackson, 1997; Topp *et al.*, 2016). To the extent that the genetics of RSA have been examined, this has mostly been at the level of QTL mapping, where many loci have been identified in crop varieties (Uga *et al.*, 2011; Topp *et al.*, 2013). Only two RSA genes have been cloned and functionally characterized, both in rice: *DEEPER ROOTING*

1 (*Dro1*) (Uga *et al.*, 2013) and *Phosphorus-Starvation Tolerance 1* (*PSTOL1*) (Gamuyao *et al.*, 2012). To our knowledge, no study has investigated RSA or its genetic basis in weedy rice.

In this study we employed comparative QTL mapping in two advanced-generation recombinant inbred line (RIL) mapping populations, derived from a BHA × *indica* cross and a SH × *indica* cross with the same *indica* parent, to examine the genetic basis of weedy rice RSA and the extent to which it has evolved through parallelism or convergence. We investigated three questions: 1) Are there RSA differences between the BHA, SH, and *indica* parents? 2) If so, are any of those differences shared by both weed ecotypes in a pattern suggesting repeated phenotypic evolution? And 3) To the extent that there are shared weed-specific RSA traits, does their genetic architecture indicate that these are controlled by similar or different underlying genetic mechanisms?

Materials and Methods

Plant materials

Seeds for all accessions from two weed × crop recombinant inbred line (RIL) mapping populations were obtained from the USDA-ARS Dale Bumpers National Rice Research Center (Stuttgart, AR), where they were advanced to the F_8 generation through single seed descent. The mapping populations were initiated in 2007-2009 at the University of Massachusetts - Amherst by crossing the Taiwanese *indica* rice variety Dee Geo Woo Gen (DGWG; PI 653419) with each of two US weedy rice ecotypes (Thurber *et al.*, 2013). The crop genotype used in our study is best known as the original source of the *sd1* semi-dwarfism allele that gained fame with the improved rice cultivars of the Green Revolution (Spielmeyer *et al.*, 2002). The first cross (source of the B mapping population below) was produced by crossing DGWG with a black-hull awned accession (MS-1996-6; GSOR 303535). The second cross (source of the S population

below) was produced by crossing DGWG with a straw-hull awnless accession (AR-2000-1135-01; GSOR 303286). Seeds for a total of 224 and 175 RILs from the B and S populations were obtained through the USDA-GRIN germplasm collection (https://www.ars-grin.gov).

Phenotyping

On a weekly basis over a two-year period in 2016-2017, replicates of the parental lines and RIL accessions were grown and phenotyped for below-ground root architecture using a modified root imaging protocol (Topp *et al.*, 2013). Two replicates per parent genotype were grown each week to serve as controls. Seeds were de-hulled and surface-sterilized with a 10-minute bath of 35% hydrogen peroxide followed by three washes with distilled and deionized water. Sterilization prevented fungal growth which would inhibit efficient imaging as described below. Sterilized seeds were placed on petri dishes with Yoshida's nutrient solution containing 0.5% Gelzan gellan gum. Seeds were then placed in a dark incubation chamber at 29 °C for three days to stimulate germination. Up to two healthy seeds per genotype were chosen for transplanting based on germination success, lack of microbial contamination, and, when applicable, maximal distance from the nearest contaminated seedling. Germinated seeds were transplanted into glass 2 L ungraduated cylinders with 1 L sterilized Yoshida's nutrient solution containing 0.25% Gelzan gellan gum using flame-sterilized forceps and a sterile pipette (one seedling per cylinder). Transplanted seeds were assigned a unique identifier and left at room temperature and ambient light for 12 hours to overcome transplanting shock. Plants were then moved into a growth chamber set for long day photoperiod (16-hour days at 28˚C and 600 µmol of light: 8-hour nights at 24˚C and 0 µmol of light) and left to grow for 10 days. On day 13 after germination, plants were removed from the growth chamber and imaged using a custom rig as described below (see also Supporting Information Fig. **S1**).

To facilitate imaging, cylinders with plants were placed individually on a turntable in a glass box filled with water (to correct for light diffraction) and backlit with a uniform green light panel. The cylinders were then rotated 360 degrees on the turntable, and images were taken every five degrees by a computer-controlled camera, resulting in 72 sequential images per plant. Plants that had become contaminated by microbial growth during the 10-day growth period were not imaged; these represented approximately 10% of all seeds planted. Plants that failed to continue growing after transplanting were also not imaged; these represented approximately 2% of all seeds planted. Wet shoot and root weights were taken immediately after imaging, and dry shoot and root weights were taken after sufficient drying time. Up to 40 plants were imaged per week.

Images were analyzed using a modified *GiaRoots* pipeline (Galkovskyi *et al.*, 2012; Topp *et al.*, 2013) which includes scaling, cropping, and thresholding the images to convert the greyscale image to a set of binary images. These binary images were then analyzed by *GiaRoots2D* to measure two-dimensional traits. A three-dimensional reconstruction of the root was produced using the *RootworkPerspective* software. The 3D reconstruction was then analyzed by *GiaRoots3D* to measure three-dimensional RSA traits. The reconstruction was then further analyzed by *DynamicRoots*, which can more finely measure traits from distinct root classes (e.g., primary, first-order lateral, second-order lateral). In total, we obtained phenotypic measurements for 98 RSA traits, many of which are strongly correlated (Supporting Information Fig. **S2**).

Phenotypic analysis and QTL mapping

To test for significant phenotypic differences between the crop and weed parental genotypes (DGWG, BHA, and SH), their phenotypic values were compared using a single factor ANOVA followed by a post-hoc Tukey HSD test in R. Equal variance and normality

assumptions were tested in R using Levene's test and the Shapiro-Wilk test respectively. When assumptions were violated, results from ANOVA were validated in R using Welch's one-way test and the Kruskal-Wallis rank-sum test. Traits that failed to show significant differences in the ANOVA, Welch's, or Kruskal-Wallis tests were considered not significantly different. Traits were binned into six patterns of significance based on a Tukey HSD significance threshold of $\alpha = 0.05$. The six bins were as follows for pairwise differences between the BHA parent (b), the *indica* crop parent (c) and the SH parent (s): $b = c = s$ (i.e., no significant differences between any lines); $b = c \neq s$ (SH different from other two parents); $b = s \neq c$ (DGWG different from other two parents); $c = s \neq b$ (BHA different from the other two parents); $b \neq c \neq s$ (all parents different from each other); and a catch-all bin for any other patterns (e.g., $b = c = s \neq b$ and other nontransitive relationships that reflected differences in confidence interval widths). The $c = s \neq b$ bin corresponds to a pattern predicted based on phylogenetic relationships alone, as *indica* rice is the putative direct ancestor of SH weedy rice whereas BHA is less closely related to these genotypes. The $b = s \neq c$ bin would be consistent with repeated phenotypic evolution of shared root phenotypes in the two independently-evolved weed strains that distinguish them from the crop. This analysis allowed us to begin describing the suite of traits that together characterize the below-ground weedy rice phenotype.

The parental phenotypes were further analyzed using the *r/randomForest* machine learning package in *R*. This analysis was performed to determine 1) if the parental genotypes were reliably distinguishable from each other (as opposed to the ANOVAs above, which assessed whether the weeds were distinguishable from the crop), and 2) if so, which traits were the most diagnostic in differentiating the parental genotypes. The random forest model built 3000 trees and was trained on two-thirds of the data. The resulting model was applied to the remaining one-third of the data to assess predictive success.

Using modified linkage maps from previously published B and S mapping populations (Qi *et al.*, 2015; D.M Goad, unpublished), QTL mapping of root phenotypes was performed in *r/qtl* using the *scanone* function and the Haley-Knott method for a balance in speed and performance (Haley & Knott, 1992; Broman *et al.*, 2003). Physical positions were determined relative to the MSU v7.0 rice genome (http://rice.plantbiology.msu.edu). LOD thresholds were calculated on a trait-by-trait basis using 10,000 permutations. LOD confidence intervals represent a drop of 1.5 LOD on either side of the maximum value. Mapping was performed using 11,853 and 4,733 single nucleotide polymorphism (SNP) markers from the F_5 generations for the B and S populations, respectively. These markers were obtained in an earlier generation of the RILs (F_5) and published in an earlier study (Qi *et al.*, 2015). QTL positions were visualized using the *r/qtlTools* package in R (Delaneau *et al.*, 2017).

Results

We imaged 671 rice plants for 98 2-dimensional and 3-dimensional RSA traits. The phenotyped plants included 30 replicates of the BHA parent, 29 replicates of the crop parent, 33 replicates of the SH parent, 237 plants from the B mapping population (BHA × DGWG RILs), and 342 plants from the S mapping population (SH × DGWG RILs). In the B population 84 RILs were phenotyped twice, and 23 RILs were phenotyped three times, yielding 107 RILs with two or more replicates. In the S population, 63 RILs were phenotyped twice, and 72 were phenotyped three times; this yielded 135 RILs represented by two or more replicates. Only RILs that were phenotyped at least twice were used in further analysis. It should be noted that the limited number of RILs analyzed in each population could potentially bias our results toward the identification of a few large-effect QTL, leading to an underestimate of the total number of small -effect loci.

Parental line assessment

For the 98 RSA traits where we tested for differences between the three parental lines, 62 of them showed no significant differences (corresponding to a pattern of b = c = s, where b is black-hull awned parent, c is crop parent, and s is straw-hull awnless parent). Among those with significant differences, eight of the traits fit the pattern that would be predicted if phylogenetic history were the primary determinant of phenotypic differences (with the closely related SH and DGWG genotypes not significantly different from each other, but both significantly different from the evolutionarily-diverged BHA genotype; i.e., c = s ≠ b). For only one trait was the opposite pattern observed (no significant difference between the BHA and crop genotype, but SH significantly different from those accessions; b = c ≠ s). Notably, 20 traits showed significant phenotypic differences in the pattern that would be predicted if the weedy rice strains had independently evolved shared root morphologies (no significant differences between the SH and BHA parents but significant difference between the weeds and DGWG; b = s ≠ c). We refer to these as "weed-specific RSA traits" below.

Weed-specific RSA traits. Because many of the root traits are highly correlated (e.g., mean root depth and median root depth (Supporting Information Fig. **S2**)) we condensed the 20 putative weed-specific RSA traits into eight summary descriptor traits: root width-depth ratio, average root width, maximum number of roots, width of the root system, specific root length, mean root depth, mean root tortuosity (i.e., degree to which roots are curved), and mean root-soil angle (i.e., degree to which roots grow horizontally or vertically). From these descriptor traits, the crop (DGWG) root system can be summarized as being different from the weeds in the following ways: it is wider and higher in the soil, with individual roots that are thinner, longer, more abundant, more curved, and at a lower angle to the soil (Fig. **1**).

To assess whether the parental lines could be reliably differentiated using root traits, a random forest machine learning approach was undertaken in R using the *r/randomForest* package. While random forest machine learning is usually used to predict unknowns, we used it here to reduce the dimensions of our data (Ramírez *et al.*, 2010). We found that the three parents were correctly identified approximately 60% of the time when each data type was analyzed separately (*GiaRoots2D*, *GiaRoots3D*, and *DynamicRoots*) (Supporting Information Table **S1**). By comparison, assignment to the correct parental genotype by chance alone would be expected 33.3% of the time. When all three data types were combined, the strains were correctly identified approximately 70% of the time. Thus, the analysis of root traits approximately doubles the probability of correct assignment compared to random chance alone.

The *r/randomForest* package also generates a rank order of diagnostic traits. These traits can be considered the most important for distinguishing root phenotypes of the crop vs. weed parents, although they should not be interpreted as necessarily related to the biology of weediness. For the 2D dataset, the most diagnostic traits were *Solidity (2D)* (density of the root system), and *Maximum Width (2D)*. For the 3D datasets, the most diagnostic traits are *Width-depth Ratio*, and *Median Lateral Root-soil* angle for *GiaRoots3D* and *DynamicRoots* traits, respectively (Fig. **2**). Although the results presented in Fig. **2** represent a typical run, highly correlated traits shifted in relative importance between individual runs. Regardless of the exact trait at the top of the list, the biological interpretation is robust between runs. All three datasets place the most importance on traits related to width and exploration.

QTL analysis

Out of 98 root phenotypes that were evaluated in the F_8 generation of the two weed × crop mapping populations, we identified a total of 65 significant QTLs distributed across 43 root traits (Supporting Information Table **S2**). In the S population (SH × *indica*), 36 QTLs were

identified, with 22 traits having one QTL apiece, and six traits mapping to more than one genomic location (up to three QTLs). In the B (BHA × *indica*) population, 29 QTLs were identified, with 19 traits mapping to one QTL apiece, and five traits mapping to two QTLs. Of the 43 traits with significant QTLs, 10 traits mapped to both populations; these traits fall into four broad trait categories (Table **1**). We describe these shared mapped traits below.

Root depth. Both *Depth (2D)* and *Major Ellipse Axis (2D)* are measures of rooting depth. *Depth (2D)* is the straight-line distance between the soil-line and the tip of the deepest root at 90 degrees from the soil-line. *Major Ellipse Axis (2D)* is the distance between the two major vertices of the smallest possible ellipse encompassing the entire root system. If the root mass is symmetrically distributed along the depth axis, this measurement is very similar to *Depth (2D)*. If not, it captures differences in root mass distribution. In the B population, both traits map to the same position in the middle of Chromosome 4 (Table **1**; Fig. **3a**), while in the S population both traits map to the same position in the middle of Chromosome 8 (Table **1**; Fig. **3b**). Both weed parents are on average deeper than the crop. Interestingly, however, all the significant QTLs for root depth have increased effects conferred by the crop allele, ranging from 10.8-16.5% effect and explaining 8.4-14.9 % of the variation. This pattern suggests that there may be many small effect loci in the weeds that are undetected by this study and that collectively cause the weeds to grow deeper roots than the crop parent.

Root system width. *Minor Ellipse Axis (2D)* and *Maximum Network Width (3D)* are both measures of the width of the root system. *Minor Ellipse Axis (2D)* is the distance between the two minor vertices of the smallest possible ellipse encompassing the entire root system. If the root mass is symmetrically distributed along the depth axis, this measurement is very similar to the *Maximum Network Width (2D)*. If not, it provides an alternative measure of differences in root mass distribution distinct from *Major Ellipse Axis (2D)*. *Maximum Network Width (3D)* is the widest span of the root system in a plane parallel to the soil line. In the B population, two width-

associated QTLs were identified, including a *Minor Ellipse Axis (2D)* QTL at the top of Chromosome 6 and a *Max Network Width (3D)* QTL at the top of Chromosome 9 (Table **1**; Fig. **3a**). In the S population, we also identified two QTLs. The first QTL is mapped with both trait measures to the middle of Chromosome 4, while the second maps with *Maximum Network Width (3D)* at the top of Chromosome 5 (Table **1**; Fig. **3b**). The crop parent has a wider root system than both weed parents. For three of the five significant QTLs, the crop alleles confer increased width, ranging from 10.6-15.2% increased effects and explaining 7.4-13.3% of the phenotypic variation. For the other two QTLs, the weed alleles confer increased width, ranging from a 10-16.6% increase and explaining 6.6-13.4% of the phenotypic variation.

Exploratory space. *Perimeter (2D)*, *Network Convex Area (2D)*, *Convex Hull Volume (3D)*, and *Solidity (3D)* are all measures to describe the volume of soil media explored by a root system. These measures approximate the extent to which the roots reach into their surroundings. *Perimeter (2D)* is calculated as the number of root pixels connected to a background pixel – an estimate of absorptive surface of the root system. *Network Convex Area (2D)* is calculated by drawing the smallest convex polygon around the root system and calculating the area inside the polygon. *Convex Hull Volume (3D)* is calculated in much the same way, but in three dimensions. *Solidity (3D)* can be thought of as the density of the root system and is calculated by dividing *Total Root Volume (3D)* by *Convex Hull Volume (3D)*. A larger solidity would be denser and thus less exploratory. The latter three traits are correlated (Supporting Information Fig. **S2**).

For all four exploratory space measures, one QTL was identified in the B population at the top of Chromosome 6, and one QTL was identified in the S population in the middle of Chromosome 4 (Table **1**; Fig. **3b**). For *Convex Hull Volume (3D)* and *Network Convex Area (2D)*, another B population QTL was identified in the middle of Chromosome 4, but statistically only the *Convex Hull Volume (3D)* QTLs overlap between the S and B population. Interestingly,

convex hull volume is the only trait of the 10 RSA traits considered in both mapping populations that maps to overlapping genomic regions in the two populations (Table **1**; Fig. **3a**). Thus, most RSA QTLs for weedy rice are not shared between the BHA and SH ecotypes.

There was high variability in these exploratory space traits, but in general the crop parent had higher exploration than either weed ecotype. Effect directions are similarly variable, with the same QTL increasing exploration in both the BHA weed and crop depending on the particular exploratory space measure calculated. This variability in effect directions is likely due to the allometric relationships between the traits which can create non-linear relationships as a function of dimensionality. Five of the other six QTLs have increased effects in the crop, conferring a 12.6-40% change in phenotype and explaining 8.7-14.3% of the phenotypic variation.

Root-soil angle: *Mean Root-soil Angle (3D)* and *Mean Lateral Root-soil Angle (3D)* describe the angle of roots relative to the soil surface. A larger angle would result in a deeper, narrower root system. Both root-soil angle traits map to the top of Chromosome 12 in the B population (Table **1**; Fig. **3a**), whereas they map to the bottom of Chromosome 2 in the S population (Table **1**; Fig. **3b**). The crop parent had a lower root-soil angle than the weeds. For both significant QTLs, the weed parent alleles confer increased effects, leading to a 9.8-23.4% phenotypic change and explaining 9-13.7% of the phenotypic variation.

Discussion

Despite its critical importance for traits such as nutrient uptake and competition for soil resources, root system architecture (RSA) remains one of the least well characterized aspects of plant growth morphology. Here we have used an integrated root imaging platform to precisely characterize RSA traits in a cultivated rice genotype and in two independently-evolved

ecotypes of weedy rice, feral descendants of the crop that aggressively outcompete it in the field. We have used this system to examine the extent to which weed-associated RSA traits have evolved in a pattern consistent with repeated phenotypic evolution, and whether comparative QTL mapping suggests that parallelism or convergence is more likely to have played a role in this process. We find clear evidence for repeated phenotypic evolution below ground, with the SH and BHA weedy rice parents independently evolving a shared suite of RSA traits (Fig. **1**). Interestingly, despite the close phylogenetic relationship of the two weed ecotypes, we find very little evidence that this has occurred through shared genetic mechanisms. Of the 10 weed-specific RSA traits with significant QTLs in both mapping populations (Table **1**), only a single trait (*Convex Hull Volume (3D)*) mapped to overlapping genomic positions in both sets of RILs (Fig. **3a, b**). Below we discuss these results in the context of RSA variation, repeated phenotypic evolution in plants, and potential implications for combatting weedy rice in crop fields.

Repeated phenotypic evolution

Our analyses reveal clear evidence of weed-specific RSA traits. Compared to the crop genotype, the two major US weedy rice ecotypes are characterized by root systems that are deeper, thinner, straighter, and less spread out, with fewer individual roots that are thicker and steeper relative to the soil line (Fig. **1**). These patterns suggest independent evolution of shared RSA traits in these weedy rice lineages. Since both weed ecotypes were being compared to the same crop variety, DGWG, one potential contributor to these patterns could be the occurrence of root traits that are unique to the crop accession. If this were the case, the traits that we are interpreting as independently-evolved in the weeds would in fact be DGWG-specific traits. Our QTL mapping results do not support this possibility, however, since we do not find shared QTL in the two mapping populations; thus, the determining genetic factors cannot be attributed to the shared crop parent. Nonetheless, our understanding of RSA trait evolution in weedy rice would

clearly benefit from follow-up studies with expanded sampling of multiple weed and crop genotypes at multiple life stages to assess the generalizability of our results.

In this study we used a clear Gelzan-based growth medium combined with a shadow imaging technique (Fig. **S1**). With this imaging technique, any amount of microbial growth in the medium would cast a shadow on the camera and alter our measurements. Therefore, all of our RSA traits are based on growth in sterile media. There is no doubt that microbial communities are important for root growth (Rolli *et al.*, 2015; Saleem *et al.*, 2016). Indeed, anecdotally, we observed that plants heavily contaminated by microbial growth (and thus not imaged for this study) had visibly different growth patterns. It is an unfortunate constraint of this root imaging technique that microbial growth cannot be considered. Follow-up studies using the 2D mature root-crown imaging software DIRT (York *et al.*, 2014), or advanced imaging techniques using x-ray computed tomography (X-ray CT) could potentially provide additional insights on the impact of microbial communities on RSA traits.

Above-ground traits have been extensively described for the domestication syndrome in crop species (Morrell *et al.*, 2012; Vigueira *et al.*, 2013b; Li & Olsen, 2016), as well as for the agricultural weed syndrome in their weedy relatives (Zhu *et al.*, 2012; Subudhi *et al.*, 2014; Qi *et al.*, 2015). In contrast, very little is known about what constitutes domestication and weediness traits for root system architecture. It is thus difficult to assess whether the repeated phenotypic evolution we observe for RSA traits in weedy rice is typical of other agricultural species. When considered in comparison to above-ground phenotypes in weedy rice, the independent evolution of RSA traits is consistent with the extensive phenotypic convergence observed in previous studies (Zhu *et al.*, 2012; Qi *et al.*, 2015). At the genetic level, the repeated detection of different underlying QTL in comparative studies of weedy rice ecotypes suggests multiple instances of independent evolution, including for emergence date, shattering, and pericarp color

(Qi *et al.*, 2015). In this respect, our RSA study parallels previous findings for above-ground traits in weedy rice.

Given the current lack of information on plant RSA traits, it is also difficult to assess the biological significance of the RSA differences we have observed in cultivated and weedy rice. Here we find that DGWG roots are more abundant and more exploratory (although not more massive) than the weedy counterparts (Fig. **1**). At face value this finding seems counterintuitive for a crop phenotype, since a more compact root system could reduce neighbor-to-neighbor competition for soil nutrients. Indeed, for above-ground traits, much of the progress that plant breeders have achieved in increasing cereal crop yields has been through breeding for traits that minimize or reduce plant-to-plant interactions (thereby reducing competition for light and growing space) (Duvick, 2005). One possible explanation for this unexpected pattern is that the growth of cultivated rice could be enhanced by root-to-root interactions. Consistent with this hypothesis, a study that examined the root growth of cultivated rice when grown the same genotype or a different genotype found that homotypic pairings led to greater intermingling of roots than heterotypic pairings (Fang *et al.*, 2013). That finding has been further supported by a recent study which suggests that below-ground kin recognition in cultivated rice plays an important role in root behavior and thus could explain the exploratory nature of the crop roots (Yang *et al.*, 2018b). In maize, modern varieties have been found to have shallower root angles than their historical progenitors (York *et al.*, 2015); this is also be consistent with selection for increased root interactions in this crop. Since our study was performed using individual plants grown alone in sterile gel media, field experiments should be undertaken to address the extent to which kin recognition may occur in crop fields and the role of the soil microbiome in mediating below-ground interactions. Expanded sampling of genotypes and plant growth stages could also be particularly insightful in this context.

Lack of parallelism

We found very little evidence for parallelism in this study, with only a single trait (*Convex Hull Volume (3D)*) mapping to overlapping genomic locations in both the S and B populations (Li *et al.*, 2017). This finding provides an interesting contrast to observations from studies of domestication traits in cereal crop species, which sometimes suggest a one gene – one trait pattern for domestication traits (reviewed in Sang, 2009). Our results in the weedy rice system show that this pattern does not necessarily extend to direct descendants of crop species. This inference has been further borne out by genome scans in weedy rice, where signatures of selection suggest little parallelism for above-ground trait QTLs or for genomic regions showing signatures of selection (Qi *et al.*, 2015; Li *et al.*, 2017). It should be noted that since the level of genetic resolution in the present study is on the scale of QTL intervals, the identity of the underlying causal genes remains unknown. Thus, it is possible that different QTL for a given RSA trait correspond to different genes within a single developmental pathway. If this is the case, then the prevalence of parallelism in RSA trait evolution may be greater than is apparent from our QTL data alone. Identification of candidate genes and confirmation of developmental pathways would be needed to definitively address this possibility.

Only two genes that directly control RSA have been cloned and functionally verified in plants, and neither of these genes appears to play a role in the RSA variation observed in the present study. *Dro1* occurs in the middle of chromosome 9 and encodes an auxin sensitive gravitropic response protein and thus controls rice root-soil angle, with plants homozygous for the upland allele developing roots with a higher angle relative to the soil (Uga *et al.*, 2011, 2013). This results in a deeper root system which is more drought tolerant. Similar phenotypes were linked to overexpression of *Dro1* in *Arabidopsis thaliana* and plum (*Prunus domestica*) (Guseman *et al.*, 2017). In the present study, although two width-associated QTL were mapped

to chromosome 9, neither QTL overlaps with the genomic region containing *Dro1*. Root-soil angle QTLs in this study localized to chromosomes 3, 7, and 12.

The other gene, *PSTOL1*, is an enhancer of early root growth in the middle of chromosome 12 which enables rice to increase intake of phosphorus in early growth stages (Gamuyao *et al.*, 2012). This gene was identified in the traditional *aus* rice variety Kasalath and was found to occur as a gene presence/absence polymorphism in other rice varieties. While it is known that DGWG lacks the *PSTOL1* gene, it is highly likely that our weedy rice parents both possess the gene since every US weed genotyped to date carries it (Vigueira *et al.*, 2016). Given that we did not find any QTL mapping to this locus, despite the probable presence/absence polymorphism in the RILs, it seems likely that *PSTOL1* is not a contributing factor to phenotypic variation in this system. This finding is consistent with a previous study of *PSTOL1* variation in cultivated and weedy rice, which detected no observable phenotypes associated with this gene (Vigueira *et al.*, 2016).

Implications for agriculture

Previous studies have linked phosphorus starvation tolerance and drought tolerance to root architecture, suggesting that breeders can select for a more optimal RSA to take advantage of soil conditions (Uga *et al.*, 2011; Gamuyao *et al.*, 2012). In this study, we identified early life-stage root depth QTL not associated with *Dro1* (Supporting Information Table **S2**). Although no test of drought tolerance was performed in this study, further experimentation would be relatively simple. If the prediction holds that plants with deep rooting-associated QTL are more drought tolerant, the weeds studied here could be a valuable resource in marker-assisted selection. In addition, our observations of differences in root system width and exploration between cultivated and weedy rice suggest that neighbor-to-neighbor root communication may

be important to growth in cultivated rice (Fang *et al.*, 2013; Yang *et al.*, 2018b). This study sheds light on potential QTLs of interest for further characterizing this trait and its potential agronomic value.

Author Contribution

K.M.O. planned and designed the research. M.J.W. performed the experiments, analyzed the data, and wrote the manuscript. C.N.T. contributed to the design of the research and provided lab equipment and space.

Acknowledgements

The authors would like to thank members of the Topp lab for assistance in working the gel imaging system and analysis pipeline. We would also like to thank David Goad of the Olsen lab for insight on generating the *R* script necessary to map QTLs. We would like to thank all members of the Olsen lab for useful comments on earlier drafts of this manuscript. MJW is supported by an NSF Graduate Research Fellowship Program (Fellow ID: 2017221153). Weedy rice research in the Olsen lab has been supported through the NSF Plant Genome Research Program (IOS-1032023).

