9-2012

The Evolutionary Genetics of Seed Shattering and Flowering Time, Two Weed Adaptive Traits in US Weedy Rice

Carrie S. Thurber
University of Massachusetts Amherst, cthurber@cns.umass.edu

THE EVOLUTIONARY GENETICS OF SEED SHATTERING AND FLOWERING TIME, TWO WEED ADAPTIVE TRAITS IN US WEEDY RICE

A Dissertation Presented

by

CARRIE S. THURBER

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2012

Plant Biology Graduate Program

THE EVOLUTIONARY GENETICS OF SEED SHATTERING AND FLOWERING TIME, TWO WEED ADAPTIVE TRAITS IN US WEEDY RICE

A Dissertation Presented

by

CARRIE S. THURBER

Approved as to style and content by:

_____
Ana L. Caicedo, Chair

_____
Samuel P. Hazen, Member

_____
Lynn S. Adler, Member

_____
Jeffrey L. Blanchard, Member

_____
Elsbeth L. Walker, Director, Plant Biology Graduate Program

DEDICATION

"Weed is not a category of nature but a human construct, a defect of our perception."
Michael Pollan, Second Nature

This dissertation is dedicated to my parents who always supported me in my academic pursuits.

ACKNOWLEDGMENTS

I would like to thank my advisor, Ana Caicedo, for her mentoring and my fellow Plant Biology graduate program and lab members for their assistance and friendship.

ABSTRACT

THE EVOLUTIONARY GENETICS OF SEED SHATTERING AND FLOWERING TIME, TWO WEED ADAPTIVE TRAITS IN US WEEDY RICE

SEPTEMBER 2012

CARRIE S. THURBER, B.S., FRAMINGHAM STATE UNIVERSTIY

M.S., UNIVERSITY AT BUFFALO

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Ana L. Caicedo

Weedy rice is a persistent weed of cultivated rice (*Oryza sativa*) fields worldwide, which competes with the crop and drastically reduces yields. Within the US, two main populations of genetically differentiated weedy rice exist, the straw-hulled (SH) group and the black-hulled awned (BHA) group. Current research suggests that both groups are derived from Asian cultivated rice. However, the weeds differ from the cultivated groups in various morphological traits. My research focus is on the genetic basis of two such traits: seed shattering ability and differences in flowering time. The persistence of weedy rice has been partly attributed to its ability to shatter (disperse) seed prior to crop harvesting. I have investigated the shattering phenotype in a collection of US weedy rice accessions and find that all US weedy rice groups shatter seeds easily. Additionally, I characterized the morphology of the abscission layer at the site where seed release occurs and find that weeds begin to degrade their abscission layers at least five days prior to wild plants. I also assessed allelic identity and diversity at the major shattering locus, *sh4*, in weedy rice and find that all cultivated and weedy rice share similar haplotypes at *sh4*. These haplotypes contain a single derived mutation associated with decreased seed

shattering during domestication. The combination of a shared cultivar *sh4* allele and a highly shattering phenotype suggests that US weedy rice have re-acquired the shattering trait after divergence from their crop progenitors through alternative genetic mechanisms. Additionally, my investigation into flowering time in weedy rice shows that weed populations differ in their flowering times. I also assessed allelic identity and diversity at two genes involved in the transition to flowering, *Hd1* and *Hd3a*, and again found haplotype sharing between weeds and cultivars with *Hd1* only accounting for some of the flowering time differences between weeds. In order to locate genomic regions containing additional candidate genes I conducted a QTL mapping study on two F_2 populations derived from crosses of weedy rice with cultivated rice. My results show sharing of QTL for flowering time between populations, yet lack of sharing of QTL for shattering.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

OVERVIEW OF DISSERTATION TOPIC

1.1 Parallel Evolution

The evolution of morphological similarities between different species can be homologous, arising from shared descent from a common ancestor, or homoplasious, independently derived from separate ancestors (Hodin, 2000). Parallel evolution, also referred to as convergence, is a type of homoplasy, where two evolving populations have acquired the same trait independently (Hodin, 2000; Bollback & Huelsenbeck, 2009). For the purposes of this thesis, parallel evolution and convergence are used interchangeably to refer to independently evolved phenotypic similarities.

Population genetic theory predicts that there will be a decrease in the probability of parallel evolution if there is an increase in the number of possible adaptive solutions (Bollback & Huelsenbeck, 2009). If only one adaptive solution is viable, then there will be high levels of parallel evolution even among highly divergent species (Bollback & Huelsenbeck, 2009). Similarly, Haldane (1931) posited that similar selective pressures lead to similar directions of heritable variation in closely related species. Additionally, closely related species tend to vary in the same way phenotypically due to shared genetic biases, which may predispose these species to utilizing similar genes, and even similar mutations, to independently arrive at convergent phenotypes (Vavilov 1922; Schluter *et al.*, 2004). Whether or not these instances of parallel evolution actually are due to similar genetic changes is of major interest in evolutionary genetics (Nadeau & Jiggins, 2010).