References

Broman KW, Wu H, Sen Ś, Churchill GA. **2003**. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.

Burgos NR, Norman RJ, Gealy DR, Black H. **2006**. Competitive N uptake between rice and weedy rice. *Field Crops Research* **99**: 96–105.

Cao Q, Lu BR, Xia H, Rong J, Sala F, Spada A, Grassi F. **2006**. Genetic diversity and origin of weedy rice (*Oryza sativa* f. *spontanea*) populations found in North-eastern China revealed by simple sequence repeat (SSR) markers. *Annals of Botany* **98**: 1241–1252.

Casper BB, Jackson RB. **1997**. Plant Competition Underground. *Annual Review of Ecology and Systematics* **28**: 545–570.

Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. **2017**. A complete tool set for molecular QTL discovery and analysis. *Nature Communications* **8**: 1–7.

Duvick DN. **2005**. The Contribution of Breeding to Yield Advances in maize (*Zea mays* L.). *Advances in Agronomy* **86**: 83–145.

Fang S, Clark RT, Zheng Y, Iyer-Pascuzzi AS, Weitz JS, Kochian L V., Edelsbrunner H, Liao H, Benfey PN. **2013**. Genotypic recognition and spatial responses by rice roots. *Proceedings of the National Academy of Sciences* **110**: 2670–2675.

Federici MT, Vaughan D, Tomooka N, Kaga A, Wang XW, Doi K, Francis M, Zorrilla G, Saldain N. **2001**. Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. *Electronic Journal of Biotechnology* **4**: 42–57.

Fishman L, Beardsley PM, Stathos A, Williams CF, Hill JP. **2015**. The genetic architecture of traits associated with the evolution of self-pollination in Mimulus. *New Phytologist* **205**: 907–917.

Fong SS, Joyce AR, Palsson BØ. **2005**. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Reserch* **858:** 1365–1372.

Galkovskyi T, Mileyko Y, Bucksch A, Moore B, Symonova O, Price CA, Topp CN, Iyer-pascuzzi AS, Zurek PR, Fang S, *et al.* 2012. GiA Roots: software for the high throughput analysis of plant root system architecture. *BMC Plant Biology* **12**: 1–12.

Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, Dalid C, Slamet-Loedin I, Tecson-Mendoza EM, Wissuwa M, Heuer S. **2012**. The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. *Nature* **488**: 535–539.

Gealy DR, Agrama H a., Eizenga GC. **2009**. Exploring genetic and spatial structure of U.S. weedy red rice (*Oryza sativa*) in relation to rice relatives worldwide. *Weed Science* **57**: 627–643.

Grimm A, Fogliatto S, Nick P, Ferrero A, Vidotto F. **2013**. Microsatellite markers reveal multiple origins for Italian weedy rice. *Ecology and Evolution* **3**: 4786–4798.

Gross BL, Steffen FT, Olsen KM. **2010**. The molecular basis of white pericarps in African domesticated rice: Novel mutations at the *Rc* gene. *Journal of Evolutionary Biology* **23**: 2747–2753.

Guseman JM, Webb K, Srinivasan C, Dardick C. **2017**. *DRO1* influences root system architecture in *Arabidopsis* and *Prunus* species. *Plant Journal* **89**: 1093–1105.

Haley CS, Knott SA. **1992**. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**: 315–324.

Larson A. **2014**. Concepts in character macroevolution: adaptation, homology, and evolvability In: Losos JB, ed. The Princeton Guide to Evolution. Princeton: Princeton University Press, 89–99.

Lenser T, Theißen G. **2013**. Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science* **18**: 704–714.

Li L-F, Li Y-L, Jia Y, Caicedo AL, Olsen KM. **2017**. Signatures of adaptation in the weedy rice genome. *Nature Genetics* **49**: 811–814.

Li LF, Olsen KM. **2016**. To have and to hold: selection for seed and fruit retention during crop domestication. *Current Topics in Developmental Biology* **119**: 63–109.

Londo JP, Schaal BA. **2007**. Origins and population genetics of weedy red rice in the USA. *Molecular Ecology* **16**: 4523–4535.

Losos JB. **2011**. Convergence, adaptation, and constraint. *Evolution* **65**: 1827–1840.

Lyu H, He Z, Wu CI, Shi S. **2018**. Convergent adaptive evolution in marginal environments: unloading transposable elements as a common strategy among mangrove genomes. *New Phytologist* **217**: 428–438.

Mispan MS, Zhang L, Feng J, Gu X-Y. **2013**. Quantitative trait locus and haplotype analyses of wild and crop-mimic traits in U.S. weedy rice. *G3 (Bethesda, Md.)* **3**: 1049–59.

Morrell PL, Buckler ES, Ross-Ibarra J. **2011**. Crop genomics: advances and applications. *Nature Reviews Genetics* **13**: 85–96.

Ng J, Smith SD. **2016**. Widespread flower color convergence in Solanaceae via alternate biochemical pathways. *New Phytologist* **209**: 407–417.

Orr HA. **2005**. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**: 119–127.

Pfenning AR, Hara E, Whitney O, Rivas M V., Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell P V., Ganapathy G, *et al.* 2014. Convergent transcriptional specializations in the brains

of humans and song-learning birds. *Science* **346:6215** 1256846.

Pichersky E, Lewinsohn E. **2011**. Convergent evolution in plant specialized metabolism. *Annual Review of Plant Biology* **62**: 549–566.

Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. **2006**. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* **38**: 107–111.

Qi X, Liu Y, Vigueira CC, Young ND, Caicedo AL, Jia Y, Gealy DR, Olsen KM. **2015**. More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Molecular Ecology* **24**: 3329–3344.

Ramírez J, Górriz JM, Segovia F, Chaves R, Salas-Gonzalez D, López M, Álvarez I, Padilla P. **2010**. Computer aided diagnosis system for the Alzheimer's disease based on partial least squares and random forest SPECT image classification. *Neuroscience Letters* **472**: 99–103.

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL. **2010**. Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evolutionary Biology* **10**.

Rolli E, Marasco R, Vigani G, Ettoumi B, Mapelli F, Deangelis ML, Gandolfi C, Casati E, Previtali F, Gerbino R, *et al.* 2015. Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. *Environmental Microbiology* **17**: 316–331.

Saleem M, Law AD, Moe LA. **2016**. Nicotiana roots recruit rare rhizosphere taxa as major root-inhabiting microbes. *Microbial Ecology* **71**: 469–472.

Sang T. **2009**. Genes and mutations underlying domestication transitions in grasses. *Plant Physiology* **149**: 63–70.

Singh V, Burgos NR, Singh S, Gealy DR, Gbur EE, Caicedo AL. **2017a**. Impact of volunteer rice infestation on yield and grain quality of rice. *Pest Management Science* **73**: 604–615.

Singh V, Singh S, Black H, Boyett V, Basu S, Gealy D, Gbur E, Pereira A, Scott RC, Caicedo A, *et al.* 2017b. Introgression of Clearfield TM rice crop traits into weedy red rice outcrosses. *Field Crops Reserach* **207**: 13–23.

Spielmeyer W, Ellis MH, Chandler PM. **2002**. Semidwarf (*sd-1*), 'green revolution' rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of the National Academy of Sciences* **99**: 9043–9048.

Subudhi PK, Singh PK, Deleon T, Parco A, Karan R, Biradar H, Cohn MA, Sasaki T. **2014**. Mapping of seed shattering loci provides insights into origin of weedy rice and rice domestication. *Journal of Heredity* **105**: 276–287.

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jørgensen K, *et al.* 2011. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant Journal* **68**: 273–286.

Thurber CS, Jia MH, Jia Y, Caicedo AL. **2013**. Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Molecular Ecology* **22**: 685–698.

Topp CN, Bray AL, Ellis NA, Liu Z. **2016**. How can we harness quantitative genetic variation in crop root systems for agricultural improvement? *Journal of Integrative Plant Biology* **58**: 213–225.

Topp CN, Iyer-Pascuzzi AS, Anderson JT, Lee C-R, Zurek PR, Symonova O, Zheng Y, Bucksch A, Mileyko Y, Galkovskyi T, *et al.* 2013. 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. *Proceedings of*

the National Academy of Sciences of the United States of America **110**: 1695–1704.

Uga Y, Okuno K, Yano M. **2011**. *Dro1*, a major QTL involved in deep rooting of rice under upland field conditions. *Journal of Experimental Botany* **62**: 2485–2494.

Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, Hara N, Kitomi Y, Inukai Y, Ono K, Kanno N, *et al.* 2013. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nature Genetics* **45**: 1097–1102.

Vigueira CC, Li W, Olsen KM. **2013a**. The role of *Bh4* in parallel evolution of hull colour in domesticated and weedy rice. *Journal of Evolutionary Biology* **26**: 1738–1749.

Vigueira CC, Olsen KM, Caicedo AL. **2013b**. The red queen in the corn: agricultural weeds as models of rapid adaptive evolution. *Heredity* **110**: 303–311.

Vigueira CC, Small LL, Olsen KM. **2016**. Long-term balancing selection at the Phosphorus Starvation Tolerance 1 (PSTOL1) locus in wild, domesticated and weedy rice (*Oryza*). *BMC Plant Biology* **16**: 1–10.

Wedger MJ, Olsen KM. **2018**. Evolving insights on weedy rice. *Ecological Genetics and Genomics* **7**–**8**: 23–26.

Weller JL, Liew LC, Hecht VFG, Rajandran V, Laurie RE, Ridge S, Wenden B, Vander Schoor JK, Jaminon O, Blassiau C, *et al.* 2012. A conserved molecular basis for photoperiod adaptation in two temperate legumes. *Proceedings of the National Academy of Sciences* **109**: 21158–21163.

Xu XM, Li G, Su Y, Wang XL. **2018**. Effect of weedy rice at different densities on photosynthetic characteristics and yield of cultivated rice. *Photosynthetica* **56**: 520–526.

Yang X-F, Li L-L, Xu Y, Kong C-H. **2018**. Kin recognition in rice (*Oryza sativa*) lines. *New*

Phytologist: 567–578.

York LM, Das A, Weitz JS, Burridge J, Bucksch A, Lynch JP, Nord E. **2014**. Image-based high-throughput field phenotyping of crop roots. *Plant Physiology* **166**: 470–486.

York LM, Galindo-Castaneda T, Schussler JR, Lynch JP. **2015**. Evolution of US maize (*Zea mays* L.) root architectural and anatomical phenes over the past 100 years corresponds to increased tolerance of nitrogen stress. *Journal of Experimental Botany* **66**: 2347–2358.

Zhu Y, Ellstrand NC, Lu BR. **2012**. Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus *sh4* played a minor role in Asian rice domestication. *Ecology and Evolution* **2**: 2106–2113.

Tables

Table 1. QTLs for identified phenotypes that map in each of two mapping populations of cultivated × weedy rice.

Trait group	Phenotype	Mapping Population	QTL‡	Chromo-some	Genomic position (cM)	Physical position‡‡	LOD drop 1.5 confidence interval	PVE* (R^2)	α = 0.05 LOD threshold	LOD	Increased effect	Effect size per allele**	RIL Average	% Effect per allele
Root depth	Depth2D	B	*qRDP4B-a*	4	72.1	23609792	21631772-27286914	14.9	3.3	3.9	Crop	-5.8	76	-7.63
		S	*qRDP8S-a*	8	51.6	10219719	5757773-21026154	8.4	3	3.1	Crop	-3.8	70.4	-5.40
	majorElipseAxis2D	B	*qRDP4B-b*	4	72.1	23609792	21631772-27286914	14.9	3.2	3.9	Crop	-5.7	68.9	-8.27
		S	*qRDP8S-b*	8	47.6	8939999	5757773-17076446	9.9	3.2	3.6	Crop	-3.7	60.3	-6.14
Root system width	MinorEllipseAxis2D	B	*qRSW6B-a*	6	9.6	2472895	2176780-2867607	13.4	3	3.5	Weed-B	2.8	33.7	8.31
		S	*qRSW4S-a*	4	101	29817550	21631772-30808550	9.6	3.1	3.5	Crop	-1.8	34.4	-5.23
	MaxNetworkWidth3D	B	*qRSW9B-a*	9	13.3	8199257	7936052-10105116	13.3	3.3	3.4	Crop	-2.6	34.2	-7.60
		S	*qRSW4S-b*	4	90.7	28026225	21746628-33128017	7.4	3.1	3	Crop	-1.8	33.8	-5.33
		S	*qRSW5S-a*	5	30.1	4184550	3090444-5478629	6.6	3.1	2.7	Weed-S	1.7	33.8	5.03
Exploratory	Perimeter2D	B	*qREX6B-a*	6	9.6	2472895	2176780-2867607	13	3.2	3.3	Weed-B	394.2	2897.8	13.60
		S	*qREX4S-a*	4	101	29817550	23747123-33386030	8.7	3.1	3.1	Crop	-195.5	3081.6	-6.34
	Solidity3D	B	*qREX6B-b*	6	9.6	2472895	2176780-25947874	14.2	3.2	3.7	Crop	-0.001	0.005	-20
		S	*qREX4S-b*	4	57	21892817	19858548-33386030	7	2.4	2.5	Weed-S	0.001	0.007	14.29
	ConvexHullVolume3D	B	***qREX4B-a***	4	84.4	27210802	24867620-27286914	14.3	3.2	4.4	Crop	-10835	57968.3	-18.69
		B	*qREX6B-c*	6	9.6	2472895	2176780-2867607	11.2	3.2	3.6	Weed-B	9873	57968.3	17.03
		S	***qREX4S-c***	4	90.7	28026225	21844394-30808550	10.8	3.2	4	Crop	-7949	53084	-14.97
	NetworkConvexArea2D	B	*qREX4B-b*	4	84.5	27247058	21631772-27286914	13.9	3.2	4.3	Crop	-399	2870.6	-13.90
		B	*qREX6B-d*	6	9.6	2472895	2176780-2867607	12.1	3.2	3.8	Weed-B	381.6	2870.6	13.29
		S	*qREX4S-d*	4	101	29817550	22176514-33631677	9.14	3.1	3.3	Crop	-230.7	2658.5	-8.68
Root-soil angle	MeanLateralRootSoilAngle3D	B	*qRSA12B-a*	12	10	1659127	148244-2023121	13.7	3.2	3.5	Weed-B	3.8	32.5	11.69
		S	*qRSA2S-a*	2	170	34784617	32697234-35053040	9	3	3.3	Weed-S	1.1	22.6	4.87
	MeanRootSoilAngle3D	B	*qRSA12B-b*	12	10	1659127	148244-2023121	13.6	3.3	3.4	Weed-B	3.9	32.8	11.89
		S	*qRSA2S-b*	2	167	34307545	32697234-35053040	9	3	3.3	Weed-S	1.1	22.8	4.82

* PVE is percent of the phenotypic variation explained by the allelic variation at the QTL.
** A positive value in Effect size represents a positive change in the weed, while a negative value represents a positive change in the crop
‡QTLs in bold are the only QTL to map to the same position in both mapping populations
‡‡ Physical positions were determined relative to the MSU v7.0 assembly http://rice.plantbiology.msu.edu

Figures

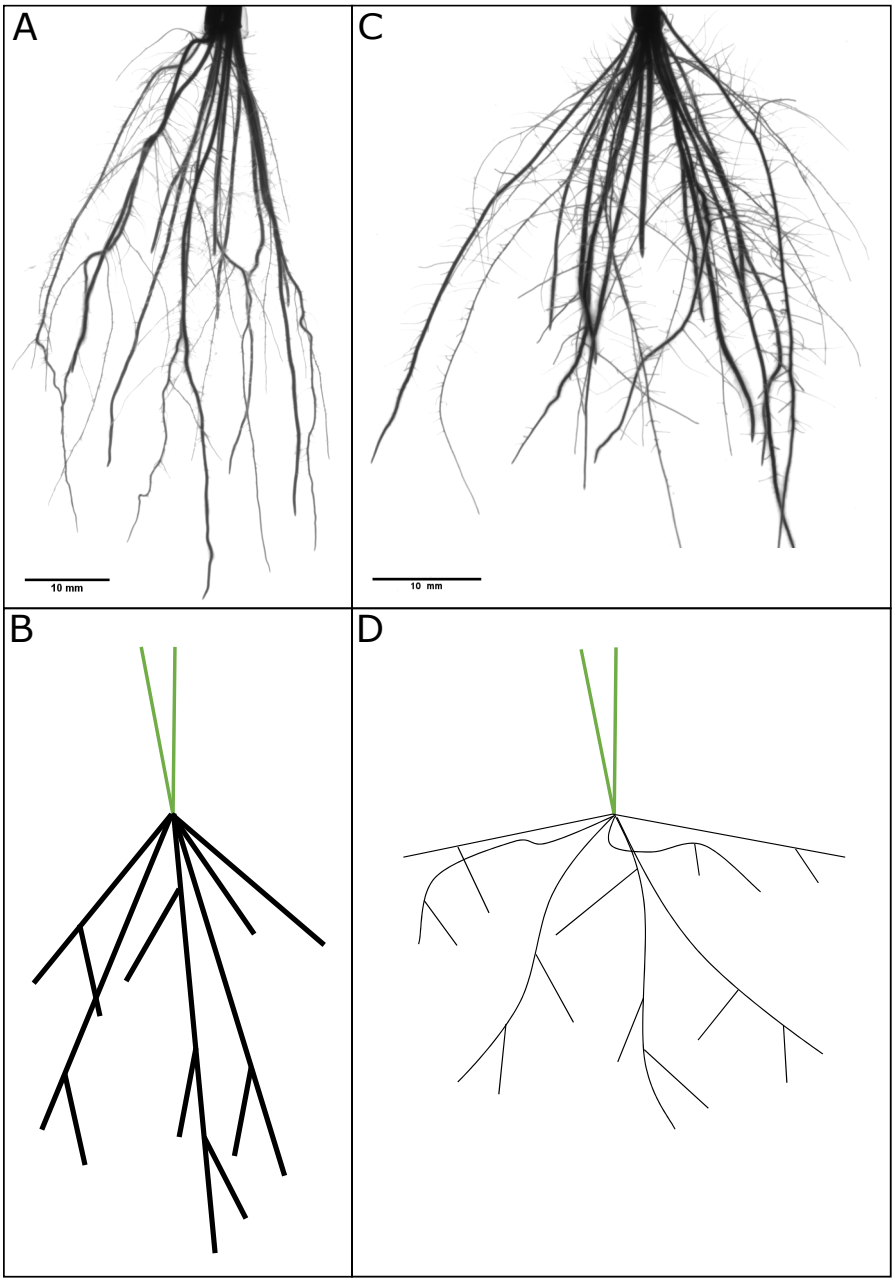


Figure 1. Characteristic differences in root system architecture between weedy and cultivated rice. Panels show digitized images (**a,c**) and schematic drawings (**b,d**) of typical weed (**a,b**) and crop (**c,d**) roots. Individual crop roots are thinner, longer, more curved, and more abundant, while the root system as a whole is higher in the soil, wider, and has shallower root-soil angles than the weed.

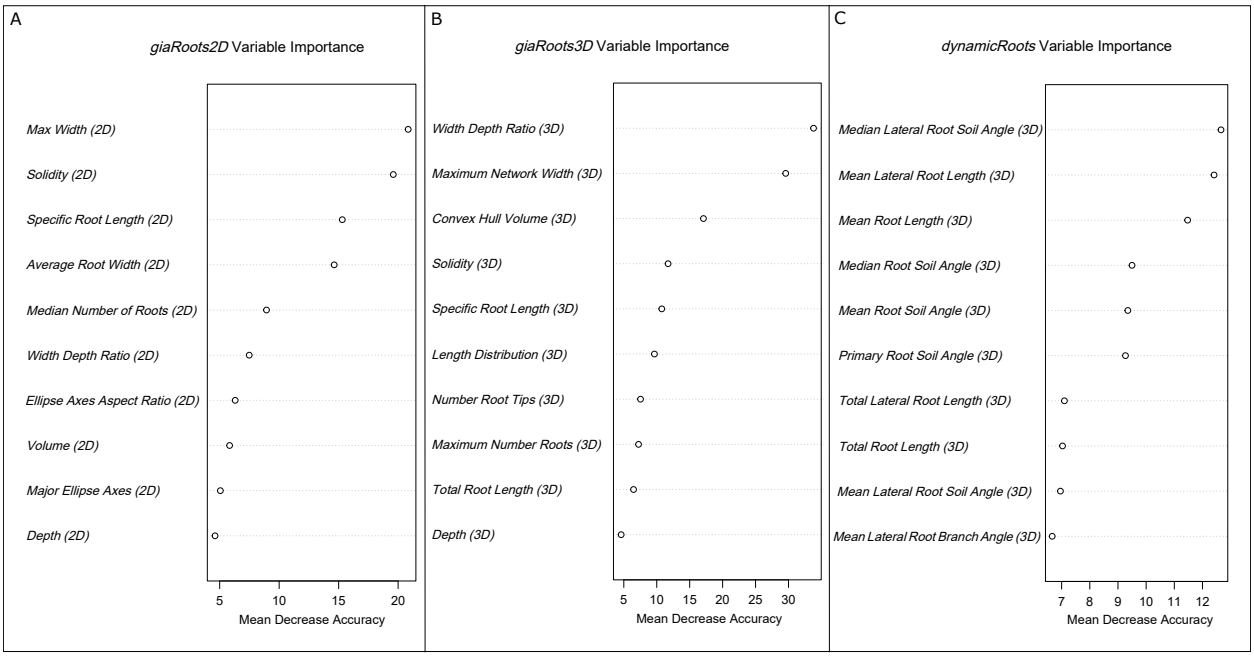


Figure 2. Diagnostic importance of rice root phenotypic variables from random forest machine learning model. *GiaRoots2D* (**a**), *GiaRoots3D* (**b**), and *DynamicRoots* (**c**) datasets put highest diagnostic importance on exploration, system width, and root-soil angle traits respectively. "Mean Decrease Accuracy" is a measure of how many extra observations would be misclassified if the trait in question were removed. Highly correlated traits will shift in importance between runs, but these changes in rank order do not change the biological interpretation.

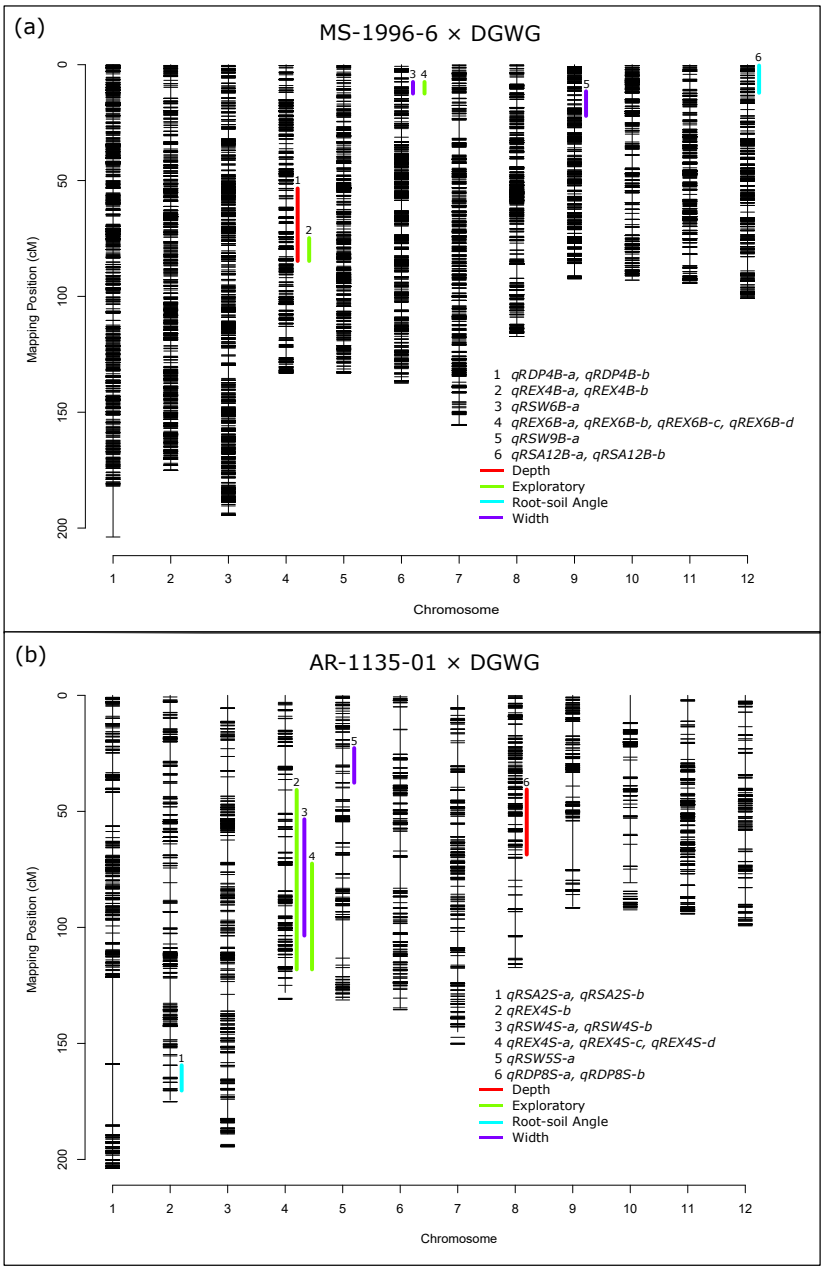


Figure 3. Rice genome linkage maps of the MS-1996-6 × DGWG "B" population (**a**) and AR-2000-1135-01 × DGWG "S" population (**b**) with QTL from four broad trait groups highlighted. Each vertical black line represents a rice chromosome, while horizonal hash marks indicate one SNP. Colored vertical lines represent the confidence intervals (LOD drop 1.5) of mapped traits. Only one trait (*Convex hull volume 3D*, an exploratory trait shown in bold font) maps to the same location in both populations.

APPENDIX I

Chapter 2 Supplementary Material

Supplemental Tables

Table S1. Success of random forest machine learning model.

	Probablity of Successful Identification*		
Trait group	BHA	Crop	SH
giaRoots2D	0.65	0.57	0.59
giaRoots3D	0.74	0.47	0.52
dynamicRoots	0.47	0.61	0.77
Combined	0.67	0.69	0.75

*Successful Identification of genotypes would be 0.33 by random chance alone

Table S2. All root QTLs identified in this study

Due to size of table, please see online supplementary materials here:
https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15791

or contact author.

Supplemental Figures

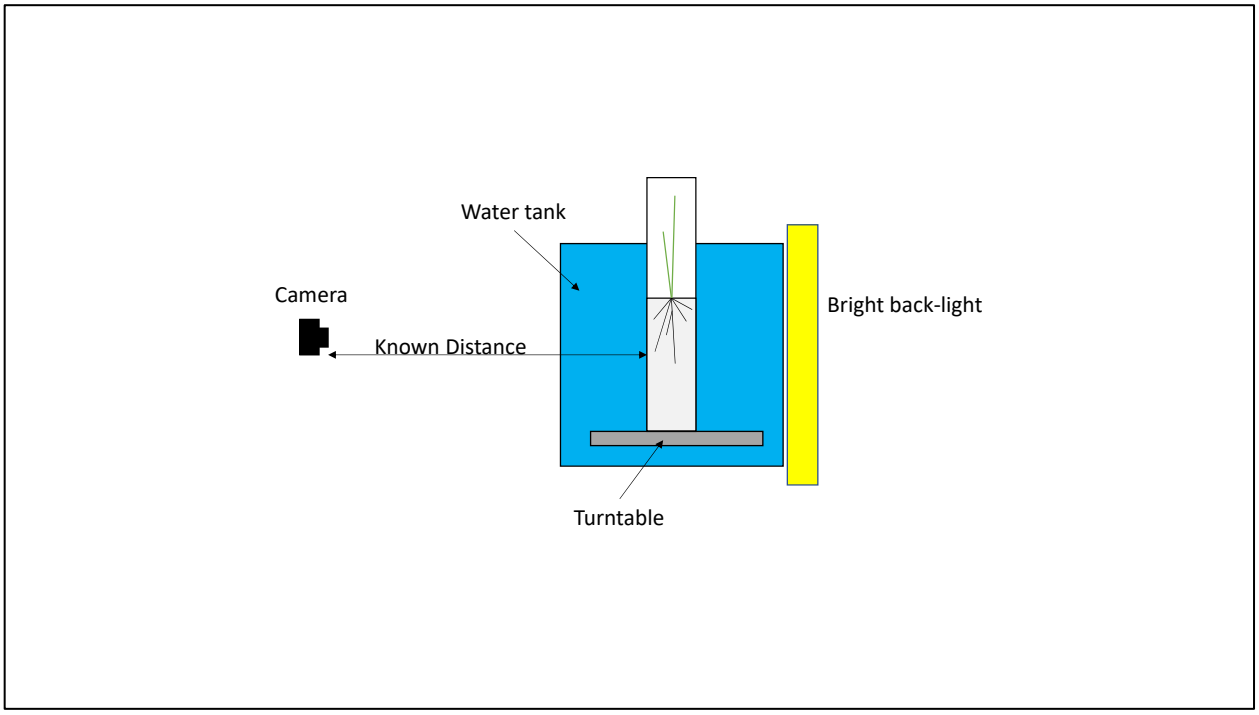


Figure S1. Schematic of the custom rig used to image the root system of rice plants growing in 2L glass cylinders. Camera and turntable are controlled with the same computer, allowing the camera to take 72 images exactly 5 degrees apart.

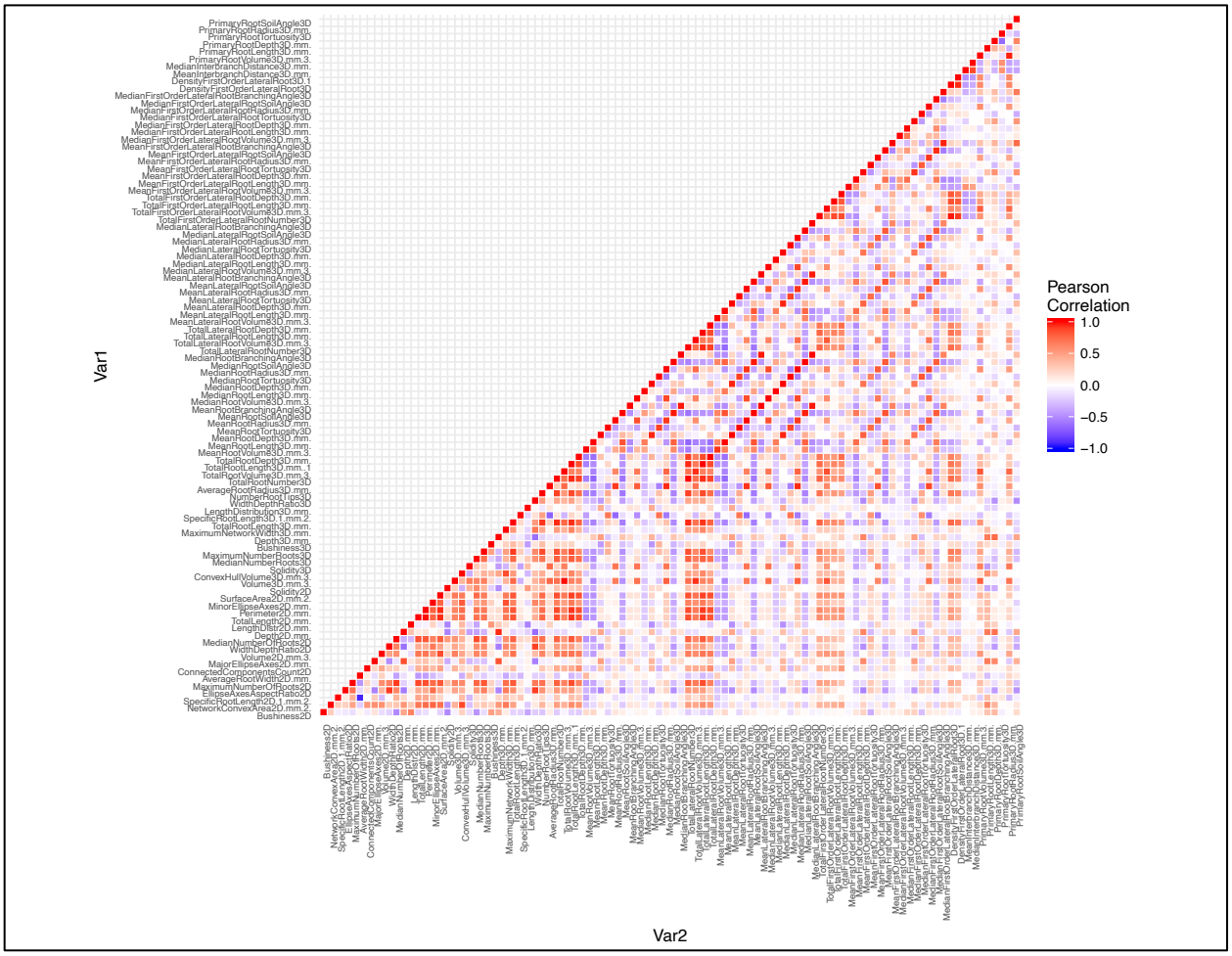


Figure S2. Correlation matrix of 98 rice root system architecture traits measured in this study.

CHAPTER THREE

Genomic revolution of US weedy rice in response to 21st century agricultural technologies

Abstract

Weedy rice is a close relative of cultivated rice that devastates rice productivity worldwide. In the southern United States, two distinct strains have been historically predominant, but the 21st century introduction of hybrid rice and herbicide resistant rice technologies has dramatically altered the weedy rice selective landscape. Here, we use whole-genome sequences of 48 contemporary weedy rice accessions to investigate the genomic consequences of crop-weed hybridization and selection for herbicide resistance. We find that population dynamics have shifted such that most contemporary weeds are now crop-weed hybrid derivates, and that their genomes have subsequently evolved to be more like their weedy ancestors. Haplotype analysis reveals extensive adaptive introgression of cultivated alleles at the resistance gene ALS, but also uncovers evidence for convergent molecular evolution in accessions with no signs of hybrid origin. The results of this study suggest a new era of weedy rice evolution in the United States.

Introduction

Understanding the genomic basis of adaptation is among the most important question in modern evolutionary biology. Crop domestication has long been recognized as a model for studying adaptive responses to selection (Darwin, 1859), and important insights in the last two decades have come from studies of evolving crop species' genomes (Lenser & Theißen, 2013; Purugganan, 2019). Recently, the evolution of agricultural weeds — which, unlike crops, evolve without intentional selection by humans — are providing additional new insights into the genomics of adaptation (Li *et al.*, 2017; Li & Olsen, 2020; Wu *et al.*, 2021). Among agricultural weeds, those that are closely related to crop species can be a particularly dynamic system because of the added potential for genetic exchange with crop cultivars as a means of weed evolution and adaptation (Ellstrand *et al.*, 1999).

The genomically best-characterized weedy crop relative is weedy rice (*Oryza* spp.), a de-domesticated form of cultivated rice (*O. sativa*) (Londo & Schaal, 2007; Li *et al.*, 2017; Roma-Burgos *et al.*, 2021b) that has evolved multiple times independently around the world (Qiu *et al.*, 2014; Li *et al.*, 2017; Wedger & Olsen, 2018). As a feral crop derivative, it is highly adapted and specialized to rice fields, where it competes aggressively with the crop. Just one weedy rice plant per square meter can lead to a >200 kg ha^{-1} loss of yield (Burgos *et al.*, 2006) and reductions in harvest quality that compromise market value (Ottis *et al.*, 2005; Cao *et al.*, 2007; Nadir *et al.*, 2017). Due to its close phenotypic and genetic similarity to the crop, weedy rice is challenging to control with herbicides and often requires additional specialized field maintenance practices. As a result, weedy rice causes annual economic losses of more than $45 million in the United States (US) (Estorninos *et al.*, 2005) and hundreds of millions of dollars worldwide (Chauhan, 2020).

Weedy rice strains worldwide are characterized by a few key shared weed-adaptive features, including a strong seed dispersal mechanism (shattering) (Thurber *et al.*, 2010; Subudhi *et al.*, 2014) and persistent seed dormancy (Gu *et al.*, 2011). In the southern US, two phenotypically and genetically distinct morphotypes have historically predominated; strawhull (SH) weedy rice is descended from from *indica* rice varieties grown in Asia, while blackhull-awned (BHA) is derived from genetically distinct *aus* Asian varietal group (Londo & Schaal, 2007; Olsen *et al.*, 2007; Reagon *et al.*, 2010). Outcrossing rates between SH, BHA, and local US cultivars (all *tropical japonica* varieties) have historically all been <1% despite their close physical proximity within US rice fields (Shivrain *et al.*, 2009; Li *et al.*, 2017).

In 2002, non-transgenic herbicide resistant (HR) rice cultivars (marketed as Clearfield™ rice) were first commercialized in the US as a means of controlling weedy rice and other agricultural weeds. These HR cultivars are resistant to the imidazolinone (IMI) class of herbicides due to one of two amino acid replacements in the acetolactate synthase enzyme

(ALS). The first Clearfield™ cultivars, CL121 and CL141, carried a $G_{654}E$ replacement (Rajguru *et al.*, 2005); they were quickly replaced in 2003 by CL161 and later cultivars, which instead carry an adjacent $S_{653}N$ replacement conferring greater herbicide resistance. US HR rice cultivation peaked at ~65% by the mid-2010s and now constitutes ~35% of rice acreage (Moldenhauer *et al.*, 2020).

Concurrent with the introduction of HR rice, US rice agriculture was further altered by the adoption of hybrid rice technology in place of traditional inbred cultivars. First commercialized in the US in 2000 and now comprising ~50% of US rice acreage (including many HR cultivars) (Moldenhauer *et al.*, 2020), hybrid rice offers the substantial advantage of enhanced yield through heterosis (Singh *et al.*, 2017b). However, an unintended consequence of this technology has been the large increase of instances of volunteering (Singh *et al.*, 2016), whereby cultivar seeds shatter in the field, overwinter and emerge in subsequent years . Allelic segregation in these hybrid-derived crop volunteers results in a wide range of phenotypic variation, including for flowering time, which increases outcrossing rates with weedy rice (Singh *et al.*, 2017b). Volunteer rice thus has the potential to serve as a gene flow bridge, allowing for the escape of HR, and other crop-derived alleles into weedy rice.

The combined adoption of HR and hybrid rice in US agriculture has thus created a two-decades long natural experiment: two genetically distinct strains of a historically self-fertilizing weedy crop relative have now been subject to strong selection for herbicide resistance, and this selective pressure has coincided with increased opportunities for crop-weed hybridization via crop volunteers. Notably, as early as 2004, farmers utilizing the Clearfield™ technology reported instances of HR weedy rice (Rajguru *et al.*, 2005). By 2010, 80% of weedy rice plants sampled in one study were classified as resistant and carrying the $S_{653}N$ allele derived from HR cultivars (Singh *et al.*, 2017d). In the decade of continued HR cultivar use that has followed, it is unclear

how weedy rice has continued to evolve and adapt, or the extent to which crop × weed hybridization has continued to shape the genetic composition of US weedy rice populations.

In this study we used whole genome resequencing to investigate how the genomic composition of southern US weedy rice has changed since the 21st century introduction of HR and hybrid rice cultivars. We addressed the following specific questions: 1) How do the genomes of contemporary weeds differ from the historic SH and BHA strains that predominated through the 20th century? 2) Following crop-weed hybridization (creating a weed with 50:50 crop-weed genomic composition), does selection over subsequent generations in weed populations lead to a genome-wide bias toward one ancestral genome or the other? And 3) Within the weed genome, does selection drive known weed- or crop-specific alleles to high frequency in a predictable pattern based on expected advantageous traits for contemporary weeds? Our findings reveal a genomic revolution in US weedy rice in the last 20 years that has irrevocably altered crop-weed dynamics and mechanisms of weed adaptation.

Results

Population genetics of contemporary US weedy rice

Seeds from 48 maternal samples across 5 Arkansas rice fields were collected during the harvest season of 2018. US weedy rice lacks geographical genetic structure(Burgos *et al.*, 2014; Li *et al.*, 2017), so this sampling may be considered representative of the southern US rice production region (Reagon *et al.*, 2010; Burgos *et al.*, 2014; Li *et al.*, 2017). Whole genome sequences (>40x average coverage) were generated using leaf tissue from one seed per maternal plant grown to the seedling stage. Genome assemblies were analyzed with 98 previously published weedy, cultivated, and wild rice samples (Upadhyaya, 2007; Huang *et al.*, 2012; Genomes, 2014; Leung *et al.*, 2015; Li *et al.*, 2017) resulting in a dataset of 146 samples and ~19.34 million SNPs. Previously published genomes included 22 historic weedy (11 SH and

11 BHA), 49 cultivated (10 *aus*, 5 *aromatic*, 12 *indica*, 12 *temperate japonica*, and 10 *tropical japonica*) and 27 wild rice accessions. Wild rice accessions were removed from analysis after they were confirmed to play no role in US weedy rice evolution, as was expected given their absence from the US agroecosystem (Supplementary Fig. 1).

To assess the overall genetic composition of contemporary weed samples in comparison to historic US weed strains, we employed principal component analysis (PCA) and ADMIXTURE analysis. The PCA revealed relatively tight within-strain grouping of cultivated and historic (pre-2000) weedy rice, with contemporary weedy rice showing a much broader dispersion (Fig. 1). PC1 (22.8% variation explained) separated the *japonica* and *indica* subspecies lineages, which is the deepest divergence in the Asian rice taxonomy. PC2 (15.6% variation explained) separated subgroups within the *indica* subspecies, with *aus* crop varieties and *aus*-like weeds distinguished from *indica* and *indica*-like samples. Aside from four contemporary weed accessions that cluster very closely with historic SH strains, all other contemporary weeds have intermediate distributions along PC1 between historical weedy rice (SH, BHA) and the US cultivated rice group (*tropical japonica*) (Fig. 1). This suggests that all but four of the contemporary US weed samples are derived from crop-weed hybridization. Among these hybrid descendants, far more appear to be related to historic BHA strains (38 accessions) than to historic SH strains (6 accessions).

For ADMIXTURE analyses of population structure, CV scores indicated K=6 as the optimal number of populations. However, we believe that K=5 makes the most biological sense since at K=6 and above, the contemporary weeds are subdivided into genetically bottlenecked subgroups, revealing no further information with respect to ancestry (Fig. 2). At K=5, the genetic groups corresponded broadly to the following: *japonica* cultivated varieties (including US cultivars), *indica* cultivated varieties, historic SH weeds, historic BHA weeds, and a genetically homogeneous subgroup within the contemporary weeds that in the PCA are grouped with other

crop-BHA hybrid descendants. This genetically homogeneous subset of BHA-like weeds may represent a derivative population of BHA × *tropical japonica* hybrids that emerged early enough after HR cultivar introduction to have evolved into a genetically homogeneous subgroup through multiple generations of inbreeding (see also genetic diversity quantifications below); it is designated the 'beta' group in reference to this inferred early origin.

Consistent with results from the PCA, ADMIXTURE analysis suggests that most contemporary US weeds are genetic admixtures descended from hybridization between the historic weed strains and US cultivated rice. At K=5, 35 of 48 contemporary weed accessions (72.9%) had membership assignment coefficients of >15% in two or more genetic populations. Most of these admixed weeds (28 of 48, or 58.3%) appear to be derived from BHA rather than SH historic weeds, which account for 6 of the 48 admixed accessions (12.5%). A single contemporary accession appears to have complex SH-BHA admixed ancestry, with >20% membership coefficients from SH, BHA, and *tropical japonica* genetic populations. Nine samples (18.8%) fell into the homogeneous beta group. As in the PCA, the remaining four contemporary samples (8.3%) were genetically indistinguishable from historical SH weeds. Thus, crop-weed hybridization appears to have given rise to most contemporary US weedy rice, with most of these hybrid derivatives descended from BHA-crop hybridization.

Genetic diversity measures were calculated at every SNP across the genome in order to gain a snapshot of the contemporary weedy rice genome. These measures allowed us to quantify the relative endurance of weed and crop ancestor genomes on a genome-wide scale, and to gauge the relative timing of emergence of the homogeneous beta weed population in comparison to the more heterogeneous contemporary weed groups. Heterozygous SNP quantification indicated that contemporary weeds collectively have a high number of heterozygous sites when compared to their crop ancestors (Supplementary Fig. 2); this is consistent with their relatively recent hybrid ancestry. Among the contemporary hybrid-derived

weeds, SH-like weeds averaged higher heterozygosity than BHA-like weeds, with the 'beta' subpopulation having significantly lower heterozygosity overall. (Supplementary Fig. 3). In the samples with clear weed and crop admixed ancestry (excluding the 'complex' accession), heterozygosity-based estimates of generations since hybridization suggest that most of our samples are five or more generations post-hybridization, with only eight samples less than three generations post-hybridization (Supplementary Fig. 4); these may be conservative estimates, as they assume a return to complete selfing after a single outcrossed generation. Accounting for the soil seed bank and seed dormancy, these results are thus in line with a 20-year-old phenomenon for HR weedy rice evolution via crop-weed hybridization.

Genome wide local ancestry

The *Loter* software package (Dias-Alves *et al.*, 2018) was used to calculate estimates of local ancestry throughout the contemporary weedy rice genome in order to reveal any bias towards crop or weed ancestry that has arisen since hybridization. Notably, the contemporary weeds have shifted away from the 50:50 ratio predicted under neutral genetic drift, and instead show an average of 74.1% and 69.2% assignment to the historical weed genome for BHA-like and SH-like groups, respectively (Table 1, Supplementary Fig. 5). The similarity of these values suggests that both of these independently evolved weed lineages are evolving back toward the historic weed genome at a similar rate. Taken together with the heterozygosity measures above, we can conclude, with high certainty, that the descendants of hybridization events that occurred soon after the introduction of HR rice cultivars have persisted and that they show a clear bias, on a genome-wide level, of evolving back towards their weedy ancestor.

F_{ST} was calculated between the hybrid-derived contemporary weeds and their inferred ancestors in a genome-wide sliding window analysis to search for evidence of adaptation via selective introgression of weed or crop alleles. We specifically compared *ALS*, the locus

conferring IMI herbicide resistance (where crop alleles are predicted to be strongly favored), with *Rc*, a locus conferring seed dormancy (where weed alleles are predicted to be strongly favored). As hypothesized, we found consistent evidence of a crop-like *ALS* region on chromosome 2 (Fig. 3a). We also identified a weed-like *Rc* region on chromosome 7, although this pattern only held for the BHA-like, and not the SH-like weeds (Fig. 3b). Consistent with the F_{ST} sliding window analysis, the *Loter* software identified a large crop-like haplotype block in the region containing *ALS*; interestingly, this was only the case for BHA-like samples (Fig. 4). For *Rc*, *Loter* identified a weed-like region around *Rc,* which could reflect selective maintenance of the dormancy-associated weed allele (or simply the overall genomic shift towards the weed-like genome).

Haplotype network analysis of ALS

To gain a finer-scale view of haplotype variation at the *ALS* HR locus, a median joining network tree was constructed from manually phased consensus nucleotide sequences retrieved from assembled raw reads (Fig. 5, Supplementary Fig. 6). The haplotype tree is structured into two diverged haplogroups, with haplotypes on the right side of the network derived from cultivar (*tropical japonica*) ancestry and those on the left side characteristic of weedy ancestry. Most of the contemporary weeds are distributed on the right side of the network and carry the $S_{653}N$ mutation and surrounding haplotype sequence present in the widely grown CL161 and later HR cultivars. Two weed samples, E08 and E09, are also on the right side of the network but instead carry the $G_{654}E$ mutation and surrounding haplotype indicative of the oldest HR cultivars (CL121 or CL141); this suggests that these two samples are descendants of the very earliest crop × weed hybridization events.

The left section of the haplotype tree, conversely, does not have *ALS* haplotypes of cultivar origin. These haplotypes are represented almost entirely by SH-like plants, consistent

with *Loter* results for the *ALS* genomic region. Samples A01 and A08 carry the older $G_{654}E$ mutation, but do not show evidence of hybrid origin and occur in a distinctly weed-like haplotype background. This allele was previously shown to have been present in the historical SH population at low frequency (Sales *et al.*, 2008), likely due to infrequent exposure to imazethapyr during IMI-resistant soybean rotations. The presence of this allele in contemporary weedy rice populations is thus most likely due to selection on standing variation. Two additional samples, A05 and A06, carry the $S_{653}N$ resistance allele, and also show no evidence for hybrid ancestry; this suggests a convergent mutation event conferring resistance. To our knowledge, this is the first report of the $S_{653}N$ resistance allele occurring in weedy rice through mutational convergence rather than crop allele introgression.

Herbicide resistance phenotyping confirmed that most of our samples showed some level of resistance following application of imazethapyr, with 34/48 samples (70.8%) classified as highly resistant (Table 2). Another 4/48 (8.3%) of samples showed moderate levels of resistance, while 8/48 (16.6%) were segregating for resistance. Thus, the vast majority of contemporary weed genotypes (46/48, or 95.8%) show some degree of herbicide resistance. Only two samples (4.1%) were completely susceptible in our phenotyping trials; both susceptible plants were of crop-weed hybrid origin and were collected from fields not utilizing the Clearfield™ technology. Thus, we suspect they are offspring of parents segregating for resistance. Most plants showing high herbicide resistance carried the common CL161 haplotype (characterized by the $S_{653}N$ mutation); additional samples carry the older resistance haplotype of CL121 and CL141 cultivars (characterized by the $G_{654}E$ mutation) (Fig. 5). HR phenotyping also confirmed resistance in the four SH weeds that are not of crop-weed hybrid origin and that appear to have evolved resistance through mutational convergence.

Discussion

One question that emerges from these results is: why are the majority of contemporary plants of crop-weed hybrid origin? The lack of hybrid persistence prior to 2000 (Reagon *et al.*, 2010) suggests low hybrid fitness. Additionally, the existence of "pre-adapted" HR weedy rice strains, even at initial low frequency, would lead one to expect those fit individuals to quickly rise to high frequency. Instead, reality suggests that those high-fitness individuals only make up ~4% of contemporary samples, while the presumably low-fitness hybrids make up 92% of the contemporary population. It could be possible that while F_1 fitness is low, fitness of later stages, after selfing, is higher as alleles segregate into favorable configurations.

Local ancestry analysis of BHA-like, 'beta', and SH-like accessions revealed genomes, regardless of ancestry, built primarily of components derived from their weedy rice ancestor. This consistency broadly suggests either a selective maintenance of weedy genome components or a selective purge of crop alleles – though these need not be mutually exclusive. The results described here could help inform discussions on crop allele escape (transgenic or otherwise) and the genome wide process of adaptive introgression in agro-ecosystems.

The shifting landscape of rice agriculture has resulted in a new generation of weedy rice. The Clearfield™ cropping system has reduced average field infestations drastically, but two decades of herbicide application in the presence of hybrid rice gene-flow bridges has resulted in weedy rice that is herbicide resistant and likely more competitive than historical populations. The rapid adaptation of weedy rice to herbicide application should serve as yet another example of the dangers of relying on single methods of control for agricultural pests.

Acknowledgements

The authors would like to thank members of the Roma-Burgos lab for their assistance with herbicide resistance trials. We would also like to thank members of the Olsen lab for helpful comments on earlier versions of this manuscript. MJW is supported by an NSF Graduate

Research Fellowship Program (Fellow ID: 2017221153). Weedy rice research in the Olsen laboratory has been supported through the NSF Plant Genome Research Program (IOS-1947609).

Materials and Methods

Plant Materials

Contemporary weedy rice plants were collected from rice fields in Greene County, AR, in late August 2018. Fields were selected based on their cropping history as reported by local farmers and are representative of the major rice growing area of the southern US. Previous population genetic studies have documented that the weed strains show no evidence of geographical population structure across the southern US rice growing region (Londo & Schaal, 2007). Samples were collected from fields representing three different rice cropping histories: HR inbred cultivars (1 field, 14 plants); HR hybrid cultivars (1 field, 15 plants), and hybrid non-HR cultivars (3 fields, 19 plants) (TABLE S1). Where applicable (43 of 48 samples), full mature panicles were clipped and collected in the field from weedy rice plants no closer than 5 meters from another collection site. For the remaining samples, where seeds had not yet reached maturity, plants were transplanted to Washington University in St. Louis (WUSTL) greenhouse facilities in 18-gallon plastic bins and brought to seed maturity in growth chambers (28°C, 16:8 hour day:night, 60% humidity).

Whole genome sequencing

One seed per field-sampled plant was brought to seedling stage, from which fresh leaf tissue was collected and ground in liquid nitrogen for DNA extraction using a modified CTAB protocol (Gross *et al.*, 2009). It should be noted that the DNA for this study was collected from

plants grown and selfed one additional generation in the greenhouse; therefore, variation observed in genome sequence data may not correspond perfectly to field-collected genotypes, particularly for segregating variants in hybrid derivatives. Illumina libraries were generated in house using a Nextera DNA Flex library prep kit with Nextera DNA CD indexes with the i5 bases recommended for HiSeq 3000/4000 (Illumina, San Diego, CA). Samples were multiplexed following recommendations from Nextera and sent to Novogene (Novogene Corporation Inc., Sacramento, CA) for paired-end short read sequencing on the HiSeq X 10 platform. Raw reads were de-multiplexed by Novogene before data return.

Data collection and SNP filtering

Whole-genome sequencing reads from contemporary weedy rice samples were combined with raw reads from previously published whole genome studies (Upadhyaya, 2007; Huang *et al.*, 2012; Genomes, 2014; Leung *et al.*, 2015) resulting in a full dataset of 146 samples representing cultivated, weedy and wild rice (Supplementary Table S1). All SNP identification and filtering was performed using the full dataset. Raw reads were trimmed for quality control using default parameters in *Trimmamatic* (Bolger *et al.*, 2014), followed by alignment to the MSU version 7.0 rice reference genome (Kawahara *et al.*, 2013) using *BWA* (Li & Durbin, 2009). Aligned sequences were sorted and converted to .bam files using *samtools*. The *mpileup* program in the *bcftools* software (Li *et al.*, 2009) was used for variant calling and conversion to the .vcf file type. Finally, *vcftools* (Danecek *et al.*, 2011) was used to filter out indels, remove variants with a minor allele frequency < 0.05, and remove sites clearly out of Hardy-Weinberg equilibrium ($p<0.0000001$). *Vcftools* was also used to remove wild samples from .vcf files for analyses where they were not required (described below).