Parallel evolution occurs across many taxa ranging from insects to plants to higher eukaryotes and within traits ranging from disease resistance to flower color to

other adaptive traits. It is unknown what the exact genetic mechanisms are that control many morphological traits as similarities can occur at different levels (i.e. the same nucleotide within the same gene, different nucleotides at the same gene, different genes within a similar pathway, or even different biochemical pathways). Although there are many phenotypic examples in nature, only recently have scientists begun examining the genetic basis of some of these including the parallel evolution of adaptation to fresh water in two different populations of Pacific Ocean threespine stickleback (same gene, Schluter *et al*., 2004), parallel changes in pigmentation in fruit flies (multiple genes, Wittkopp *et al*., 2003), and parallel shifts in flower color and patterning across several angiosperm species (similar enzymes, similar gene families and similar cis-regulatory changes, Schwinn *et al.*, 2006; Des Marias & Rausher 2010; Streisfeld & Rausher 2009; Smith & Rausher 2011).

Plants evolving in the agricultural environment, including domesticated crops and weeds, are ideal systems for the study of parallel evolution. The evolutionary process of domestication often involves selection for similar traits in different crops (i.e. artificially-driven parallel evolution), which increase their usefulness to humans; likewise, plants invading crop fields experience selection for weedy traits that allow them to succeed in the agricultural environment. Many popular grass crops such as rice, corn, barley, and wheat have been selected for similar traits (e.g. increased selfing, uniform germination, and decreased seed shedding) despite having different centers of domestication worldwide (Purugganan & Fuller, 2009). These adaptations may have allowed for increased germination in disturbed and deep soils as well as easier harvesting and higher yields.

Highly competitive invasive weeds, especially those that colonize agricultural fields, have also been selected for similar fitness related traits that help them out-compete neighboring domesticated plants for resources and produce more offspring (Basu *et al*., 2004). Traits such as rapid growth, high seed production, increased seed dispersal and deep roots have been characterized at the phenotypic level in several agricultural weeds including weedy rice, wild turnip, and Johnson grass, but the genetic mechanisms behind these traits is still not fully understood (Basu et al., 2004). Discovering the genes and alleles underlying weedy traits would allow us to determine, at the genetic level, what makes a plant weedy and to characterize the weedy niche (Basu et al., 2004). By characterizing the genes underlying weedy traits, we can also determine to what extent common genetic mechanisms have evolved in the parallel evolution of weediness.

1.2 *Oryza* as a Model System

Cereal grasses are important both economically and scientifically, as the study of cereals has been a major driving force for research in genetics, development, and the evolution of plants (Paterson *et al*., 2005). Large databases and resources for different cereal species have been developed to answer questions in fields such as molecular evolution, diversity, and crop productivity (Paterson *et al*., 2005). Rice is a model system for the grasses due to its small genome size (389 Mbp), ability to be transformed, and the availability of two fully sequenced genomes of two subspecies of cultivated rice (*O. sativa indica* and *O. sativa japonica*; Goff *et al.*, 2002; Yu *et al.*, 2002). Additionally, there are large amounts of germplasm and molecular resources available (Paterson *et al*., 2005). Benefits to having these data available include construction of detailed physical

maps, assistance in understanding the biological function of rice genes and improvement of current rice cultivars (Wing *et al.*, 2005). Research into rice has also benefited from the creation of large expressed sequence tags (EST) databases along with the development of copious amounts of marker data including sequence tagged sites (STS), simple sequence repeats (SSR), and single nucleotide polymorphisms (SNP).

Within the genus *Oryza* there are twenty-two wild species that represent ten distinct genome types (Ammiraju *et al*., 2006). The two species of cultivated rice, *O. sativa* and *O. glaberrima*, are AA diploids, as are their wild ancestors *O. rufipogon* and *O. barthii* (Ammiraju *et al.*, 2006). Parallel evolution during rice domestication is possible, as rice domestication is thought to have occurred independently in at least three geographic locations, and cultivars share many domestication-related traits. In the West African savanna, *O. glaberrima* was domesticated less than 3,000 years ago (Sweeny & McCouch, 2007). In Asia, wild *O. rufipogon* gave rise to two separate domesticated divisions of *O. sativa*: *indica/aus* and *japonica/aromatic* (Figure 1.1; Caicedo *et al*., 2007; Garris *et al.*, 2005; Londo *et al.*, 2006). Within these two domestication events there are multiple genetically differentiated varieties, each of which has been selected for different grain traits (shorter, fatter versus longer, slender), fragrance, and appearance, with the *aus* groups showing the most phenotypic variation (Oka, 1988; Bhattacharjee *et al*., 2002). These genetic and phenotypic variations make rice a good system for answering evolutionary questions such as: What genes were under selection in the transition from wild to domesticated rice? What are the molecular mechanisms that control variety specific traits, such as flowering time variability?