Population Genetic Analyses

The *pca* flag within *plink* (Purcell *et al.*, 2007) was used in conjunction with *ADMIXTURE* (Alexander *et al.*, 2009) to determine population structure (supplementary Fig.1). Wild rice was found to show little-to-no overlap with contemporary US weedy rice and was removed from further analysis. Principal component analysis (PCA) and *ADMIXTURE* analysis showed that grouping contemporary weeds by field type was uninformative for explaining population structure; they were therefore grouped and analyzed based on their predominant weedy rice ancestry in subsequent analyses. From the ADMIXTURE and PCA results combined, contemporary weeds were categorized into three groups: 'SH-like', defined as >10% SH ancestry in ADMIXTURE (without BHA contribution) or placement in the PCA output as intermediate between historic SH and cultivated *tropical japonica* strains; 'BHA-like', defined as >10% BHA ancestry in ADMIXTURE (without SH contribution); and the 'beta' group, defined based on placement with BHA-like weeds in the PCA (and in the ADMIXTURE analysis at K=4), but with assignment to its own unique genetic population in the ADMIXTURE analysis at $K \geq 5$. A single contemporary weed accession with complex admixed ancestry was assigned its own category ('complex').

Heterozygous sites among genome-wide SNPs were calculated per accession using the *-het* flag in the *plink* software. Wright's F_{ST} was calculated for each contemporary weed group in relation to its weed and crop ancestors using the *–weir-fst-pop* flag in the *vcftools* software, with a window size of 500 kb and a 250 kb step size. The first F_{ST} calculation measured differentiation between a given contemporary weed group (as identified in population structure analyses) and the predominant weed ancestor of that group (SH or BHA), while the second measured differentiation between that weed group and the rice variety group representing US cultivars (*tropical japonica*). These F_{ST} values were then plotted together across the 12 chromosomes of the rice genome to identify genomic regions with differential contributions of the weed or crop ancestor. Average pairwise nucleotide diversity (π) values for contemporary

weed groups were calculated and visualized in the same way using the *–site-pi* flag in the *vcftools* software. As a measure of inbreeding, homozygous locus counts were performed in *vcftools* using the *-het* flag and converted to fraction of heterozygous loci using the formula ((N_sites – O(HOM))/N_sites). Lastly, a custom Python script was developed to identify ancestrally-informative SNPs (defined here as sites that are fixed differences between the presumptive ancestors of a hybrid individual). This script then calculated observed heterozygous genotype counts at those sites only. This analysis allowed us to estimate the number of generations since hybridization, assuming a 50% reduction in the number of heterozygous genotypes per generation and a return to a strictly selfing mating system.

Local Ancestry

To complement F_{ST} analyses, local ancestry across the genome was calculated for weed groups using the *Loter* software (Dias-Alves *et al.*, 2018) and visualized with *matplotlib* (Hunter, 2007). A custom python script was used to quantify the amount of ancestral genomes (crop vs. weed) found in the contemporary hybrids. A second custom Python script was written to convert an MSU-7.0 genomic location to the corresponding bin of the *Loter* output. This allowed us to pinpoint potential candidate genes for weed adaptation.

ALS haplotype network analyses

The *samtools* (Li *et al.*, 2009) software package was used to retrieve raw reads mapping to the *ALS* gene region from sorted .bam files. These raw reads were retrieved by using the *index*, *view* (specifying the known gene boundaries) and *fasta* commands. Raw reads were then exported to the *Geneious* 8.1.6 software (https://www.geneious.com) for assembly to a reference *ALS* sequence obtained from GenBank (accession MH636577). After assembly, sequences were trimmed to match the reference sequence and manually phased to remove heterozygous calls from consensus sequences. Phased and trimmed sequences were exported

to the *PopART* software (Bandelt *et al.*, 1999; Leigh & Bryant, 2015) for haplotype network visualization.

Herbicide Resistance Phenotyping

Weedy rice seeds were planted at 1.27 cm depth into pots (15.24-cm top diameter) filled with 50:50 mixture by volume of field soil and Sunshine potting mix. Up to 12 seeds were planted per pot, depending on the quantity of seeds available per sample. The pots were placed in a greenhouse with supplemental lighting to achieve a 16-h daylength. The temperature was set at a minimum of 25 ºC and maximum of 35 ºC. At the 3-leaf stage, the plants were treated with 70 g ai ha^{-1} imazethapyr two times, 10 days apart. Imazethapyr was applied in 187 L ha^{-1} spray volume, in a spray chamber with a motorized spray boom fitted with two 800067 flat fan nozzles spaced 46 cm apart. The herbicide treatment was replicated three times and a nontreated check for each sample served as reference for evaluation of plant response. Visible injury was evaluated 3 weeks after the second application of imazethapyr on a scale of 0 to 100% where 0 indicated no injury and 100 indicated a dead plant. The level of injury reflects the level of resistance to imazethapyr.

Bibliography

1. Darwin, C. *On the origin of Species by Means of natural Selection, or Preservatoin of Favoured Races in the Struggle for Life*. (John Murray, 1859).

2. Lenser, T. & Theißen, G. Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* **18**, 704–714 (2013).

3. Purugganan, M. D. Evolutionary Insights into the Nature of Plant Domestication. *Curr. Biol.* **29**, R705–R714 (2019).

4. Li, L.-F. F., Li, Y.-L. L., Jia, Y., Caicedo, A. L. & Olsen, K. M. Signatures of adaptation in the weedy rice genome. *Nat. Genet.* **49**, 811–814 (2017).

5. Li, L.-F. & Olsen, K. M. Population Genomics of Weedy Crop Relatives: Insights from Weedy Rice. (2020) doi:10.1007/13836_2020_77.

6. Wu, D., Lao, S. & Fan, L. De-Domestication: An Extension of Crop Evolution. *Trends Plant Sci.* **26**, 560–574 (2021).

7. Ellstrand, N. C., Prentice, H. C. & Hancock, J. F. Gene Flow and Introgression from Domesticated Plants into Their Wild Relatives. *Annu. Rev. Ecol. Syst.* **30**, 539–563 (1999).

8. Londo, J. P. & Schaal, B. A. Origins and population genetics of weedy red rice in the USA. *Mol. Ecol.* **16**, 4523–4535 (2007).

9. Roma-Burgos, N., San Sudo, M. P., Olsen, K. M., Werle, I. & Song, B.-K. Weedy Rice (*Oryza spp.*): What's In a Name? *Weed Sci.* 1–36 (2021) doi:10.1017/wsc.2021.22.

10. Wedger, M. J. & Olsen, K. M. Evolving insights on weedy rice. *Ecol. Genet. Genomics* **7**–**8**, 23–26 (2018).

11. Qiu, J. *et al.* Genome re-sequencing suggested a weedy rice origin from domesticated indica-japonica hybridization: a case study from southern China. *Planta* **240**, 1353–1363 (2014).

12. Burgos, N. R., Norman, R. J., Gealy, D. R. & Black, H. Competitive N uptake between rice and weedy rice. *F. Crop. Res.* **99**, 96–105 (2006).

13. Cao, Q. J., Li, B., Song, Z. P., Cai, X. X. & Lu, B. R. Impact of weedy rice populations on the growth and yield of direct-seeded and transplanted rice: Research paper. *Weed Biol. Manag.* **7**, 97–104 (2007).

14. Nadir, S. *et al.* Weedy rice in sustainable rice production. A review. *Agron. Sustain. Dev.* **37**, (2017).

15. Ottis, B. V., Smith, K. L., Scott, R. C. & Talbert, R. E. Rice yield and quality as affected by cultivar and red rice (*Oryza sativa*) density. *Weed Sci.* **53**, 499–504 (2005).

16. Estorninos, L. E., Gealy, D. R., Gbur, E. E., Talbert, R. E. & McClelland, M. R. Rice and red rice interference. II. Rice response to population densities of three red rice (Oryza sativa) ecotypes. *Weed Sci.* **53**, 683–689 (2005).

17. Chauhan, B. S. Grand Challenges in Weed Management. *Front. Agron.* **1**, 1–4 (2020).

18. Subudhi, P. K. *et al.* Mapping of seed shattering loci provides insights into origin of weedy rice and rice domestication. *J. Hered.* **105**, 276–287 (2014).

19. Thurber, C. S. *et al.* Molecular evolution of shattering loci in U.S. weedy rice. *Mol. Ecol.* **19**, 3271–3284 (2010).

20. Gu, X. Y. *et al.* Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice.

Genetics **189**, 1515–1524 (2011).

21. Olsen, K. M., Caicedo, A. L. & Jia, Y. Evolutionary genomics of weedy rice in the USA. *J. Integr. Plant Biol.* **49**, 811–816 (2007).

22. Reagon, M. *et al.* Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evol. Biol.* **10**, (2010).

23. Shivrain, V. K., Burgos, N. R., Gealy, D. R., Sales, M. A. & Smith, K. L. Gene flow from weedy red rice (*Oryza sativa* L.) to cultivated rice and fitness of hybrids. *Pest Manag. Sci.* **65**, 1124–1129 (2009).

24. Rajguru, S. N. *et al.* Mutations in the Red Rice ALS Gene Associated with Resistance to Imazethapyr. *Weed Sci.* **53**, 567–577 (2005).

25. Moldenhauer, K. K., Scott, B. & Hardke, J. *B.R. Wells Arkansas Rice Research Studies 2019*. https://scholarworks.uark.edu/aaesser/163 (2020).

26. Singh, V. *et al.* Impact of volunteer rice infestation on yield and grain quality of rice. *Pest Manag. Sci.* **73**, 604–615 (2017).

27. Singh, V. *et al.* Herbicide and winter flood treatments for controlling volunteer rice off-season. *Crop Prot.* **79**, 87–96 (2016).

28. Singh, V. *et al.* Introgression of Clearfield™ rice crop traits into weedy red rice outcrosses. *F. Crop. Res.* **207**, 13–23 (2017).

29. Burgos, N. R. *et al.* The Impact of Herbicide-Resistant Rice Technology on Phenotypic Diversity and Population Structure of United States Weedy Rice. *Plant Physiol.* **166**, 1208–1220 (2014).

30. Genomes, P. T. 3000 rice. The 3,000 rice genomes project. *Gigascience* **3**, 7 (2014).

31. Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).

32. Leung, H. *et al.* Allele mining and enhanced genetic recombination for rice breeding. *Rice* **8**, 1–11 (2015).

33. Upadhyaya, N. M. *Rice functional genomics: Challenges, progress and prospects*. *Rice Functional Genomics: Challenges, Progress and Prospects* (2007). doi:10.1007/0-387-48914-2.

34. Dias-Alves, T., Mairal, J. & Blum, M. G. B. Loter: A software package to infer local ancestry for a wide range of species. *Mol. Biol. Evol.* **35**, 2318–2326 (2018).

35. Sales, M. A., Shivrain, V. K., Burgos, N. R. & Kuk, Y. I. Amino Acid Substitutions in the Acetolactate Synthase Gene of Red Rice (*Oryza sativa*) Confer Resistance to Imazethapyr . *Weed Sci.* **56**, 485–489 (2008).

36. Gross, B. L., Skare, K. J. & Olsen, K. M. Novel *Phr1* mutations and the evolution of phenol reaction variation in US weedy rice (*Oryza sativa*). *New Phytol.* **184**, 842–850 (2009).

37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

38. Kawahara, Y. *et al.* Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 3–10 (2013).

39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,

2078–2079 (2009).

41. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

42. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

43. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

44. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

45. Bandelt, H. J., Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).

46. Leigh, J. W. & Bryant, D. POPART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

Tables

Table 1: Average proportions of contemporary genome called by the Loter software as derived from the weedy ancestor for BHA-like and SH-like samples across each of the 12 rice chromosomes..

Chromosome	BHA-like	SH-like
chr1	0.652485	
chr2	0.795564	0.658229
chr3	0.792826	0.658175
chr4	0.775346	0.696298
chr5	0.773475	0.732717
chr6	0.609169	0.769602
chr7	0.796445	0.652486
chr8	0.704422	0.736543
chr9	0.766718	0.544975
chr10	0.765744	0.791307
chr11	0.648896	0.6561
chr12	0.810042	0.727845
average	0.740928	0.692221

Table 2: Resistance levels of contemporary weedy rice samples. Samples are binned into four categories: high (0-32% average tissue damage), moderate (33-67% average tissue damage), susceptible (68-100% average tissue damage), and segregating. CHY, HYB, and CLF represent field cropping histories representing fields that historically grew Clearfield™ hybrid, non-Clearfield™ hybrid, and Clearfield™ inbred cultivars, respectively.

	Field Type			
Resistance (mean injury)	CHY (%)	HYB (%)	CLF (%)	Total (%)
High	14 (93.3)	12 (63.2)	8 (57.1)	34 (70.8)
Moderate	0 (0)	2 (10.5)	2 (14.3)	4 (8.3)
Susceptible	0 (0)	2 (10.5)	0 (0)	2 (4.2)
Segregating	1 (6.6)	3 (15.7)	4 (28.6)	8 (16.7)

Figures

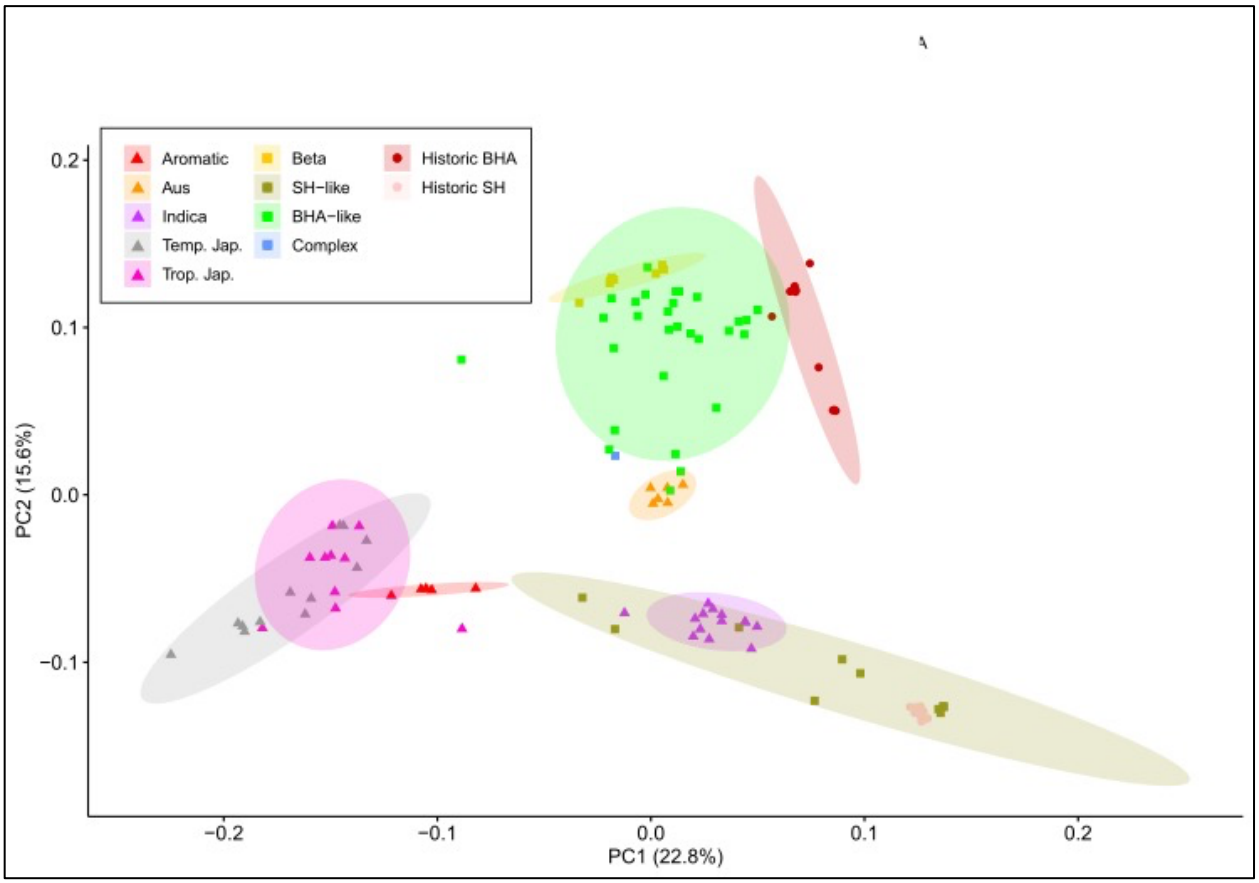


Figure 1: Principal component analysis of genome-wide SNPs in cultivated, historical weedy, and contemporary weedy rice. The first component seperates the japonica groups on the left and the indica groups on the right. The second component separates SH and SH-like weeds at the bottom from BHA and BHA-like weeds at the top. All hybrid weeds fall in between their presumed crop and weed ancestor, consistent with a hybrid origin. Shaded regions represent 95% confidence interval of placement of a theoretical sample.

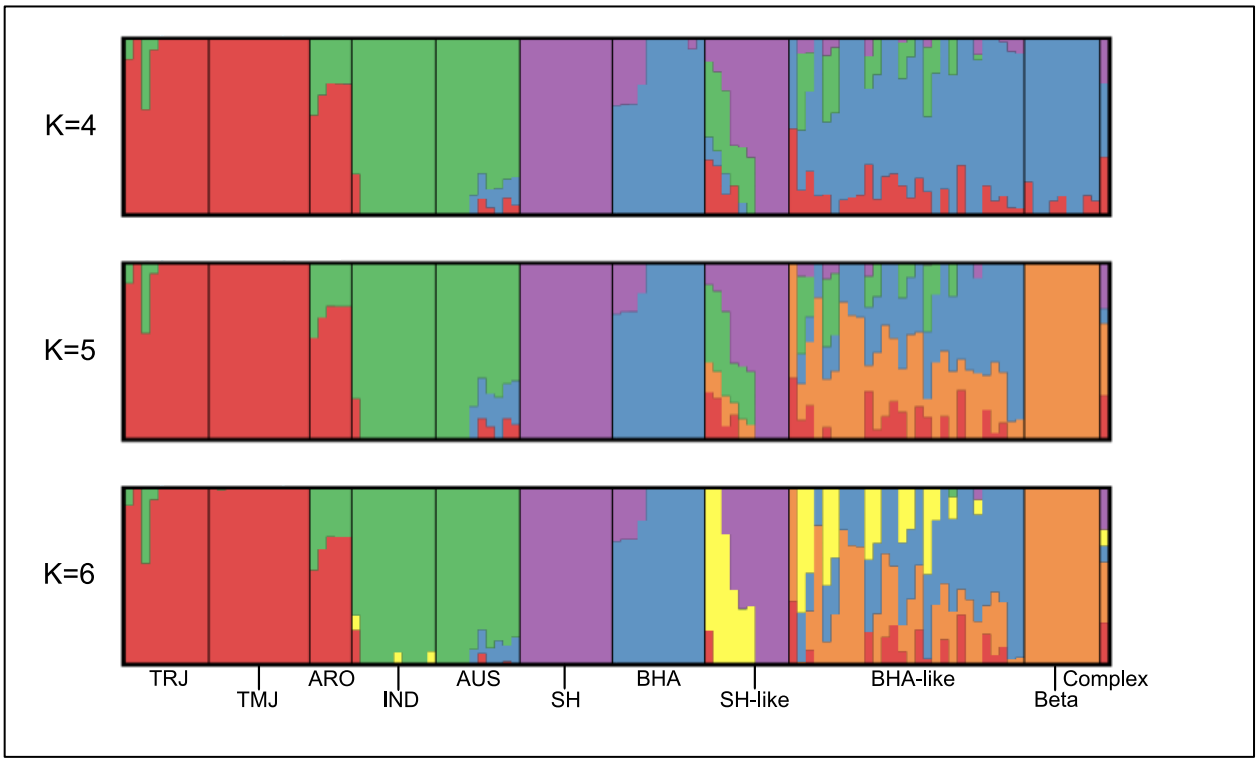


Figure 2: Results of ADMIXTURE analysis of contemporary weedy rice (SH-like, BHA-like, Beta, Complex) in comparison to historic weedy rice (SH, BHA) and cultivated rice (TRJ, tropical japonica; TMJ temperate japonica; ARO, aromatic; IND, indica; AUS, aus). Values of K at 4, 5, and 6 are shown; K = 6 is the optimal value based on cross-validation error. Categories for contemporary weeds are based on predominant weedy ancestry.

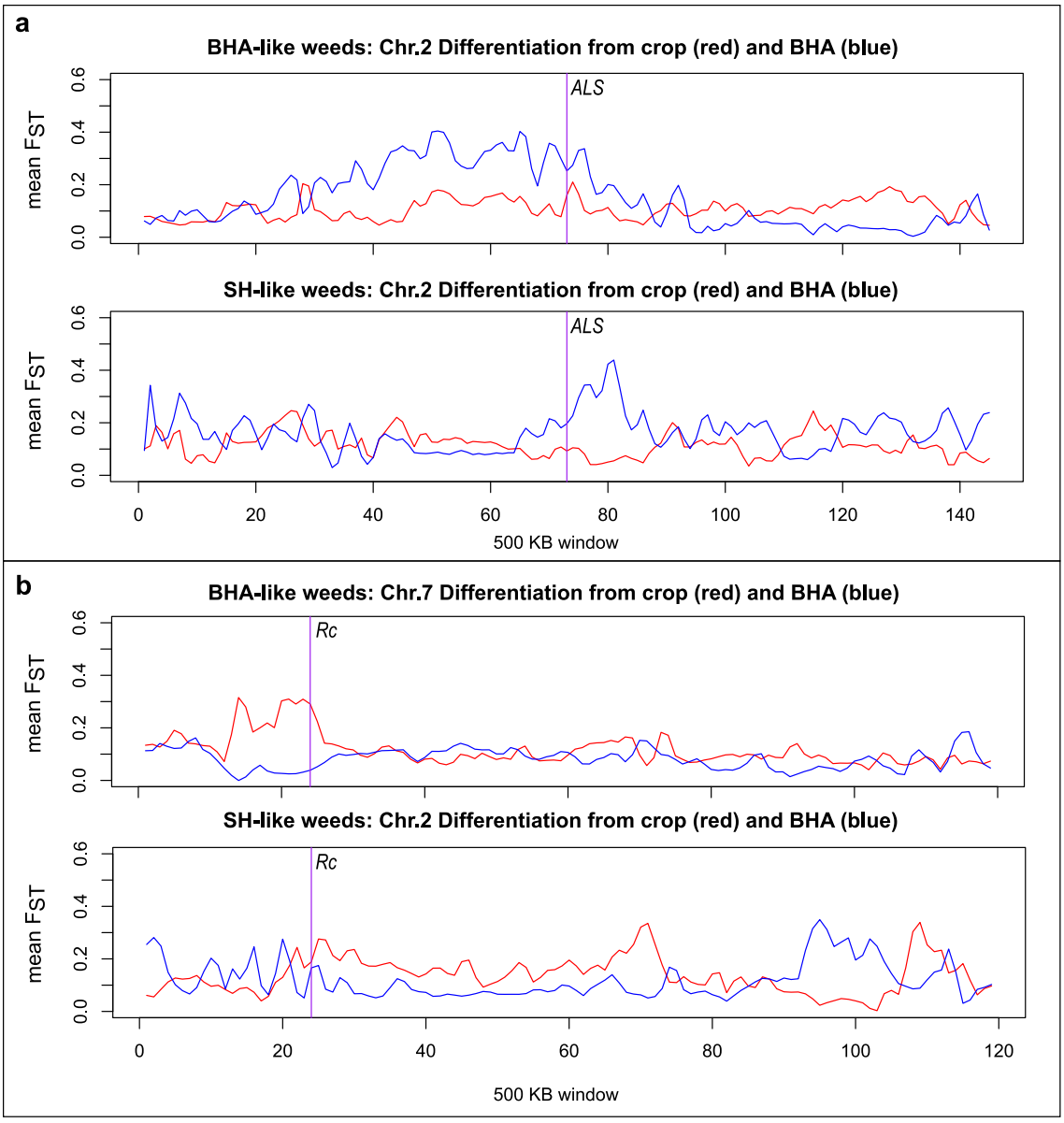


Figure 3: FST between contemporary weeds and their presumed ancestors in two chromosomes (chr. 2, a and chr. 7, b) containing genes associated with contemporary weed adaptation (ALS, herbicide resistance; Rc, seed dormancy). Red lines represent the FST between cultivated and contemporary weedy populations; blue lines represent FST between historical and contemporary weedy populations. The vertical purple lines denote the 500-kb window containing the focal gene.

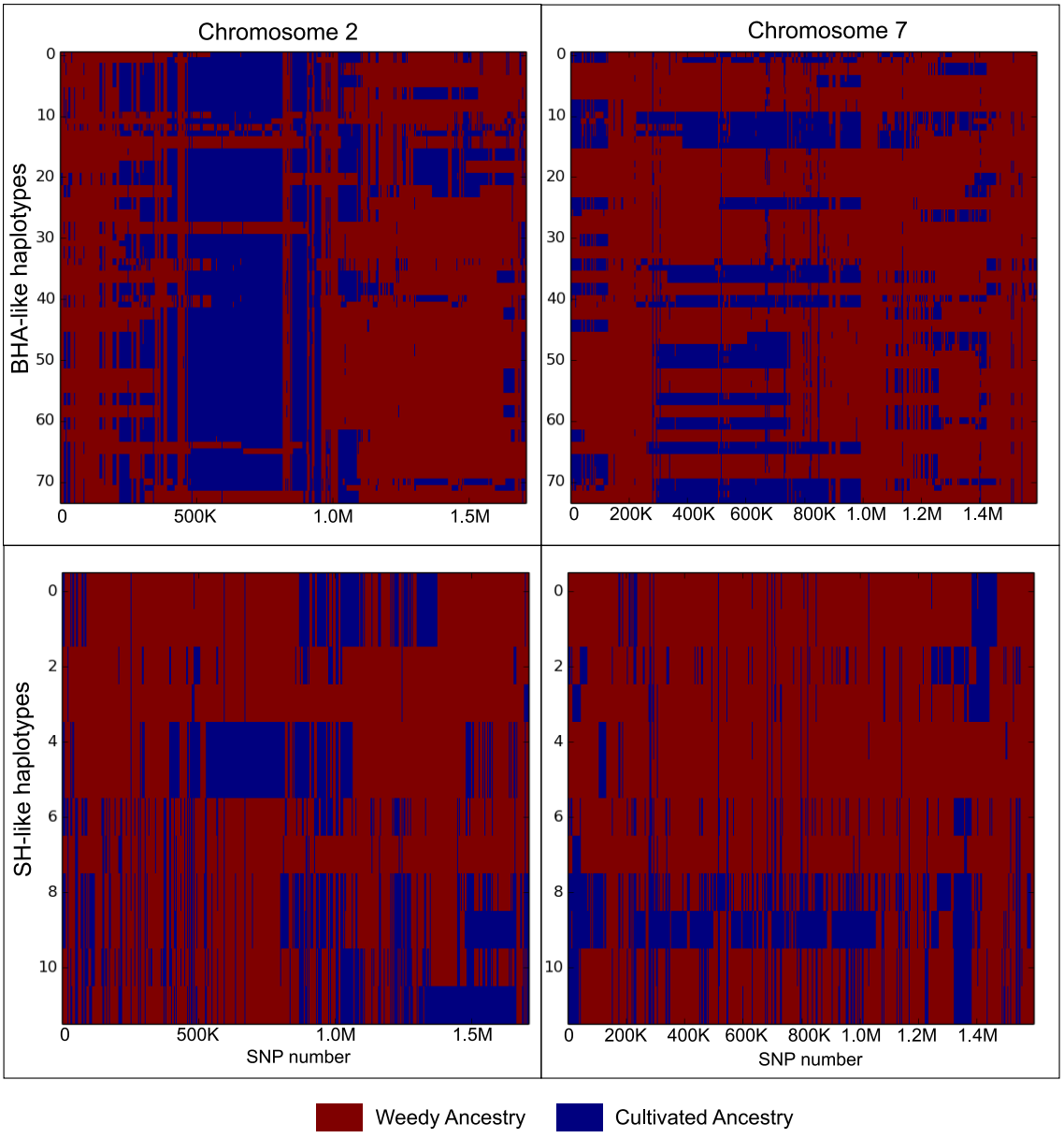


Figure 4: Local ancestry estimations based on Loter analysis across two rice chromosomes (2, 7) for each of two populations of hybrid-derived weedy rice. Each haplotype is plotted horizontally across the relevent chromosome. Blue areas denote crop-like regions of the genome, while red areas represent weed-like regions.

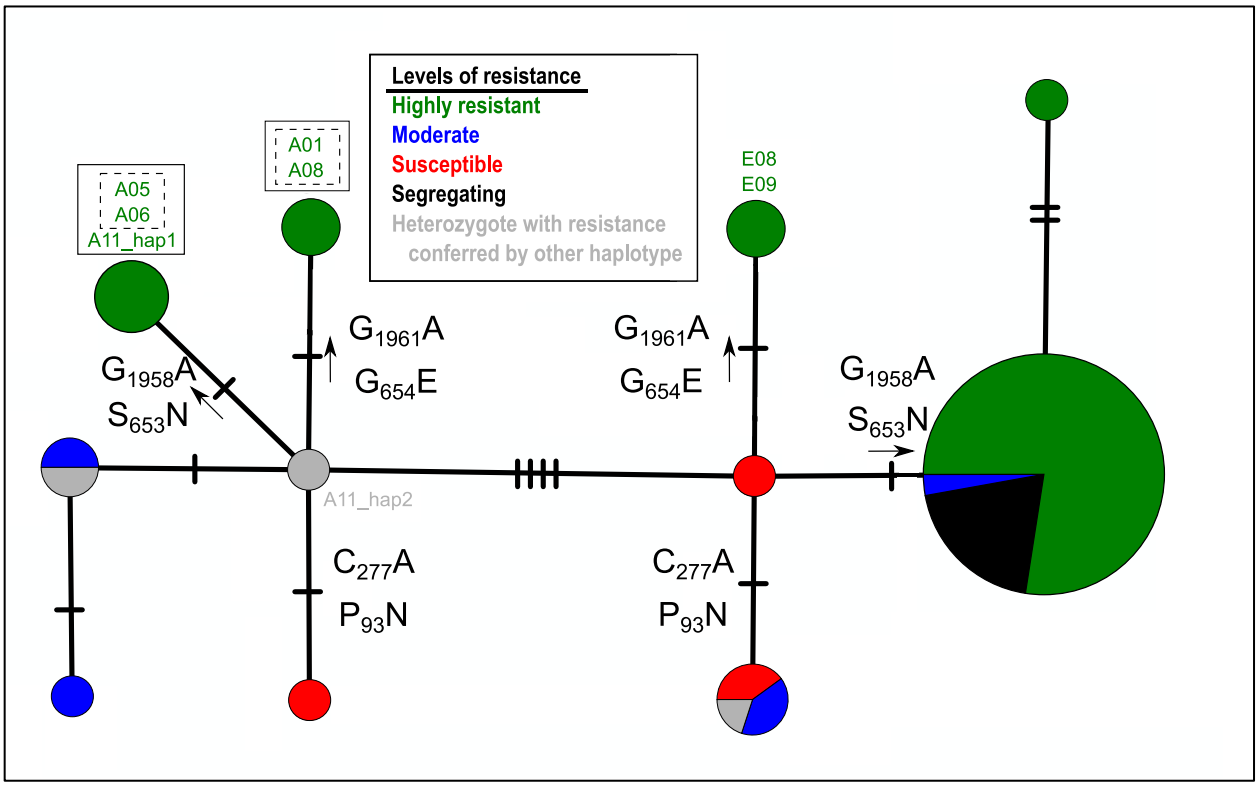


Figure 5: Median joining haplotype tree of the ALS herbicide resistance locus from contemporary weedy rice samples. Tree shown is one of four equally parsimonious arrangements (see Supplementary Fig. 6 for alternative topologies). Labeled mutational steps with arrows indicate gain-of-resistance mutations (nucleotide change and corresponding amino acid replacement). Sample names in boxes (A05, A06, A01, A08) are contemporary weed accessions that are not of crop × weed hybrid origin. Sizes of pie chart circles are proportional to haplotype numbers, and colors indicate proportions of herbicide resistance levels.

APPENDIX II

Chapter 3 Supplementary Material

Supplementary Table

Supplementary Table 1: List of all samples collected and used in this study.

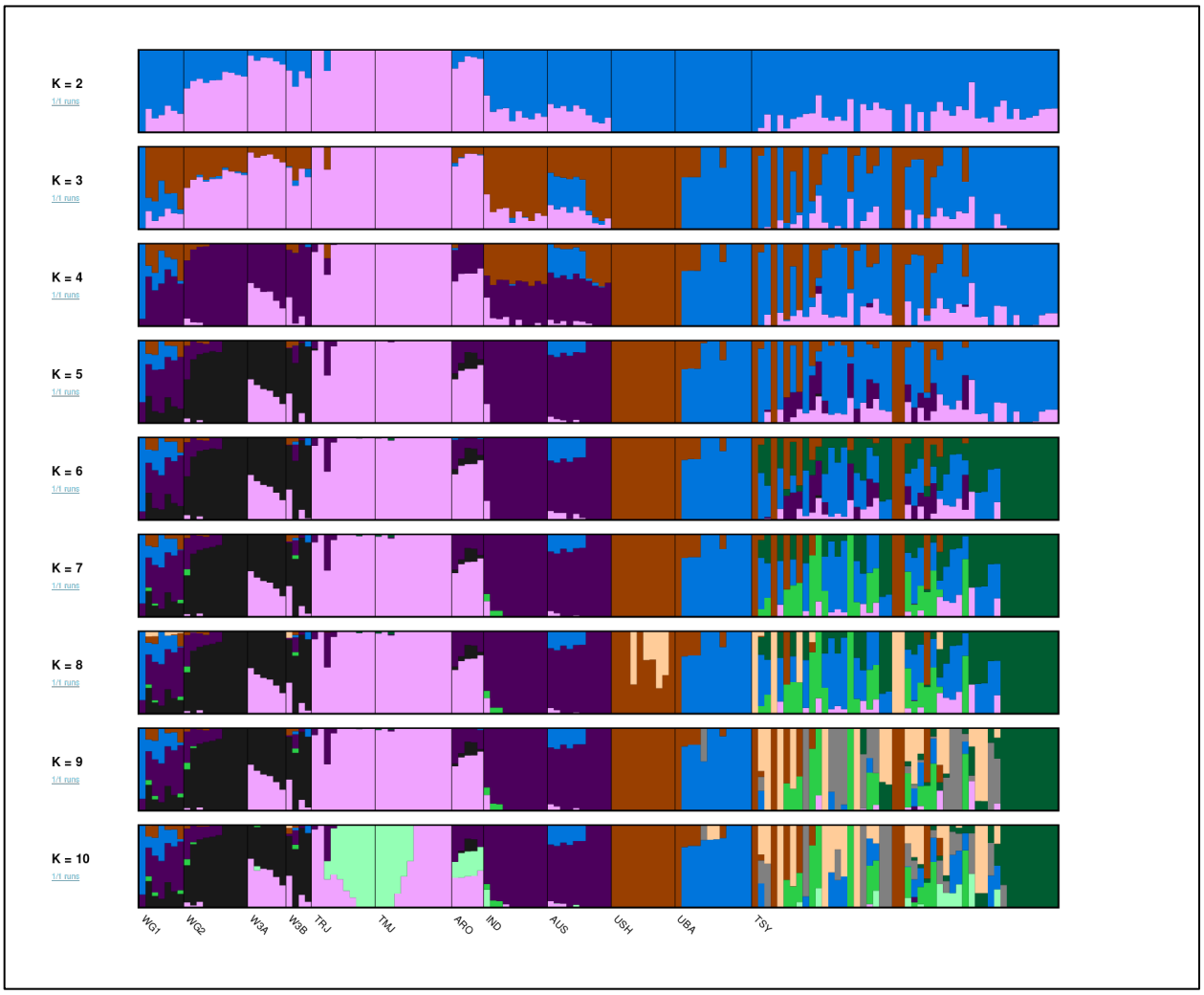
Sample Name	Rice Type	Sequence Platform	Reference	GenBank Accession Number
RR_A1	SH-like	HiSeq X 10	This Study	TBD
RR_A2	BHA-like	HiSeq X 10	This Study	TBD
RR_A3	BHA-like	HiSeq X 10	This Study	TBD
RR_A4	BHA-like	HiSeq X 10	This Study	TBD
RR_A5	SH-like	HiSeq X 10	This Study	TBD
RR_A6	SH-like	HiSeq X 10	This Study	TBD
RR_A7	BHA-like	HiSeq X 10	This Study	TBD
RR_A8	SH-like	HiSeq X 10	This Study	TBD
RR_A9	SH-like	HiSeq X 10	This Study	TBD
RR_A10	BHA-like	HiSeq X 10	This Study	TBD
RR_A11	SH-like	HiSeq X 10	This Study	TBD
RR_A12	Complex	HiSeq X 10	This Study	TBD
RR_A13	SH-like	HiSeq X 10	This Study	TBD
RR_A14	BHA-like	HiSeq X 10	This Study	TBD
RR_A16	BHA-like	HiSeq X 10	This Study	TBD
RR_B1	BHA-like	HiSeq X 10	This Study	TBD
RR_B2	BHA-like	HiSeq X 10	This Study	TBD
RR_C1	BHA-like	HiSeq X 10	This Study	TBD
RR_C2	SH-like	HiSeq X 10	This Study	TBD
RR_C3	SH-like	HiSeq X 10	This Study	TBD
RR_C4	BHA-like	HiSeq X 10	This Study	TBD
RR_C5	BHA-like	HiSeq X 10	This Study	TBD
RR_C6	BHA-like	HiSeq X 10	This Study	TBD
RR_C7	BHA-like	HiSeq X 10	This Study	TBD
RR_C8	BHA-like	HiSeq X 10	This Study	TBD
RR_C9	BHA-like	HiSeq X 10	This Study	TBD
RR_C10	SH-like	HiSeq X 10	This Study	TBD
RR_C11	BHA-like	HiSeq X 10	This Study	TBD
RR_D1	BHA-like	HiSeq X 10	This Study	TBD
RR_D2	BHA-like	HiSeq X 10	This Study	TBD
RR_D3	BHA-like	HiSeq X 10	This Study	TBD
RR_D4	BHA-like	HiSeq X 10	This Study	TBD
RR_D5	BHA-like	HiSeq X 10	This Study	TBD
RR_D6	BHA-like	HiSeq X 10	This Study	TBD
RR_E1	Beta	HiSeq X 10	This Study	TBD
RR_E2	Beta	HiSeq X 10	This Study	TBD

RR_E3	Beta	HiSeq X 10	This Study	TBD
RR_E4	BHA-like	HiSeq X 10	This Study	TBD
RR_E5	Beta	HiSeq X 10	This Study	TBD
RR_E6	Beta	HiSeq X 10	This Study	TBD
RR_E7	Beta	HiSeq X 10	This Study	TBD
RR_E8	BHA-like	HiSeq X 10	This Study	TBD
RR_E9	BHA-like	HiSeq X 10	This Study	TBD
RR_E10	Beta	HiSeq X 10	This Study	TBD
RR_E11	Beta	HiSeq X 10	This Study	TBD
RR_E12	BHA-like	HiSeq X 10	This Study	TBD
RR_E13	Beta	HiSeq X 10	This Study	TBD
RR_E14	BHA-like	HiSeq X 10	This Study	TBD
1995-15	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513411
10A	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513410
1333-02	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513409
1199-01	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513408
1996-05	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513406
1210-05	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513405
1025-01	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513404
1344-02	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513403
1995-12	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513402
TX4	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513401
LA3	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513400
StgS	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513399
PrCoTall1	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513398
PrCoSrt1	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513397
1995-14	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513395
1995-13	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513394
1214-02	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513393
1190-01	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513392
1188-01	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513391
1179-01	WeedBHA	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513390
1160-01	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513388
1141-01	WeedSH	Illumina HiSeq 2000	Li *et al*. 2017	SRR5513387
IRIS 313-11483	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR629787
IRIS 313-11602	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR630361
IRIS 313-11461	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR629691
IRIS 313-11453	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR629653
IRIS 313-11600	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR630355
IRIS 313-11603	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR630367
IRIS 313-11456	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR629673
IRIS 313-11453	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR629661

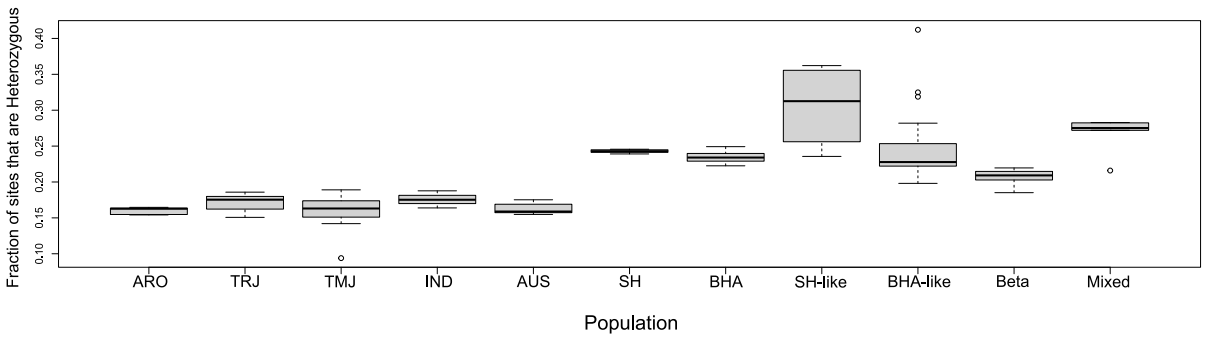
IRIS 313-9449	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR616368
IRIS 313-11741	*aus*	Illumina HiSeq 2000	3000 rice genomes project	ERR631936
IRIS 313-9682	*aromatic*	Illumina HiSeq 2000	3000 rice genomes project	ERR616657
IRIS 313-9170	*aromatic*	Illumina HiSeq 2000	3000 rice genomes project	ERR616053
IRIS 313-11626	*aromatic*	Illumina HiSeq 2000	3000 rice genomes project	ERR630481
IRIS 313-11625	*aromatic*	Illumina HiSeq 2000	3000 rice genomes project	ERR630475
IRIS 313-11630	*aromatic*	Illumina HiSeq 2000	3000 rice genomes project	ERR630505
IRIS 313-9066	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR615914
IRIS 313-11722	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR631605
IRIS 313-9262	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR616172
IRIS 313-8703	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608238
IRIS 313-11665	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630903
IRIS 313-11646	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630661
IRIS 313-11656	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630771
IRIS 313-11668	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630949
IRIS 313-11669	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630969
IRIS 313-9922	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR638200
DGWG	*indica*	Illumina HiSeq 2000	Leung et al., 2015	SRR6322106
IRIS 313-11681	*indica*	Illumina HiSeq 2000	3000 rice genomes project	ERR631100
IRIS 313-9048	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR615882
IRIS 313-11651	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR630713
IRIS 313-9239	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR616135
IRIS 313-8755	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608292
IRIS 313-8105	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR636780
IRIS 313-8202	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR637116
IRIS 313-8102	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR636774
IRIS 313-8856	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608364
IRIS 313-8669	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608202
IRIS 313-8180	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR637062
IRIS 313-8665	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608196
IRIS 313-8204	*temperate japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR637122
IRIS 313-11690	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR631238
IRIS 313-9470	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608780
IRIS 313-8400	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR607956
IRIS 313-8381	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR607904
IRIS 313-9996	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR609056
IRIS 313-8434	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR607994
IRIS 313-9550	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608804
IRIS 313-8072	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR636702
IRIS 313-8658	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608190
IRIS 313-8502	*tropical japonica*	Illumina HiSeq 2000	3000 rice genomes project	ERR608062
W0171	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068632

W1087	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068678
W1093	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068681
W1096	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068682
W1559	wild group I	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068722
W1683	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068733
W1715	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068741
W1725	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068748
W1748	wild group IIIb	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068764
W1757	wild group I	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068771
W1804	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068791
W2024	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068878
W2197	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068898
W2198	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068899
W3046	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068985
W3048	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068987
W3049	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068988
W3070	wild group IIIb	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR069009
W3072	wild group II	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR069011
W3105	wild group I	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR069038
W1943	wild group IIIa	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068843
W593	wild group IIIb	Illumina Genome Analyzer IIx	Huang et al., 2012	SRR1016473
W1963	wild group IIIb	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068849
W630	wild group I	Illumina Genome Analyzer IIx	Huang et al., 2012	SRR1016489
W1866	wild group I	Illumina Genome Analyzer IIx	Huang et al., 2012	ERR068819
IRGC102463	wild group I	Illumina HiSeq 2000	OMAP clone end sequence project	SRR1450138
IRGC105751	wild group I	Illumina HiSeq 2000	OMAP clone end sequence project	SRR1450141

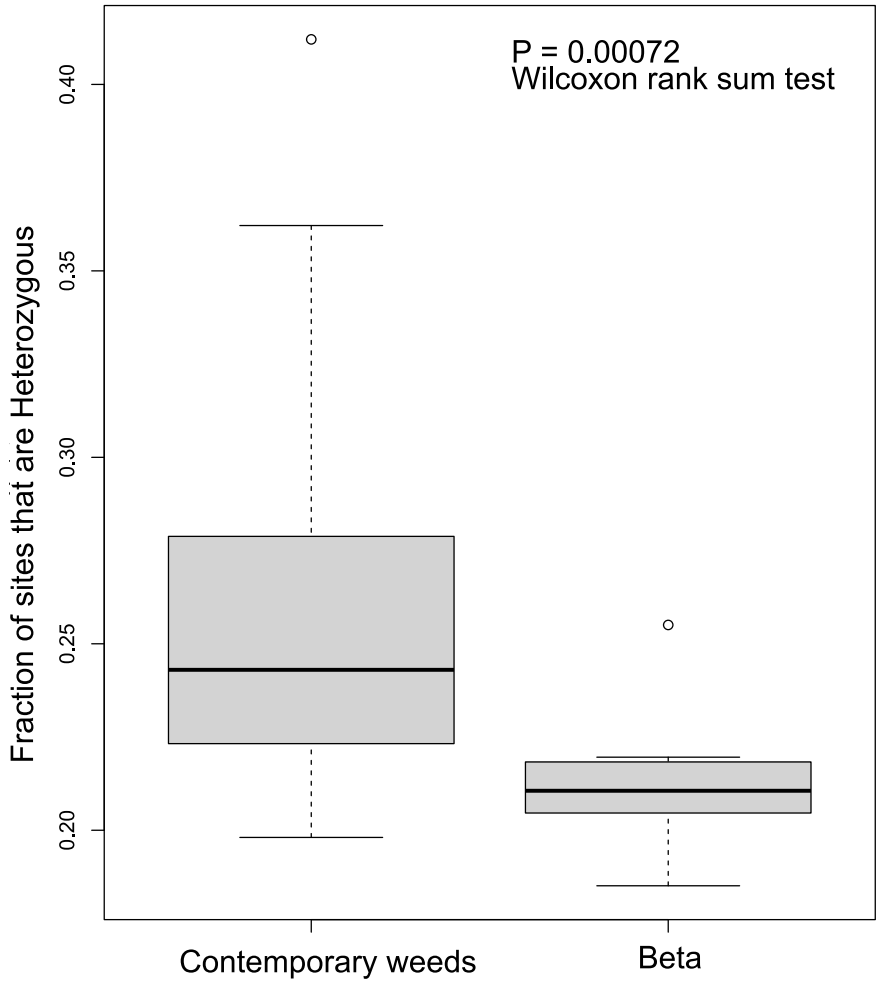
Supplementary Figures



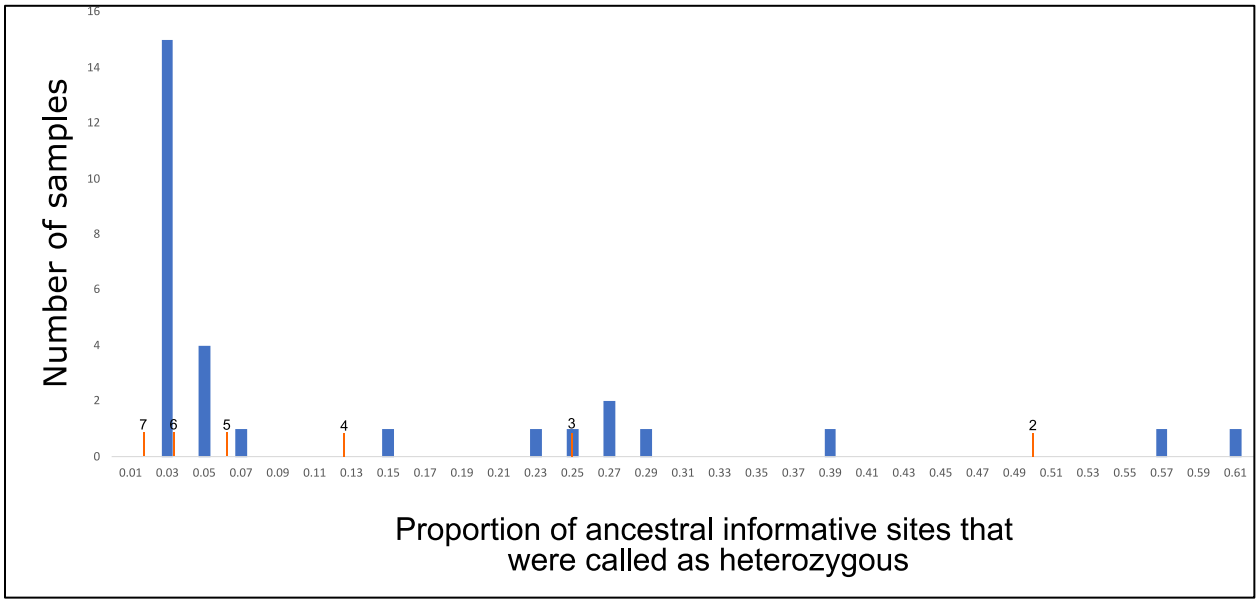
Supplementary Figure 1: ADMIXTURE results at K=2-10 for contemporary weeds collected for this study (TSY) in comparison to historic weedy rice (SH, BHA), cultivated rice (TRJ, tropical japonica; TMJ temperate japonica; ARO, aromatic; IND, indica; AUS, aus), and wild rice (WG1, WG2, W3A, W3B).



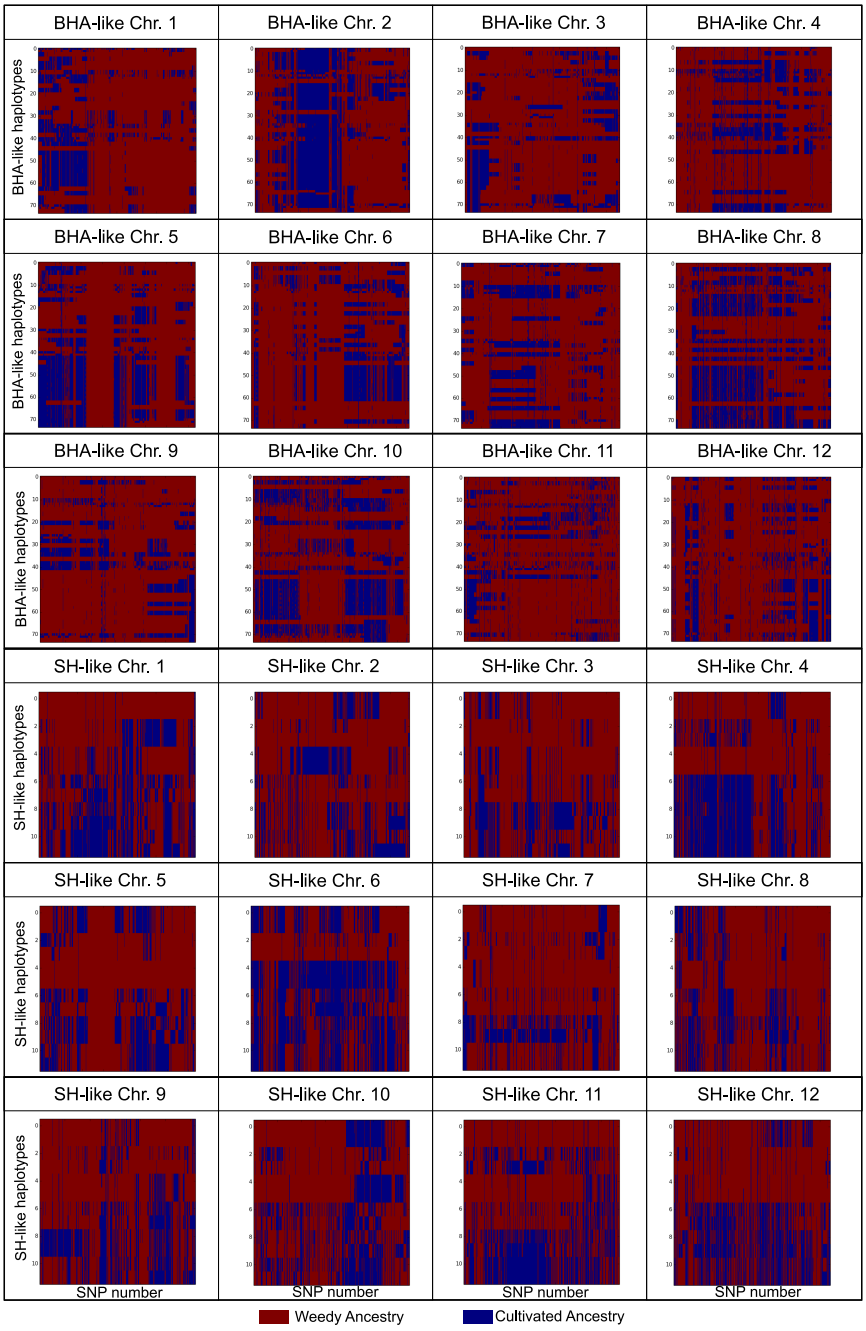
Supplementary Figure 2: Proportion of SNPs that were called as heterozygous among populations of cultivated rice (TRJ, tropical japonica; TMJ temperate japonica; ARO, aromatic; IND, indica; AUS, aus), historical weedy rice (SH, BHA), and contemporary weedy rice of hybrid origin (excluding the 'complex' sample) (SH-like, BHA-like, and Beta). Box plots show median, inter-, and outerquartile ranges of all samples in the population.



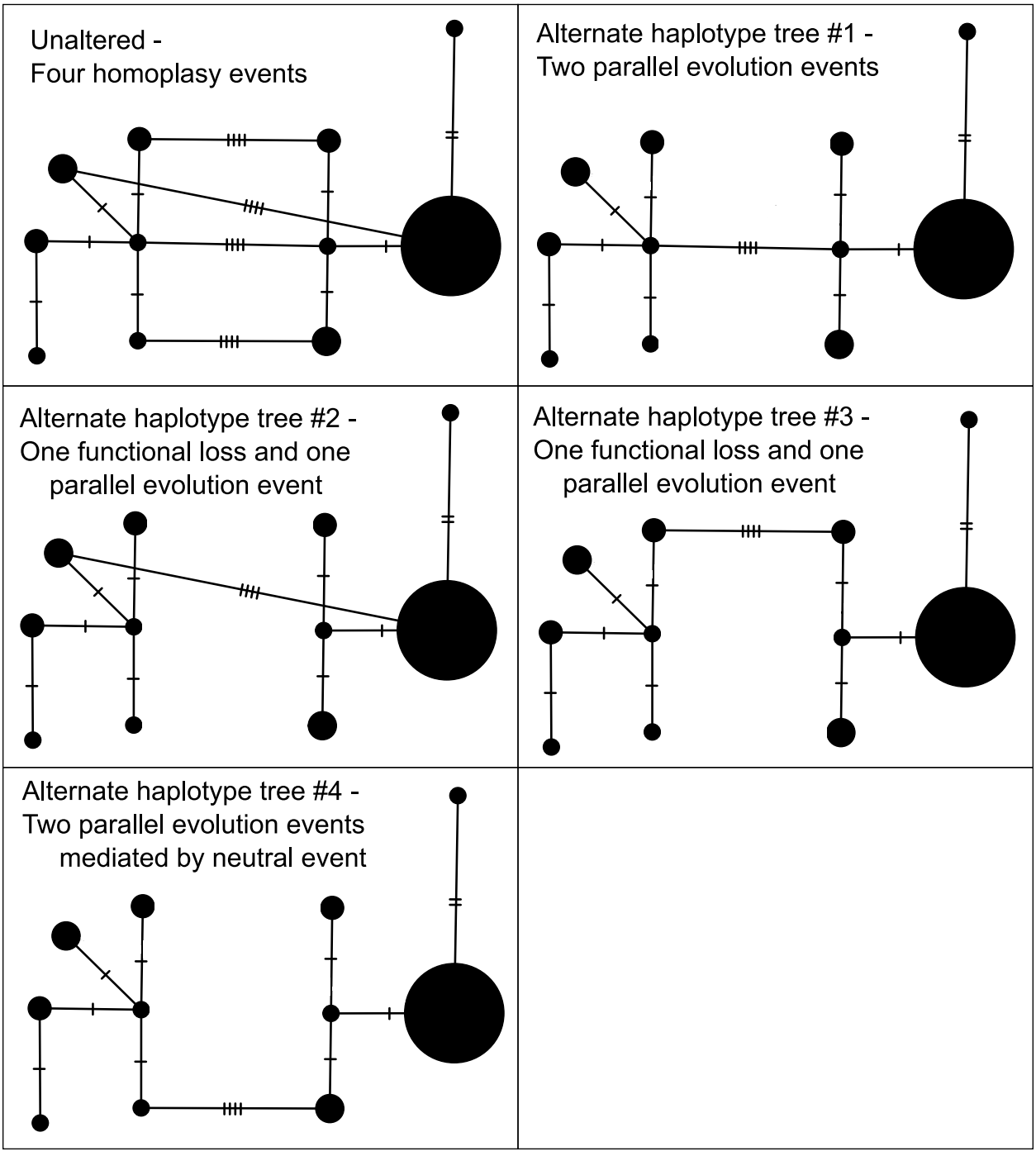
Supplementary Figure 3: Fraction of sites called as heterozygous in Beta population against the rest of the contemporary weeds. Beta weeds have significantly fewer heterozygous sites, which suggests many more generations since the hybridization event.



Supplementary Figure 4: Histogram of the proportion of ancestrally informative sites that were called as heterozygous. Thick blue bars represent the number of samples that fell into 0.02-unit wide bins. Thin orange lines represent the expectation of heterozygosity loss per generation with selfing under neutral genetic drift, where 50% of heterozygosity is lost per generation. The number above the thin orange lines represents the expected generation.



Supplementary Figure 5: Loter output for BHA-like and SH-like weeds across all 12 rice chromosomes. Red represents genomic locations derived from ancestral weedy rice, while blue represents ancestry from cultivated rice.



Supplementary Figure 6: All possible median joining haplotype trees for the ALS locus including the unaltered tree and four alternative trees with homoplasies removed. Mutational scenarios are provided for each tree.

CHAPTER FOUR

Discordant patterns of introgression suggest historical gene flow into Thai weedy rice from domesticated and wild relatives

Abstract

Weedy relatives of crop species infest agricultural fields worldwide, reducing harvests and threatening global food security. These weeds can potentially evolve and adapt through gene flow from both domesticated crop varieties and reproductively-compatible wild relatives. We studied populations of weedy rice in Thailand to investigate the role of introgression from cultivated and wild rice in their evolution. We examined two complementary sources of genetic data: allelic variation at three rice domestication genes (*Bh4*, controlling hull color; *Rc*, controlling pericarp color and seed dormancy; and *sh4*, controlling seed shattering), and 12 previously-published SSR markers. Sampling spanned three major rice growing regions in Thailand (Lower North, North East, and Central Plain) and included 124 cultivated rice accessions, 166 weedy rice accessions, and 98 wild rice accessions. Weedy rice strains were overall closely related to the cultivated varieties with which they co-occur. Domestication gene data revealed potential adaptive introgression of *sh4* shattering alleles from wild rice. Introgression of potentially maladaptive *rc* crop alleles (conferring reduced dormancy) was also detected, with the frequency of the crop allele highest in northern populations. Although SSR markers also indicated introgression into weed populations from wild and cultivated rice, there was little overlap with domestication genes in the accessions showing admixed ancestry. This suggests that much of the introgression we detected at domestication genes most likely reflects past introgression rather than recent gene flow. This finding has implications for understanding long-term gene flow dynamics between rice and its weedy and wild relatives, including potential risks of transgene escape.

Introduction

Agricultural weeds that are closely related to crop species are present in agroecosystems worldwide and pose a major threat to sustainable crop production (Ellstrand *et al.,* 2010; Singh *et al.,* 2013; Ziska *et al.,* 2015). These weedy crop relatives are commonly

restricted to agricultural habitats, where they aggressively outcompete crop varieties and can reduce harvests by 80% or more (Diarra *et al.,*1985; Singh *et al.,* 2013). An important question in the study of weedy crop relatives is the extent to which their evolution and adaptive fitness is shaped by gene flow from co-occurring domesticated varieties and/or nearby populations of reproductively compatible wild relatives (Beebe *et al.,* 1997; Warwick *et al.,* 2008; Ellstrand *et al.,* 2010; Engku *et al.,* 2016). This question is often examined in the context of transgene escape, with recent studies largely focused on the contemporary movement of herbicide resistance alleles from transgenic crops into nearby wild and weedy populations (Warwick *et al.,* 2008; Singh *et al.,* 2017b). Most such studies document transgene escape but do not assess the multigeneration impact of this crop-to-weed introgression (Morrell *et al.,* 2005). Thus, less is known about the longer-term consequences of hybridization and gene flow between cultivated, weedy and wild populations. From a practical perspective, introgression into weeds can elevate their competitive advantage, leading to strains that are much more difficult to control. It is therefore imperative to understand the evolutionary influence of these types of introgression and the timescale over which they occur.

A potentially useful approach for studying the long-term dynamics of gene flow into weedy crop relatives is to examine allelic variation at genes that control domestication-related traits. Because weedy relatives are specifically adapted to agroecosystems, some domestication traits would be expected to confer fitness benefits to weed strains; these include erect plant growth architecture and short stature (allowing weeds to grow competitively and inconspicuously in agricultural fields), as well as herbicide resistance. For such traits, the domestication (crop) alleles at the genes controlling these traits would be adaptive in weed populations. For other domestication traits, including reduced seed shattering, reduced seed dormancy and loss of structures promoting secondary seed dispersal (e.g., awns and barbs), crop alleles are likely maladaptive. For such traits, introgression from wild populations rather

than crop varieties could be adaptive for allowing the weeds to persist and proliferate in agricultural fields. Comparisons of the distributions of crop vs. wild alleles at multiple domestication genes can thus provide insights on patterns of adaptive introgression into weed populations from domesticated and wild relatives (Song *et al.,* 2014; Cui *et al.,* 2016; Huang *et al.,* 2018).

As a complement to the gene-specific insights provided by domestication genes, genome-wide neutral markers can help to elucidate the broader genomic consequences of gene flow into weedy crop relatives. Depending on the frequency at which hybridization has occurred, and whether hybridization occurred recently or in the more distant past, the genetic composition of weedy relatives is expected to show greater or lesser overall levels of relatedness to the hybridizing source populations. Neutral markers can thus be informative for determining whether gene flow occurred extensively and in the recent past — in which case the weeds would show genome-wide evidence of admixture from the source population — or whether introgression occurred enough generations ago that evidence of the hybridization event is no longer apparent on a genome-wide scale.

In recent years, weedy rice (*Oryza sativa*) has emerged as a genomic model system for studying the evolution of weedy crop relatives (Guo *et al.,* 2018; Wedger and Olsen, 2018). Weedy rice is a conspecific form of cultivated Asian rice that is present in almost every world region where rice is cultivated, including both areas where the wild crop ancestor (*O. rufipogon*) is present (South and Southeast Asia), and areas without reproductively compatible wild relatives (e.g., Japan, North America, Europe) (Cao et al., 2006b; Londo & Schaal, 2007; Grimm et al., 2013). Weedy rice has evolved multiple times independently from different cultivated rice varieties, making the system highly amenable to studies on the parallel evolution of weediness (Qi *et al.,* 2015; Li *et al.,* 2017) In some rice growing regions, including Japan, Italy, and China (Akasaka *et al.,* 2009; Grimm *et al.,* 2013; Sun *et al.,* 2013), weedy rice strains are closely

related to local rice varieties, suggesting *in-situ* origins by de-domestication. In other regions, such as the United States, the weeds are genetically distinct from local crop varieties and likely evolved through de-domestication in Asia, with subsequent unintentional introductions into their present range (Reagon *et al.,* 2010; Li *et al.,* 2017). In areas of tropical Asia where wild rice is present, weedy rice strains have typically been found to show some evidence of introgression from wild populations, although they are still primarily descended from domesticated rice (Cao *et al.,* 2006; Song *et al.,* 2014; Huang *et al.,* 2017).

Comparative analyses of domestication genes and neutral markers have proved particularly insightful in evolutionary studies of weedy rice (Song *et al.,* 2014; Cui *et al.,* 2016; Huang *et al.,* 2018). These analyses have largely relied on three well-characterized rice domestication genes: *sh4* (controlling loss of shattering in the crop), *Rc* (controlling loss of pericarp pigmentation and seed dormancy in the crop) and *Bh4* (controlling loss of dark-pigmented hulls in the crop). In the case of *sh4*, strong selection during rice domestication led to the fixation in the crop of a nonsynonymous substitution in exon 1 that results in a reduction in grain shattering (Li *et al.,* 2006; Zhang *et al.,* 2009). Most weedy rice strains examined to date carry this domestication allele, confirming descent from domesticated ancestors (Thurber *et al.,* 2010; Zhu *et al.,* 2012). Despite carrying the reduced-shattering allele, however, weedy rice strains are typically highly shattering, and the re-emergence of the shattering phenotype appears to have occurred through multiple compensatory mutations throughout the genome (Qi *et al.,* 2015; Li *et al.,* 2017). In Southeast Asia, some weedy rice strains carry the wild *sh4* allele, a pattern consistent with adaptive introgression from local wild rice populations (Song *et al.,* 2014; Huang *et al.,* 2018).

Rc encodes a bHLH protein that pleiotropically controls both the proanthocyanidin pigment synthesis pathway and abscisic acid-mediated seed dormancy (Sweeney *et al.,* 2006; Gu *et al.,* 2011). Most modern cultivated rice varieties carry a 14-bp frameshift deletion in exon

7 that generates a nonfunctional gene product and non-pigmented or 'white' pericarps (bran) (Sweeney *et al.,* 2007). Unlike *sh4*, this *rc* domestication allele is not present in most weedy rice; instead, most weed strains carry functional *Rc* alleles (Gross *et al.,* 2010; Cui *et al.,* 2016). This suggests that most weedy rice strains are not descended from modern rice varieties, but rather that they evolved from dark-pericarp landraces that pre-date modern light-pericarp varieties. The high frequency of the functional *Rc* allele in weedy rice populations has been proposed to reflect strong selection for seed dormancy, as this is a critical trait for weed persistence in crop fields (Cui *et al.,* 2016).

Bh4 encodes an amino acid transporter that is expressed in maturing rice hulls and generates the dark hull pigmentation that characterizes wild *Oryza* species. Most cultivated rice varieties carry a 22-bp frameshift deletion in exon 3 that results in the straw-hull phenotype of domesticated rice (Zhu *et al.,* 2011). Among weedy rice strains, both straw- and black-hull strains occur widely (Reagon *et al.,* 2010; Grimm *et al.,* 2013; Song *et al.,* 2014; Merotto *et al.,* 2016), with the former carrying the domestication allele and the latter carrying the functional *Bh4* allele of wild *Oryzas*. The widespread occurrence of both phenotypes in weedy rice has led to the hypothesis that this variation may represent two adaptive weed strategies: a crop-mimic (straw-hull) form, and a more wild-like (black-hull) form (Federici *et al.,* 2001). Alternatively, this variation may simply reflect a lack of strong selection on hull color in weedy rice, with the two forms present as a legacy of independent weed origins from straw-hull and black-hull rice ancestors (Vigueira *et al.,* 2013; Li *et al.,* 2017).

In this study, we examined the distributions of wild and crop alleles at *Bh4*, *Rc*, and *sh4* to study patterns of adaptive introgression into weedy rice in Thailand, a region where both cultivated rice and local wild rice populations may contribute to weed evolution (Pusadee *et al.,* 2013; Wongtamee *et al.,* 2017). We then compared these patterns to genome-wide patterns inferred from previously reported neutral SSR loci (Wongtamee *et al.,* 2017) to assess the time

frame over which introgression has occurred. Thailand lies in the center of diversity for rice domestication. Additionally, Thailand is among the few rice growing countries where the wild progenitor is still abundant and present at the margins of fields. This wild-weed-crop complex allows for interactions among the three components and suggests that rice in Thailand forms an evolutionarily dynamic system.

We specifically asked the following questions: 1) Is there evidence of gene flow from wild or cultivated rice into co-occurring Thai weedy rice populations? 2) Do weeds that show evidence of crop allele introgression at domestication genes show increased genome-wide similarity to the crop based on SSR markers? 3) Do weeds that show evidence of adaptive introgression of wild alleles show genome-wide evidence of wild rice ancestry? Our results suggest that introgression into weedy rice has occurred from both wild and cultivated rice, but that this is likely a historical process with relatively little gene flow occurring on a contemporary time scale.

Methods

Sampling. *Oryza* leaf samples were obtained from three geographical regions of rice cultivation in Thailand: the North East (NE), Lower North (LN), and Central Plain (CP) (Fig. 1). Samples included 166 weedy rice accessions (40 NE, 77 LN, 49 CP), 104 co-occurring cultivated rice accessions (10 NE, 54 LN, 40 CP), and 28 common wild rice accessions collected from natural habitats spanning the three geographical regions (Supporting information, Table S1). Here we use the term 'accession' to refer to individual rice plants and their derived seed. We reserve 'populations' for genetically distinct subgroups inferred from *STRUCTURE* analyses described below. Weedy rice accessions were collected by randomly selecting plants separated by 5-10 m intervals (to avoid collecting close relatives) from heavily infested

agricultural fields (>50% infestation by visual inspection). Only *indica* rice varieties are cultivated in the sampled rice growing regions.

For weedy rice and wild rice collections, leaves and panicles of sampled plants were collected in the field and silica-dried following the method of Chase and Hills (2013). For cultivated rice samples, seeds were collected and germinated in petri dishes for one week and then transplanted to outdoor field plots at Chiang Mai University, with ten plants per plot. Four weeks after transplanting, leaves of ten individuals of each variety were harvested and dried in silica gel for DNA extraction.

Genotyping. DNA was extracted from leaf tissue at Chiang Mai University using a modified cetyltrimethyl ammonium bromide (CTAB) protocol from Doyle and Doyle (1987). Genotyping of domestication genes was performed at Washington University in St. Louis as described below.

Bh4. PCR genotyping was used to score all plants in the study for the presence/absence of the 22-bp deletion that distinguishes straw-hull rice (the common phenotype in most cultivated rice) from the black-hull phenotype (characteristic of wild rice). Four PCR primers, *Bh4-22F1*, *Bh4-22R1*, *Bh4_gt2F*, and *Bh4_gt2R*, were designed for this purpose (Supplementary Information, Table S2). Thermocycler conditions were as follows: denaturation at 94 ˚C for 2 minutes, followed by 40 cycles of denaturation at 94 ˚C for 30 seconds, annealing at 53 ˚C for 30 seconds, and elongation at 72 ˚C for 30 seconds. PCR was finished with elongation at 72 ˚C for 7 minutes and held at 4 ˚C. Reactions were conducted at standard PCR concentrations with GoTaq (Promega) and 1M betaine added to reduce secondary structure formation. PCR amplifications were visualized and scored with ethidium bromide on a 2.5% agarose gel. A functional ‘black hull’ allele would appear as a 114 bp band, whereas a non-functional ‘straw hull’ allele would appear as a 92 bp band. Results were spot

checked for accuracy by direct Sanger sequencing of PCR products (using primers *Bh4_gt2F* and *Bh4_gt2R*). Sequencing was performed on an ABI 3130 capillary sequencer in the sequencing facility of the Washington University Biology Department.

***Rc*.** A 14-bp frameshift deletion allele is the primary cause of the non-pigmented ('white') pericarp seen in most cultivated rice. Samples were genotyped for the presence/absence of the 14-bp deletion in one of two ways. For the first method, three primers, *Rc_wtF*, *Rc_delF*, and *Rc_gtR* (Table S2), were designed and used together in PCR. Thermocycler conditions were as follows: denaturation at 94 ˚C for 2 minutes followed by 40 cycles of denaturation at 94 ˚C for 30 seconds, annealing at 48 ˚C for 30 seconds, and elongation at 72 ˚C for 30 seconds. PCR was finished with elongation at 72 ˚C for 7 minutes and held at 4 ˚C. Reactions were performed with PlatinumTaq (Invitrogen) and 1M Betaine for precision and stability. PCR amplifications were visualized and scored with Ethidium Bromide on a 2.5% agarose gel. A functional 'red' *Rc* allele would appear as a 175 bp band, a non-functional 'white' rc allele would appear as a 155 bp band, and any heterozygous genotypes would amplify both products.

The second method for scoring *Rc* was based off the protocol of Rysbekova *et al.,* (2017) and used two sets of primer pairs: *Rc_wtF1* with *Rc_wtR1*, and *Rc_delF3* with *Rc_delR3* (Table S2). Thermocycler conditions for both reactions were as follows: denaturation at 94 ˚C for 2 minutes followed by 40 cycles of denaturation at 94 ˚C for 30 seconds, annealing at 54 ˚C for 30 seconds, and elongation at 72 ˚C for 30 seconds. PCR was finished with elongation at 72 ˚C for 7 minutes and held at 4 ˚C. Reactions for each primer set were conducted separately. Reactions with *Rc_wtF1* and *Rc_wtR1* were conducted with ExTaq and 2mM $MgCl_2$ to increase amplification. Reactions with *Rc_delF3* and *Rc_delR3* were conducted with ExTaq and 3 mM $MgCl_2$ to further increase amplification. PCR products were visualized and scored on a 0.8%

agarose gel. A non-functional 'white' allele would appear as a 400 bp band, while a functional 'red' allele would appear as an 800 bp band.

sh4. Two primers, *Sh4_00F* and *Sh4_00R* (Table S2) were used to PCR-amplify a portion of the gene for Sanger sequencing to genotype the domestication SNP (a G to T substitution in exon 1). The T substitution results in reduced shattering in cultivated rice and is present at 100% frequency in the crop. Thermocycler conditions for both initial PCRs were as follows: denaturation at 94 ˚C for 2 minutes, followed by 40 cycles of denaturation at 94 ˚C for 30 seconds, annealing at 58 ˚C for 30 seconds, and elongation at 72 ˚C for 1 minute. PCR was finished with elongation at 72 ˚C for 7 minutes and held at 4 ˚C. Reactions were conducted at standard PCR concentrations with ExTaq and 1M Betaine for precision and stability. Resultant PCR products underwent a further sequencing reaction consisting of 5 µl template, 2 µl of forward or reverse primer, and betaine. Thermocycler conditions were as follows: 96°C for 1 minute followed by 30 cycles of 96°C for 10 seconds, 50°C for 5 seconds, and 60°c for 1 minute. Samples were then held at 4°C until sequencing. PCR products were sequenced on an ABI 3130 capillary sequencer at Washington University and visualized using Geneious v. 8.0 (http://www.geneious.com, Kearse *et al.,* 2012).

***Data analysis*.**

SSR loci. Genotypes from twelve microsatellite loci, distributed across 10 of the 12 rice chromosomes, were obtained for all cultivated and weedy rice samples in this study from a previously published dataset (Wongtamee *et al.,* 2017) (Table S2); these data were used to assess population structure and genetic relationships among accessions (Table S3). Samples used in the study were chosen based on data availability from the previous study. Of the sampled accessions in Wongtamee *et al.* (2017), only those from fields with more than 10 accessions in that study were analyzed. SSR genotypes for an additional 20 cultivated and 70

wild rice SSR genotypes were obtained from the same study for inclusion in analyses. Population structure was first assessed using the Bayesian analysis in *STRUCTURE* (Pritchard *et al.,* 2000) at K values ranging from 2-10 with a burn-in of 50,000 MCMC replicates and a run length of 50,000 replicates. Default parameters were used to identify the optimal number of populations (K), with the delta-K statistic (Evanno *et al.,* 2005) used as the selection criterion for optimal K. A final *STRUCTURE* run was performed at the optimal K with a 500,000 burn-in length and 500,000 runs for final determination of population membership coefficients. Population membership coefficients were used as an indicator of ancestry to determine the extent to which a given accession unambiguously belonged to a population or showed evidence of genetic introgression from another group. Accessions with <80% membership assignment to a single population were considered to be admixed. As a complement to the *STRUCTURE* analysis, genetic relationships among accessions were further assessed by principal coordinates analysis (PCoA), using default parameters in GenAlEx (Peakall & Smouse, 2006, 2012).

Domestication genes. To assess the degree of concordance between domestication genes and SSRs for inferred introgression into weedy rice, weed accessions were separated into mutually exclusive groups based first on inferred population membership coefficients from *STRUCTURE*, and then on the distributions of wild and crop alleles at the three domestication genes. This allowed us to test the hypothesis that plants that showed introgression at domestication loci would also show differential similarity to the corresponding population at neutral loci.

Results

Domestication genes.

***Bh4*.** All cultivated rice plants that were genotyped for *Bh4* variation (104 of 104 accessions) carried the 22-bp deletion allele that encodes the straw-hull phenotype found in most cultivated rice (Table 1). Similarly, nearly 100% of the genotyped weedy rice plants (165 of 166 accessions) also carried the crop allele, consistent with weed descent from domesticated ancestors. The sole weedy rice plant with a wild *Bh4* allele (conferring black hull color) was collected in the Central Plain; this accession does not appear to be a descendant of recent wild-to-weed introgression (see *STRUCTURE* results below). Among the genotyped wild samples, most accessions carried the wild allele (25 of 28 accessions), with the remaining three carrying the domestication allele. This pattern suggests a low level of unidirectional gene flow from cultivated into wild rice, a pattern that has been previously reported and is likely prevalent in wild rice populations (Wang *et al.,* 2017).

***Rc*.** All but one of the genotyped cultivated rice accessions (100 of 101 accessions) carried the 14-bp deletion domestication allele that confers light-colored pericarps and is found in most rice varieties (Table 1). In the weedy rice samples, 134 of 158 genotyped accessions (84.8%) carried the functional wild *Rc* allele that confers dark pericarp pigmentation and seed dormancy, with the remaining 24 accessions carrying the light-pericarp *rc* allele. The frequency of the crop allele in weedy rice showed a general north-to-south decrease across the sampled regions (21.3% in the Lower North, 12.5% in the North East, and 7% in the Central Plain) (Fig. 1). This occurrence of the *rc* allele in weedy rice suggests that there has been introgression of the crop allele that is likely maladaptive for the weeds, since the functional *Rc* allele confers seed dormancy, an important trait for weed fitness. In the wild samples, genotyping could only be successfully performed in seven accessions; among these, two accessions (28.6%) carried the wild allele. As with *Bh4*, the identification of crop alleles in the wild samples supports an inference of crop-to-wild gene flow.

***sh4*.** The reduced-shattering *sh4* domestication allele, which is fixed in all cultivated rice (59 of 59 accessions), was present and homozygous in the majority of weedy rice accessions (95 of 111 accessions, or 85.6%) (Table 1); this pattern is consistent with descent from domesticated ancestors. The remaining 16 accessions (14.4%), one of which was a heterozygote, carried the wild allele. The presence of the wild *sh4* allele in weedy rice has been observed in other regions of tropical Asia where wild rice is present (Song *et al.,* 2014; Cui *et al.,* 2016), and is potentially consistent with adaptive introgression of the free-shattering allele into the weeds. The presence of wild *sh4* alleles varies widely by region, ranging from zero instances in the North East to 24% in the Central Plain. Because of difficulties in amplifying *sh4* gene in wild rice, no wild samples were genotyped at this locus.

SSR loci. The SSR genotype data from Wongtamee *et al.,* (2017) were highly polymorphic in the study populations, with expected heterozygosity values ranging from H_e = 0.347 to 0.544 among the 12 loci (Table S4). *STRUCTURE* analysis and delta-K assessments revealed an optimum at K=3 populations, with a smaller secondary peak at K=6 (Fig. S1). At K=3, wild rice formed its own unique group while cultivated and weedy rice were grouped by geography rather than plant type (Fig. 2). These patterns of differentiation were also broadly supported by the principal coordinates analysis (PCoA); the first coordinate (accounting for 17.4% of the total variation) primarily distinguishes wild rice from weedy and cultivated rice, while the second coordinate (accounting for 12.1% of the variation) broadly separates out the two geographical population groups that are present within cultivated and weedy rice (Fig. 3).

For individual accessions, *STRUCTURE* membership coefficient values revealed no evidence of admixed ancestry in cultivated rice (all membership coefficient values >98%). Weedy rice accessions showed the greatest evidence of admixture, with 29 accessions (17.5%) showing <80% assignment to a single population (Table S1). Among these, more than half (15 accessions) showed >20% assignment to the ‘blue’ population characteristic of wild rice. This

pattern is consistent with previous reports of introgression into Thai weedy rice from local wild rice populations (Pusadee *et al.*, 2013). In addition, 10 weed accessions that were assigned primarily to the 'green' population showed >20% assignment to the 'red' population, and one weed accession that was assigned primarily to the 'red' population showed 27.8% assignment to the 'green' population. As both cultivated and weedy rice are assigned to the red and green populations, this evidence of red-green admixture in the weeds could either represent crop-to-weed introgression or admixture between the two weed groups. For wild rice, two accessions showed potential evidence of introgression by the <80% membership assignment criterion; these accessions both showed evidence of shared ancestry with the 'green' population present in cultivated and weedy rice (Fig. 2; see also Table S1). However, as wild rice is genetically more diverse than either cultivated or weedy rice (both of which are ultimately derived from this wild species), the apparent admixture in the wild accessions could also be reflecting its more heterogeneous gene pool rather than introgression per se.

Comparison of domestication genes and SSRs. If the introgression of alleles at domestication genes were the result of hybridization in the recent past, weed accessions with introgressed alleles would be expected to show evidence of admixed ancestry in the genome-wide SSR markers. Instead, we found very little overlap between the patterns of introgression from neutral and domestication loci. For *Bh4*, only a single plant carried the wild allele (see above), despite 17.5% of the weedy rice accessions showing some potential evidence of wild introgression at the neutral loci. The sole plant with the *Bh4* wild allele has a membership coefficient of 98.5% to the same population as majority of weed and crop accessions in the region where it was collected (Table 2), suggesting no recent inter-population hybridization in its ancestry. Similar results were found at *Rc*. The 17% of weedy rice plants that carried crop-like *rc* alleles were genetically indistinguishable from other local weed accessions by the SSR markers; membership assignments to the local majority population were 92.1% and 92.5% for

putatively introgressed and non-introgressed weeds, respectively (Table 2). Weedy rice plants that carried the putatively introgressed (wild) *sh4* allele also showed little evidence of recent admixture from wild rice at the neutral loci; their average membership assignment to their local weed populations was 95.0% (Table 2). Among the 15 weed accessions with the wild *sh4* allele, only one accession (2205A, from the Central Plain) appears to be derived from recent weed-wild admixture; this accession has a 40% membership assignment to the wild rice population, consistent with descent from a recent wild-weed hybridization event (Table S1).

Discussion

The long-term evolutionary consequences of gene flow into agricultural weeds from cultivated and wild relatives has important implications for weed adaptation and competitive success. Here we used a combination of data from domestication genes and neutral SSR loci to assess the history of introgression into weedy rice in the major rice growing regions of Thailand. This combination of data sources has allowed us to analyze complementary aspects of gene flow in this system. Analysis of allele frequencies at domestication genes revealed a low level of potentially adaptive introgression from wild rice at the *sh4* locus, where the wild allele confers seed shattering, and potentially maladaptive introgression from cultivated rice at the *Rc* locus, where the light-pericarp *rc* allele is associated with reduced seed dormancy (Table 1). Interestingly, comparison with genome-wide neutral SSR loci reveals that very few if any of these putative introgression events at domestication loci involve recent hybridization; plants showing admixture at neutral loci are by and large not the same plants that show introgression at domestication genes (Table 2; Table S1). Thus, introgression at the domestication genes appears to reflect past hybridization events more than contemporary gene flow dynamics.

Below we discuss these findings and their implications for understanding processes of evolution and adaptation in weedy rice.

Gene flow into Thai weedy rice. One clear finding from these analyses is that introgression into weedy rice is detectable at both the domestication genes and the neutral SSR markers. Pooling across the three domestication genes, 23% of the weedy rice plants examined had putatively introgressed alleles at one or more loci (38 out of 165 accessions). Similarly, by the <80% membership coefficient criterion in the *STRUCTURE* analysis, 29 weedy rice plants (17.6%) were inferred to have introgression from wild or cultivated rice (Table S1). These results are consistent with previous studies in Thailand which report evidence for gene flow as a major force driving the evolution of the *Oryza* complex (Pusadee *et al.,* 2013; Wongtamee *et al.,* 2017).

Nonetheless, only six weedy rice plants show evidence of introgression at both domestication and neutral loci (Table S1). Taken together, these results suggest the following: first, there is a low, yet detectable, level of contemporary gene flow in this system; and second, the majority of introgression detected at the domestication genes is historical, with enough generations having passed since the hybridization event for recombination to break up any genome-wide signatures of introgression. Thus, while we detect relatively low levels of hybridization in the very recent past, our insights from the domestication genes suggest that past introgression — even if at low levels — can have a lasting effect on the composition of the weedy rice genome.

The first reported observation of invasive weedy rice in Thailand was in the Central Plain in 2001. After just five to six cropping seasons, weedy rice had overtaken entire production areas in this region. Weedy rice has since spread to every region of Thailand where high-yielding varieties are grown. Additionally, Thai weedy rice has become insensitive to

photoperiod, a trait presumably inherited from modern rice varieties (Maneechote 2004). The rapid expansion and apparent selection for introgressed individuals could help explain the results described above.

Adaptive and maladaptive introgression. One potential benefit in focusing on well-characterized domestication genes, including the three loci examined here, is that the allelic variation at these genes can in principle provide insights into patterns of adaptive or maladaptive introgression into weedy relatives. In the present study, our ability to draw definitive inferences in this regard are fairly limited. The strongest evidence for adaptive introgression comes from *sh4*, where the wild rice (G) allele (conferring freely shattering seeds) is present in nearly one-quarter of the weedy rice plants sampled in the Central Plain (Table 1). This frequency is far higher than has been reported in most weedy rice populations worldwide, the majority of which carry the reduced-shattering (T) allele as a legacy of descent from domesticated ancestors (Thurber *et al.*, 2010; Zhu *et al.*, 2012). Given the importance of seed dispersal for the persistence of weedy rice seeds in crop fields, the presence of the wild allele seems a plausible case of adaptive introgression. However, most weedy rice strains worldwide have highly shattering seeds despite carrying the domestication allele (Thurber *et al.*, 2010; Zhu *et al.*, 2012); the presence of the shattering phenotype in weedy rice appears to reflect the combined effects of multiple other shattering loci throughout the genome (Qi *et al.*, 2015). Thus, allelic variation at *sh4* may not by itself have major phenotypic or fitness impacts in weedy rice. Empirical studies that explicitly measure seed shattering in the Thai weed samples as related to *sh4* variation would be useful for assessing the potential adaptive significance of wild rice *sh4* introgression.

In the case of *Rc*, we find potential evidence of maladaptive introgression of crop alleles into the Thai weed populations. Whereas most modern rice varieties carry the 14-bp loss-of-function mutation at *Rc*, most weedy rice worldwide carries the functional *Rc* allele that is

associated with dark-pigmented pericarps and seed dormancy (Sweeney *et al.,* 2006; Gu *et al.,* 2011). In the United States, for example, weedy rice is nearly fixed for the functional *Rc* allele, and the dark-pericarp phenotype is so closely associated with weedy rice that it is commonly referred to by farmers as 'red rice' (Gross *et al.,* 2010). Seed dormancy is generally considered a critical fitness trait for agricultural weeds, as it promotes weed persistence in crop fields over multiple growing seasons. Thus, one would expect there to be strong selection against the *rc* allele in weedy rice populations. Nonetheless, we found this allele to be present in Thai weedy rice at an overall frequency of 15.2% (Table 1). Interestingly, this *rc* allele frequency is similar to that observed in weedy rice in a neighboring Southeast Asian country, Malaysia, where it was found to be present in a homozygous state in 17% of genotyped weeds (Cui *et al.,* 2016). One possible explanation for the higher *rc* allele frequency in Southeast Asia is that selection for dormancy may be weaker in this climate compared to temperate climates. In a tropical climate, the cycle of wet and dry seasons rather than summer-winter determine the period of rice cultivation. In this type of climate, water availability directly coincides with favorable periods of weed growth, as the arrival of the wet season triggers rice cultivation at the same time weedy rice would be germinating anyway. In contrast, weedy rice seeds in temperate climates must remain dormant through periods of wet but cold weather in order to survive. Thus, it is plausible that dormancy could be more strongly favored in temperate than tropical climates. Another possible explanation for the apparent lack of strong selection against *rc* in Southeast Asia is that, similar to *sh4*, there are other genes and pathways that contribute to seed dormancy (Marzougui *et al.,* 2012; Zhang *et al.,* 2017). Follow up studies that explicitly measure seed dormancy levels in Southeast Asian weedy rice would be useful for testing these hypotheses.

Another interesting feature of *Rc* allelic variation in our samples is the apparent north-to-south decrease in frequency of the domestication allele (Fig. 1). This cline could be due to a number of factors. Cultural and agricultural practices in Central Thailand are much different

than in the Lower North and North East (Pusadee *et al.,* 2013). In Central Thailand, some high-yielding modern rice varieties are direct seeded with up to 3-4 crop plantings per year. Conversely, rice in the North East is planted only once per year, which coincides with the wet season. It is possible that the more intensive agricultural practices of Central Thailand could impose stronger selection for dormancy in local weeds. Additionally, geography is much different from one region to the other. Soil quality and elevation differences might also contribute to the observed pattern for unknown reasons.

Potential limitations and avenues for future research. A potential limitation of our study is the relatively limited genetic sampling (three domestication loci and 12 SSRs). Although having more markers is always better, several aspects of these data suggest that they are sufficient to detect and analyze gene flow in this study system. For the SSR loci, polymorphism is quite high for a self-fertilizing species, with H_e ranging from 0.347-0.544 (Table S4; see also Pusadee *et al.*, 2013). Additionally, we were able to successfully detect introgression and admixture in both weedy rice and wild rice using both SSRs and domestication genes (Fig. 2). Thus, these data allow us to successfully infer that both historical and contemporary gene-flow have contributed to the evolution of weedy rice in Thailand. Follow-up studies using whole genome resequencing or reduced-representation SNP genotyping will be useful at answering these questions at finer-scale resolution.

In many world regions where weedy rice is present, there are two or more independently evolved strains of weedy rice that coexist (Wedger & Olsen, 2018). Interestingly, we have detected a similar pattern in the sampled Thai weedy rice populations, with two genetically distinct weed groups that are closely related to the crop varieties with which they co-occur (Figs. 2 and 3). As only *indica* rice varieties are cultivated in the region of our sampling, the two weed strains appear to represent two independent domestications from *indica* backgrounds. Independent weedy rice origins from *indica* rice have also been detected in a number of other

regions, including China (Qiu *et al.*, 2017), Korea (Vigueira *et al.*, 2019), Malaysia (Song *et al.*, 2014), and the United States (where the weeds are of Asian origin) (Reagon *et al.*, 2010; Li *et al.*, 2017). Pooled analyses that compare these different *indica*-derived weeds could be especially insightful for understanding the genetic mechanisms that underlie de-domestication, and the role of introgression from modern crop varieties and wild relatives in this process.

Lastly, transgene escape is a serious issue for crop breeding and sustainable crop production. Although transgenic rice has not been commercialized, the requisite technology is well advanced and could be rapidly put into practice on a large scale, for example with the production of herbicide-resistant rice cultivars. Both cultivated and weedy rice are primarily selfing, but outcrossing and hybridization does occur (Singh *et al.,* 2017a; Singh *et al.,* 2017b), and our analyses in the present study indicate that this hybridization can have multi-generation impacts on the composition of the weedy rice genome (Fig. 2; Table S1). Recent studies have suggested that weedy rice can act as a bridge for gene flow between cultivated and wild rice due to its extended period of flowering and weak postzygotic barriers to reproduction (Qiu *et al.,* 2017; Singh *et al.,* 2017a). Based on our results, one can conclude that if transgenic rice were to be introduced in Thailand, eventual escape into wild rice would be likely, and weedy rice could well serve as the conduit.

Funding

MJW is supported by an NSF Graduate Research Fellowship Program (Fellow ID: 2017221153). Weedy rice research in the Olsen lab has been supported through the NSF Plant Genome Research Program (IOS-1032023).

Data availability Primary data used in the analysis of this paper is deposited in supplementary table S1. SSR genotypes taken from Wongtamee *et al.*, (2017) are deposited in supplementary table S3.

Acknowledgements

The authors would like to thank members of the Olsen lab group for helpful comments on earlier drafts of this manuscript.

References

Akasaka M, Ushiki J, Iwata H, Ishikawa R, Ishii T. **2009**. Genetic relationships and diversity of weedy rice (Oryza sativa L.) and cultivated rice varieties in Okayama Prefecture, Japan. *Breeding Science* **59**: 401–409.

Beebe S, Toro Ch. O, González AV, Chacón MI, Debouck DG. **1997**. Wild-weed-crop complexes of common bean (Phaseolus vulgaris L., Fabaceae) in the Andes of Peru and Colombia, and their implications for conservation and breeding. *Genetic Resources and Crop Evolution* **44**: 73–91.

Cao Q, Lu B-R, Xia H, Rong J, Sala F, Spada A, Grassi F. **2006**. Genetic diversity and origin of weedy rice (Oryza sativa f. spontanea) populations found in north-eastern China revealed by simple sequence repeat (SSR) markers. *Annals of Botany* **98**: 1241–1252.

Chase MW, Hills HH. **1991**. Silica gel: An ideal material for field preservation of leaf samples for DNA studies. *Taxon* **40**: 215–220.

Cui Y, Song BK, Li L-F, Li Y-L, Huang Z, Caicedo AL, Jia Y, Olsen KM. **2016**. Little white lies: Pericarp color provides insights into the origins and evolution of southeast Asian weedy rice. *Genes|Genomes|Genetics* **6**: 4105–4114.

Diarra A, Smith RJ, Talbert RE. **1985**. Interference of Red Rice (Oryza sativa) with Rice (O. sativa). *Weed Science* **33**: 644–649.

Doyle JJ, Doyle JL. **1987**. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**: 11–15.

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE. **2010**. Crops gone wild: Evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications* **3**: 494–504.

Engku AK, Norida M, Juraimi AS, Rafii MY, Abdullah SNA, Alam MA. **2016**. Gene flow from clearfield® rice to weedy rice under field conditions. *Plant, Soil and Environment* **62**: 16–22.

Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14**: 2611–2620.

Federici MT, Vaughan D, Tomooka N, Kaga A, Wang XW, Doi K, Francis M, Zorrilla G, Saldain N. **2001**. Analysis of Uruguayan weedy rice genetic diversity using AFLP molecular markers. *Electronic Journal of Biotechnology* **4**: 42–57.

Grimm A, Fogliatto S, Nick P, Ferrero A, Vidotto F. **2013**. Microsatellite markers reveal multiple origins for Italian weedy rice. *Ecology and Evolution* **3**: 4786–4798.

Gross BL, Reagon M, Hsu SC, Caicedo AL, Jia Y, Olsen KM. **2010**. Seeing red: The origin of grain pigmentation in US weedy rice. *Molecular Ecology* **19**: 3380–3393.

Gu XY, Foley ME, Horvath DP, Anderson J V., Feng J, Zhang L, Mowry CR, Ye H, Suttle JC, Kadowaki KI, *et al.* 2011. Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. *Genetics* **189**: 1515–1524.

Guo L, Qiu J, Li LF, Lu B, Olsen K, Fan L. **2018**. Genomic clues for crop–weed interactions and evolution. *Trends in Plant Science* **23**: 1102–1115.

Huang Z, Kelly S, Matsuo R, Li L-F, Li Y, Olsen KM, Jia Y, Caicedo AL. **2018**. The role of standing variation in the evolution of weediness traits in south Asian weedy rice (Oryza spp.). *Genes*|*Genomes*|*Genetics* **8**: g3.200605.2018.

Huang Z, Young ND, Reagon M, Hyma KE, Olsen KM, Jia Y, Caicedo AL. **2017**. All roads lead to weediness: Patterns of genomic divergence reveal extensive recurrent weedy rice origins from South Asian Oryza. *Molecular Ecology* **26**: 3151–3167.

Kearse M, Moir R, Wilson A, Stones-havas S, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, *et al.* 2012. Geneious Basic : An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Li LF, Li YL, Jia Y, Caicedo AL, Olsen KM. **2017**. Signatures of adaptation in the weedy rice genome. *Nature Genetics* **49**: 811–814.

Li C, Zhou A, Sang T. **2006**. Rice domestication by reducing shattering. *Science* **311**: 1936–1939.

Londo JP, Schaal BA. **2007**. Origins and population genetics of weedy red rice in the USA. *Molecular Ecology* **16**: 4523–4535.

Marzougui S, Sugimoto K, Yamanouchi U. **2012**. Mapping and characterization of seed dormancy QTLs using chromosome segment substitution lines in rice. *Theoretical and Applied Genetics* **124**: 893–902.

Merotto A, Goulart ICGR, Nunes AL, Kalsing A, Markus C, Menezes VG, Wander AE. **2016**. Evolutionary and social consequences of introgression of nontransgenic herbicide resistance from rice to weedy rice in Brazil. *Evolutionary Applications* **9**: 837–846.

Morrell PL, Williams-Coplin TD, Lattu AL, Bowers JE, Chandler JM, Paterson AH. **2005**. Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. *Molecular Ecology* **14**: 2143–2154.

Peakall R, Smouse PE. **2006**. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.

Peakall R, Smouse PE. **2012**. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**: 2537–2539.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pusadee T, Schaal BA, Rerkasem B, Jamjod S. **2013**. Population structure of the primary gene pool of Oryza sativa in Thailand. *Genetic Resources and Crop Evolution* **60**: 335–353.

Qi X, Liu Y, Vigueira CC, Young ND, Caicedo AL, Jia Y, Gealy DR, Olsen KM. **2015**. More than one way to evolve a weed: parallel evolution of US weedy rice through independent genetic mechanisms. *Molecular Ecology* **24**: 3329–3344.

Qiu J, Zhou Y, Wang Y, Mao L, Ye C, Wang W, Zhang J, Yu Y, Fu F, Wang Y, *et al.* 2017. Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nature Communications* **8**: 1–12.

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL. **2010**. Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evolutionary Biology* **10**.

Singh V, Burgos NR, Singh S, Gealy DR, Gbur EE, Caicedo AL. **2017a**. Impact of volunteer rice infestation on yield and grain quality of rice. *Pest Management Science* **73**: 604–615.

Singh K, Kumar V, Saharawat YS, Gathala M, Ladha JK, Chauhan BS. **2013**. Weedy Rice: An Emerging Threat for Direct-seeded Rice Production Systems in India. *Journal of Rice Research* **1**: 1–6.

Singh V, Singh S, Black H, Boyett V, Basu S, Gealy D, Gbur E, Pereira A, Scott RC, Caicedo A, *et al.* 2017b. Introgression of Clearfield™ rice crop traits into weedy red rice outcrosses. *Field Crops Research* **207**: 13–23.

Song BK, Chuah TS, Tam SM, Olsen KM. **2014**. Malaysian weedy rice shows its true stripes: Wild Oryza and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular Ecology* **23**: 5003–5017.

Sun J, Qian Q, Ma D, Xu Z, Liu D, Du H, Chen W, Chen W. **2013**. Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New Phytologist* **197**: 290–299.

Sweeney MT, Thomson MJ, Pfeil BE, McCouch S. **2006**. Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *The Plant Cell* **18**: 283–294.

Sweeney MT, Thomson MJ, Yong GC, Yong JP, Williamson SH, Bustamante CD, McCouch SR. **2007**. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genetics* **3**: 1418–1424.

Thurber CS, Reagon M, Gross BL, Olsen KM, Jia Y, Caicedo AL. **2010**. Molecular evolution of shattering loci in U.S. weedy rice. *Molecular Ecology* **19**: 3271–3284.

Vigueira CC, Li W, Olsen KM. **2013**. The role of Bh4 in parallel evolution of hull colour in domesticated and weedy rice. *Journal of Evolutionary Biology* **26**: 1738–1749.

Vigueira CC, Qi X, Song BK, Li LF, Caicedo AL, Jia Y, Olsen KM. **2019**. Call of the wild rice: Oryza rufipogon shapes weedy rice evolution in Southeast Asia. *Evolutionary Applications*: 1–12.

Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. **2017**. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genome Research* **27**: 1029–1038.

Warwick SI, Légère A, Simard MJ, James T. **2008**. Do escaped transgenes persist in nature? The case of an herbicide resistance transgene in a weedy Brassica rapa population. *Molecular Ecology* **17**: 1387–1395.

Wedger MJ, Olsen KM. **2018**. Evolving insights on weedy rice. *Ecological Genetics and Genomics* **7**–**8**: 23–26.

Wongtamee A, Maneechote C, Pusadee T, Rerkasem B, Jamjod S. **2017**. The dynamics of spatial and temporal population genetic structure of weedy rice (Oryza sativa f. spontanea Baker). *Genetic Resources and Crop Evolution* **64**: 23–39.

Zhang L, Lou J, Foley ME, Gu X. **2017**. Comparative mapping of seed dormancy loci between tropical and temperate ecotypes of weedy rice (Oryza sativa L.). *Genes*|*Genomes*|*Genetics* **7**: 2605–2614.

Zhang L-B, Zhu Q, Wu Z-Q, Ross-Ibarra J, Gaut BS, Ge S, Sang T. **2009**. Selection on grain shattering genes and rates of rice domestication. *New Phytologist* **184**: 708–720.

Zhu Y, Ellstrand NC, Lu BR. **2012**. Sequence polymorphisms in wild, weedy, and cultivated rice suggest seed-shattering locus sh4 played a minor role in Asian rice domestication. *Ecology and Evolution* **2**: 2106–2113.

Zhu B-F, Si L, Wang Z, Jingjie Zhu YZ, Shangguan Y, Lu D, Fan D, Li C, Lin H, Qian Q, *et al.* 2011. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiology* **155**: 1301–1311.

Ziska LH, Gealy DR, Burgos N, Caicedo AL, Gressel J, Lawton-Rauh AL, Avila LA, Theisen G, Norsworthy J, Ferrero A, *et al.* 2015. An emerging constraint to global rice production: Weedy (red) rice. *Advances in Agronomy* **129**: 181–228.

Tables

Table 1. **Distributions of domestication alleles in the sampled rice groups**. Numerators indicate the number of genotyped accessions that carry the domestication allele at each gene; denominators indicate the total number of genotyped accessions.

	Bh4 22-bp deletion	***Rc*** 14-bp deletion	***sh4*** T substitution
Rice type			
Cultivated	104/104 (100%)	100/101 (99.0%)	59/59 (100%)
Weedy:			
Lower North (LN)	77/77 (100%)	16/75 (21.3%)	45/53 (84.9%)
North East (NE)	40/40 (100%)	5/40 (12.5%)	33/33 (100%)
Central Plain (CP)	48/49 (98.0%)	3/43 (7.0%)	19/25 (76.0%)
All regions	165/166 (99.0%)	24/158 (15.2%)	96[a]/111 (86.5%)
Wild	3/28 (10.7%)	2/7 (28.6%)	—

a – includes one heterozygote

Table 2. Comparison of *STRUCTURE* membership coefficients for weedy rice accessions with and without putatively introgressed alleles at domestication genes. Membership assignment values in the left column would be expected to be significantly lower than values in the right column if the domestication gene introgression occurred through recent hybridization. A two-sample, equal variance t-test indicates no significant differences ($p>0.75$ in all cases).

	Membership coefficients[a]	
	Accessions with putatively introgressed allele ± SE (N)	Accessions with majority allele ± SE (N)
Bh4	0.980 ± n/a (1)	0.923 ± 0.009 (163)
Rc	0.921 ± 0.025 (24)	0.925 ± 0.009 (132)
sh4	0.950 ± 0.026 (15)	0.899 ± 0.012 (95)

a - Values are shown with respect to the population that the majority of weed accessions in a given region are assigned to.

Figures

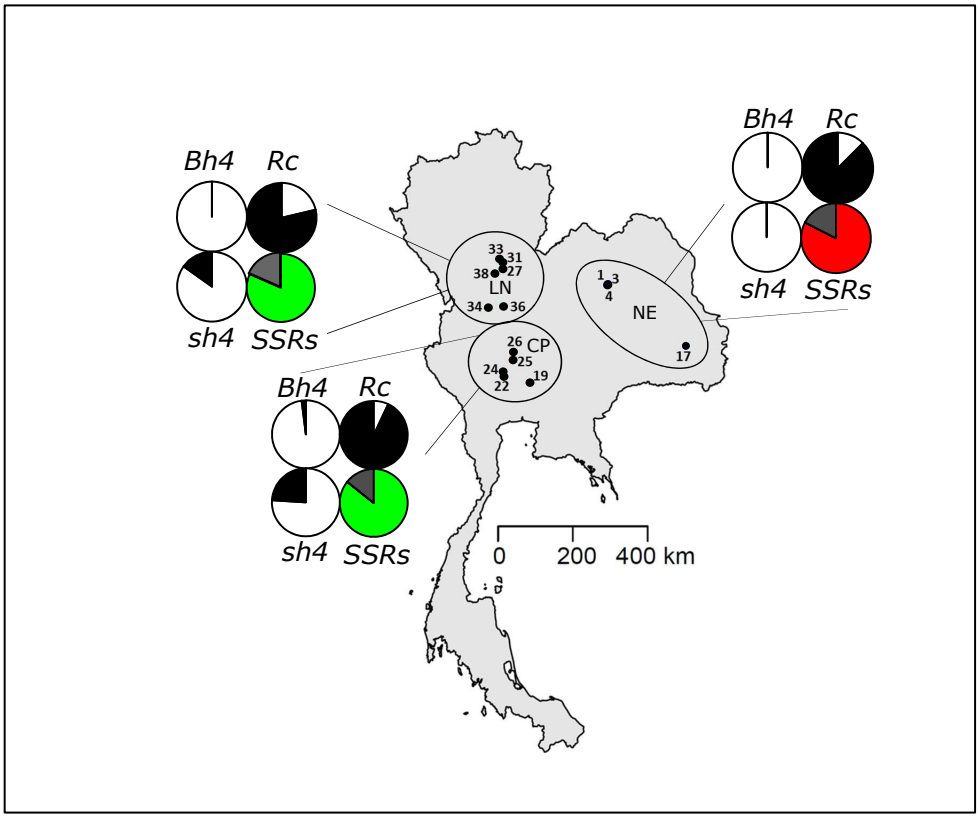


Figure 1. Sampling locations in Thailand. Black dots represent collection sites with numbers representing the field number. North East (NE) samples represent collections from Khon Kaen (1,3,4) and Ubon Ratchathani (17) provinces. Central Plain (CP) samples represent collections from Phra Nakhon Si Ayutthaya (19), Suphan Buri (22,24), Sing Buri (25), and Nakhon Sawan (26) provinces. Lower North (LN) samples represent collections from Phitsanulok (27), Uttaradit (31,33), Phichit (34,36), and Sukhothai (38) provinces. Pie charts labeled *Bh4*, *Rc*, and *sh4* represent allele frequencies in weedy rice, with white representing the domestication allele and black representing the wild allele. The pie chart labeled "SSRs" represents the proportion of weedy rice samples that *STRUCTURE* has unambiguously assigned to a population based on ≥0.80 membership assignment. The green and red colors represent the green and red *STRUCTURE* populations, while the grey represents plants that show evidence of admixture (<0.80 assignment to a single population).

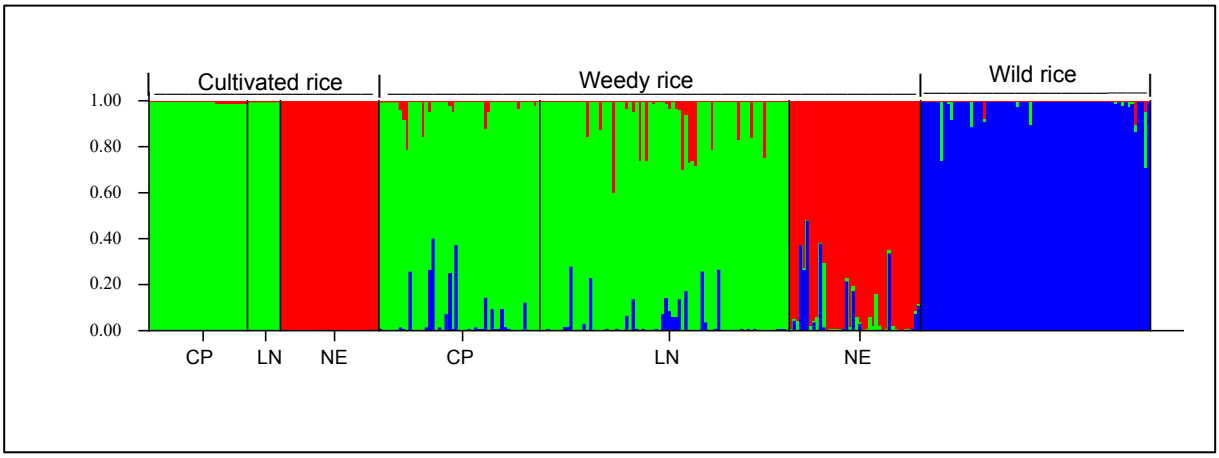


Figure 2. *STRUCTURE* output at K=3 populations. Accessions analyzed include 67 cultivated, 165 weedy, and 70 wild rice plants. Cultivated and weedy rice plants are separated into 3 geographical regions; North East (NE), Central Plain (CP), and North East (NE). Accessions with a population membership assignment <0.80 to any one population are considered admixed.

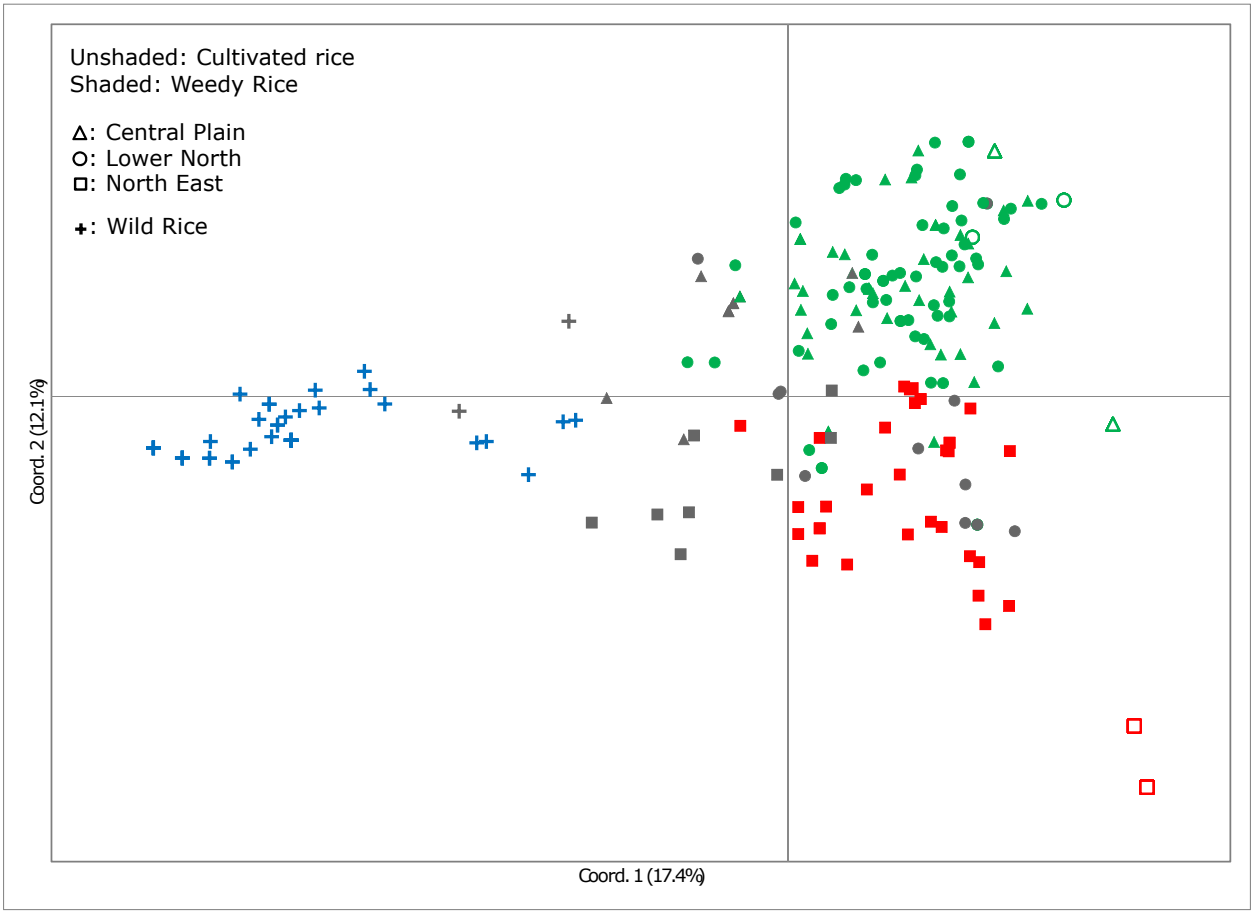


Figure 3. Principal Coordinates Analysis (PCoA) based on SSR markers. Symbol shape represents collection location. Colors correspond to *STRUCTURE* populations in Figure 2, with gray denoting admixed individuals. Filled shapes are weedy rice, while open shapes represent cultivated rice accessions.

APPENDIX III

Chapter 4 Supplementary Material

Supplementary Tables

Table S1. List of accessions used in the study. Instances of inferred introgression of alleles at domestication genes (columns F-H) are highlighted in yellow. In columns I-K, accessions are color-coded according to their primary *STRUCTURE* membership coefficient (>80% membership assignment). Colors indicate the red, green, or blue populations in Figure 2. Plants with a membership coefficient < 80% to any one population are highlighted in gray.

Due to size of table, please see online supplementary materials here:

https://academic.oup.com/jhered/article/110/5/601/5486338

or contact author.

Table S2. List of primer names and sequences used in this study.

PrimerName	Type	Sequence
Rc_gtR	Domestication gene	TCAAATACGGCTTTATAGAAATAGAGG
Rc_delF	Domestication gene	GATGCCATCCAAGGTGATTT
sh4_004F	Domestication gene	ACGGGCACCTGACTGCTACG
sh4_004R	Domestication gene	GAGGTGGGTGGTGGTGATGG
Rc_delR3	Domestication gene	GAAATCACCTTGGATGGCATCC
Rc_wtR1	Domestication gene	TTCCAATGTTCGTTAGAGGC
Rc_wtF1	Domestication gene	GCAAGTGGAACGCGAAAAGT
Bh4_gt2R	Domestication gene	CCATGAGGTCCGCGAAGAAC
Bh4_gt2F	Domestication gene	AAGCAAGTTATAAACAATCTGGTGCA
Rc_delF3	Domestication gene	ACAACACTGACACTGAAAGG
Rc_wtF	Domestication gene	CAAGTGGAACGCGAAAAGTC
Bh4-22F1	Domestication gene	CAACCAGATGCTAGTGATATGC
Bh4-22R1	Domestication gene	AGGTTGAGCGTCACCTG
RM1F	SSR	GCGAAAACACAATGCAAAAA
RM1R	SSR	GCGTTGGTTGGACCTGAC
MR109F	SSR	GCCGCCGGAGAGGGAGAGAGAG
MR109R	SSR	CCCCGACGGGATCTCCATCGTC
RM211F	SSR	CCGATCTCATCAACCAACTG
RM211R	SSR	CTTCACGAGGATCTCAAAGG
RM251F	SSR	GAATGGCAATGGCGCTAG
RM251R	SSR	ATGCGGTTCAAGATTCGATC
RM280F	SSR	ACACGATCCACTTTGCGC
RM280R	SSR	TGTGTCTTGAGCAGCCAGG
RM133F	SSR	TTGGATTGTTTTGCTGGCTCGC
RM133R	SSR	GGAACACGGGGTCGGAAGCGAC
RM234F	SSR	ACAGTATCCAAGGCCCTGG
RM234R	SSR	CACGTGAGACAAAGACGGAG
RM481F	SSR	TAGCTAGCCGATTGAATGGC
RM481R	SSR	CTCCACCTCCTATGTTGTTG
RM477F	SSR	TCTCGCGGTATAGTTTGTGC
RM477R	SSR	ACCACTACCAGCAGCCTCTG
RM316F	SSR	CTAGTTGGGCATACGATGGC
RM316R	SSR	ACGCTTATATGTTACGTCAAC
RM206F	SSR	CCCATGCGTTTAACTATTCT
RM206R	SSR	CCCATGCGTTTAACTATTCT
RM247F	SSR	TAGTGCCGATCGATGTAACG
RM247R	SSR	CATATGGTTTTGACAAAGCG

Table S3. **Raw SSR data from Wongtamee et al. (2017) used in this study.**

Due to size of table, please see online supplementary materials here:

https://academic.oup.com/jhered/article/110/5/601/5486338

or contact author.

Table S4. Diversity statistics calculated from the neutral loci used in this study. Values are shown for expected heterozygosity (H_e), observed heterozygosity (H_o), number of alleles per locus (N_a), and average number of alleles per group in Fig. 2 (N_p).

	Diversity Measure			
Locus	**H_e**	**H_o**	**N_a**	**N_p**
1	0.543	0.033	10	4
206	0.414	0.026	6	3.1
481	0.5	0.023	9	3.4
280	0.347	0.006	7	3.4
255	0.403	0.006	5	2.9
341	0.397	0.008	5	2.9
586	0.376	0.162	5	2.4
588	0.409	0	4	2.4
164	0.544	0.027	15	4.4
167	0.51	0.031	12	3.7
273	0.548	0.009	14	3.6
232	0.501	0.26	13	3.9

Supplemental Figures

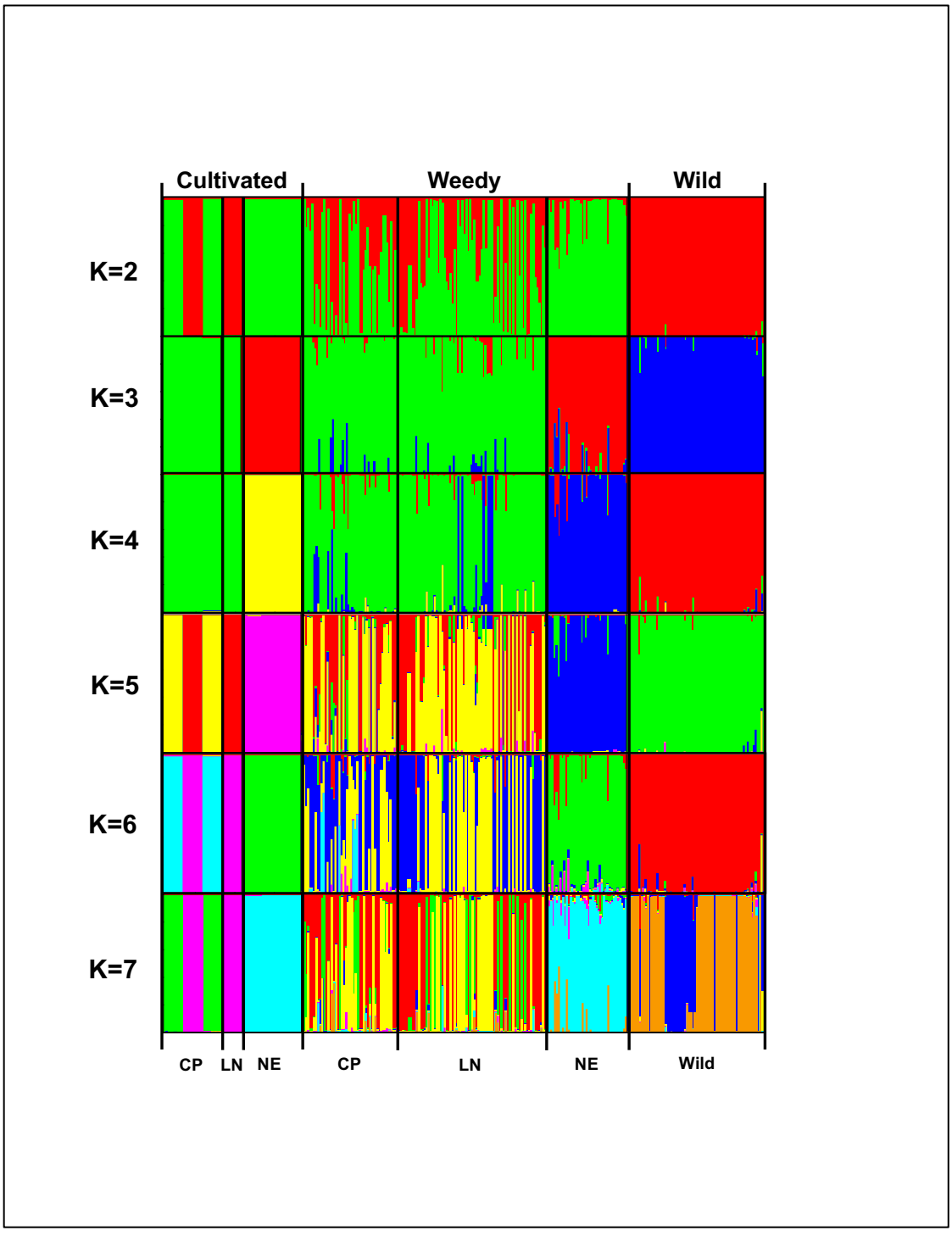


Figure S1. Population membership assignment using *STRUCTURE*. Rows represent different inferred number of populations at K=2-7.

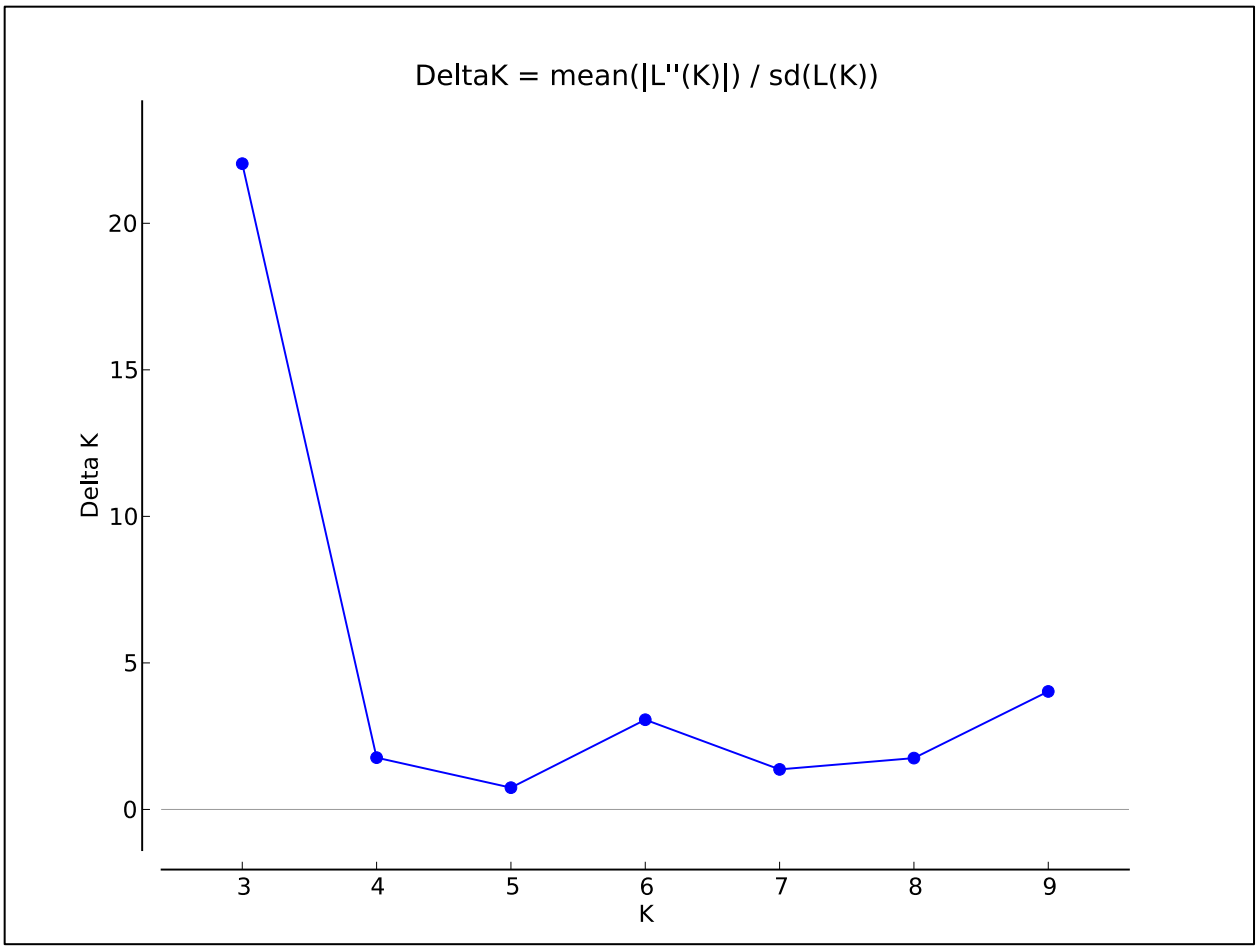


Figure S2. Delta-K plot from *STRUCTURE Harvester*. A primary peak at K=3 denotes the optimal K with a smaller secondary peak at K=6.

CONCLUSIONS

Within this dissertation, I present the body of work I have conducted using weedy rice as a model system for understanding how microevolutionary forces interact and drive evolution in para-agricultural contexts. This research has directly contributed to a growing body of literature using weedy crop relatives to elucidate the processes underlying natural selection and adaptation, genetic drift, and gene flow. Furthermore, the chapters herein aimed to advance our understanding of weedy rice as a devastating agricultural weed and document responses to human-mediated selection. The insights gained from this thesis will help inform farmers and biotech companies as rice agriculture continues to move forward.

Chapter One reviews the history of weedy rice evolutionary research. We discuss weedy rice as a world-wide problem and evaluate how technological advances in molecular biology have uncovered novel aspects of weedy rice evolutionary history. We conclude this chapter with speculation on future avenues of research including advanced gene identification methods and a combined “pan-omics” approach.

The first data chapter, Chapter Two, utilizes two previously generated QTL mapping populations to map root architecture traits potentially responsible for the competitive ability of weedy rice. Both mapping populations share an *indica* crop parent while differing in the SH or BHA weedy parent. Thus, the parents of the SH-*indica* mapping population are more closely related than the SH and BHA parents. This quirk of the mapping populations was utilized, in conjunction with root system architecture (RSA) phenotyping, to determine a suite so-called ‘weed-specific RSA traits’. From these descriptor traits, the crop root system can be summarized as being different from the weeds in the following ways: it is wider and higher in the soil, with individual roots that are thinner, longer, more abundant, more curved, and at a lower angle to the soil. Next, we utilized a random forest machine learning model to determine whether root phenotypes could reliably be used to distinguish plant genotype and identify the most diagnostic traits. We found that random forest modeling is ~70% effective at correctly

identifying the genotype, more than doubling the chance of random identification (33%). We also find that root system width and exploratory factors are the most distinguishing features of weedy rice. Lastly, we performed QTL mapping in both populations using > 47000 images from ~650 plants. Comparative QTL mapping revealed little evidence for shared underlying genetic mechanisms. We identified 65 significant QTLs distributed across 43 root traits. Of those 43 traits, only 10 mapped to both populations with only a single trait, convex hull area, mapping to the same genomic location in both populations. Thus, while we find clear evidence of convergent phenotypic evolution, we find that the genetic mechanisms underlying those traits are very different.

In Chapter Three we perform whole-genome sequencing on 48 weedy rice plants collected in 2018. In 2002, non-transgenic herbicide resistant rice cultivars were commercialized in the US. By 2004 farmers were reporting herbicide resistant weedy rice, which was soon confirmed as crop-weedy hybrid derived (Burgos *et al.*, 2014). Nearly simultaneously, in 2000, hybrid rice cultivars were released. Through a process called 'volunteering' the descendants of these hybrid rice cultivars acted as a bridge for gene flow between cultivated and weedy rice, aiding in the escape of herbicide resistance alleles. Nearly two decades later, our whole genome sequencing project reveals nearly ubiquitous resistance phenotypes and four unique haplotypes in contemporary weedy rice capable of resisting herbicide application. Two of these haplotypes are crop-derived and, unsurprisingly, appear in plants with clear evidence of crop-weed hybridization. These haplotypes make up the majority of our samples (75/96 haplotypes). Nine additional haplotypes carry the same nucleotide substitutions leading to herbicide resistance while appearing in haplotypic backgrounds without evidence of crop introgression. These nine haplotypes, representing the other two unique haplotypes, are probably instances of parallel evolution via molecular convergence. Next, using a local ancestry approach we find that selection has shifted contemporary hybrid weedy rice genomes to be ~70% weed derived and

30% crop derived. This suggests a selective advantage to purging the crop derived portions of the genome. Lastly, we find that US weedy rice demographics and population genomic composition have been altered by this bout of strong selection. BHA-derived contemporary weeds account for 37/48 (77%) contemporary samples but no historic-unaltered BHA plants were recovered. Conversely, SH-like hybrid contemporary weeds only make up 6/48 (12.5%) but 4 (8.3%) historic-unaltered SH plants were found. The results from this study show a monumental shift in US weedy rice populations in response to selection for herbicide resistance and document yet another example of natural selection coming out ahead in the evolutionary arms race with human farmers.

The final data chapter, Chapter Four, sought to investigate the gene flow dynamics of weedy rice in Thailand in the presence of the rice wild progenitor, *Oryza rufipogon*. Rice accessions from Thailand were collected from three major rice growing regions and included 124 cultivated, 166 weedy, and 98 wild rice samples. Each sample was genotyped at 12 SSRs and 3 known domestication genes (*Bh4*, controlling hull color; *Rc*, controlling pericarp color and seed dormancy; and *sh4*, controlling seed shattering). While the domestication genes gave us insight into the selective history of these plants, the complementary SSR dataset would inform us to the neutral history of each population. The SSR dataset revealed significantly more crop-weed introgression than expected if the wild ancestor was not present (e.g. the United states (Reagon *et al.*, 2010)). In concert with the SSR data, the domestication gene data set found evidence of crop-specific alleles in both weedy and wild rice. In fact, we found evidence of putatively maladaptive crop-to-weed gene flow at the *Rc* gene, suggesting significant and recent gene flow in the complex. Interestingly, the gene flow identified by these two presumably complementary datasets are discordant. Indeed, we find very little overlap between the patterns of introgression from neutral and domestication loci. We conclude that this discordance is likely temporal in nature, with the wild-weed hybridization occurring long before the crop-weed

introgression. This work offers insights into the gene flow dynamics of weedy crop relatives and the long-term selective consequences of adaptive and maladaptive gene flow.

Taken together, the four chapters presented in this dissertation proved insights into weedy rice as a model system for microevolutionary interactions. This dissertation has added to growing documentation from across the world on the propensity for weedy rice to adaptively introgress useful alleles from cultivated and wild rice (Wongtamee *et al.*, 2017; De Leon *et al.*, 2019; Qiu *et al.*, 2020). This accumulated body of information suggests a role for the maintenance of low levels of outcrossing, despite the strongly selfing nature of the species.

An additional insight from this research is on the diverse genetic mechanisms underlying adaptation in response to natural selection. Interestingly, we find examples of both convergent evolution (e.g., herbicide resistance, where molecular convergence is evident at the nucleotide level in the *ALS* gene), and non-convergent mechanisms of evolution (e.g., root growth, where the independently-evolved SH and BHA weeds have entirely different genetic architectures for most RSA traits). This combination of findings points to there being very little inherent genetic (or developmental) constraint on the mechanisms by which weed-adaptive phenotypes can evolve. This genetic flexibility has undoubtedly facilitated the repeated evolution of weedy rice that we see worldwide.

Lastly, my work has continued to elevate weedy rice as a model species and further opened avenues of future research. We find 65 QTL related to root system architecture in 13-day old rice seedlings. These QTL could be further mapped to identify the genes underlying phenotypes of interest. Furthermore, older plants could be imaged to investigate how root system architecture continues to develop and determine whether or not seedling root systems can predict late-stage resource uptake or even yield. In Chapter Three, the 48 plants were all collected from just a single county of Arkansas. More samples from the southern US or other

rice growing regions, such as France (Bourdineaud, 2020), could be sequenced to determine the impact of Clearfield™ rice in those areas. Additional studies into the long-term genomic response of crop-weed hybridization would also be insightful. We find that ancestral weedy rice genomes make up ~70% of contemporary hybrid weedy rice genomes. It would be interesting to investigate similar crop-weed hybrids to determine the degree to which one ancestral genome dominates the other.

Works Cited

Bourdineaud JP. **2020**. Toxicity of the herbicides used on herbicide-tolerant crops, and societal consequences of their use in France. *Drug and Chemical Toxicology* **0545**.

Burgos NR, Singh V, Tseng TM, Black H, Young ND, Huang Z, Hyma KE, Gealy DR, Caicedo AL. **2014**. The Impact of Herbicide-Resistant Rice Technology on Phenotypic Diversity and Population Structure of United States Weedy Rice. *Plant Physiology* **166**: 1208–1220.

De Leon TB, Karn E, Al-Khatib K, Espino L, Blank T, Andaya CB, Andaya VC, Brim-DeForest W. **2019**. Genetic variation and possible origins of weedy rice found in California. *Ecology and Evolution* **9**: 5835–5848.

Qiu J, Jia L, Wu D, Weng X, Chen L, Sun J, Chen M, Mao L, Jiang B, Ye C, *et al.* 2020. Diverse genetic mechanisms underlie worldwide convergent rice feralization. *Genome Biology* **21**: 70.

Reagon M, Thurber CS, Gross BL, Olsen KM, Jia Y, Caicedo AL. **2010**. Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice. *BMC Evolutionary Biology* **10**.

Wongtamee A, Maneechote C, Pusadee T, Rerkasem B, Jamjod S. **2017**. The dynamics of spatial and temporal population genetic structure of weedy rice (*Oryza sativa* f. *spontanea* Baker). *Genetic Resources and Crop Evolution* **64**: 23–39.