1.3 Evolution of Weedy Rice

In addition to cultivated rice, the *Oryza* genus also contains weedy forms of rice that have arisen in multiple regions of the world and as such are also suitable for parallel evolution studies. The American southern rice belt, which includes Mississippi, Arkansas, Louisiana, southern Missouri, and Texas, is the largest producer of rice in the United States, with smaller crops grown in California (FAOSTAT, 2008). All of these rice-growing areas, along with others worldwide, are infested with weedy rice, a weedy type of rice that invades cultivated rice monocultures. Weedy rice is a major agricultural pest, as it is an aggressive competitor that spreads rapidly and reduces the quality of the rice harvest (Shivrain *et al*., 2010). Weedy rice is highly adapted to the agroecosystem and is hard to fight with herbicides, as it is closely related to cultivated rice. Additionally, some weedy rice populations have evolved resistance to herbicides either *de novo* or from crossing with genetically modified rice crops (Shivrain *et al*., 2009). The emergence of weedy rice is often associated with direct seeding, rather than hand transplanting of seedlings, and intensive irrigation (Bres-Patry *et al.*, 2001).

Several hypotheses have been put forth to explain the evolutionary origin of US weedy rice. These include contamination of cultivated fields with wild rice relatives, introgression of wild alleles into cultivated rice, or de-domestication of cultivated rice — the reversion of domesticated phenotypes into wild phenotypes — through feralization or accidental selection (Olsen *et al.*, 2007). The contamination of rice fields in the US with wild rice relatives is highly unlikely, as no wild *Oryza* are native to North America. Additionally, data has shown that US weedy rice as a whole shares most of its alleles with cultivated rice (Reagon *et al.*, 2010). This implies that US weedy rice has arisen

through a process of de-domestication, and has reacquired many traits enhancing weediness after their loss during domestication. Interestingly, weedy rice collected in the US does not share a recent evolutionary origin with cultivars grown in the United States, although there is evidence of low-level hybridization with US cultivars in some minor weedy groups (Reagon *et al*., 2010).

Within the US, there are at least two independent origins of weedy rice. A study of 48 STS markers found that two subpopulations of weedy rice most likely originated from *O. sativa indica* and *O. sativa aus* cultivars (Figure 1.1; Reagon *et al*., 2010). The Straw Hulled (SH) group most closely resembles cultivated rice with a straw-colored hull with no awns and slightly larger grains, and likely originated from *O. sativa indica.* The *O. sativa aus* descended Black Hulled & Awned (BHA) group often resembles *O. rufipogon*, with a black or brown colored hull, small grains, and long awns. Often both types of weedy rice can be found in the same rice field (Shivrain *et al*., 2010). The recent origin of weedy rice in the US (within the last 200 years) and the presence of population bottlenecks and multiple introductions makes US weedy rice a prime system for studying evolutionary processes. Weedy rice is also ideal for investigating the genetic basis of parallel evolution as some traits have evolved to similar ends despite differences in population origin, and it is possible that the de-domestication of weedy rice involved alternative mutations in loci under selection during domestication.

1.4 Definition of Weedy Traits

There are several traits that could potentially enhance of the ability of weedy rice to invade and persist in rice fields. These traits include increased seed shattering (seed

dispersal) compared to the crop as well as increased seed dormancy, differences in plant height, altered flowering times, hull coloration, and the presence of awns (Burgos *et al*. 2006; Delouche *et al*., 2007; Shivrain *et al*., 2010). Although no formal studies have shown that these traits increase weed fitness, instances can be imagined where these traits could be beneficial.

Nearly all weedy rice shatters its seeds while cultivated rice does not, as they have been selected to retain their seeds during the domestication process to make them easier to harvest (Purugganan 2009). At least two loci of large effect influencing degree of shattering have been cloned in cultivated rice (*sh4* (Li *et al*., 2006) and *qsh1* (Konishi *et al*., 2006), yet nothing is known about weedy rice alleles at these loci. Additionally, if shattered seeds that enter the soil prior to crop harvest can remain dormant in the soil, they can increase the likelihood of perpetuating infestation of a field by persisting in the seed bank and contaminating fields as they are disturbed. Studies have shown that although weedy rice can show a range in seed dormancy, most strains are moderately to highly dormant compared to most cultivars, where dormancy has been selected to be mild so as to prevent pre-harvest sprouting and encourage uniform germination (Gu *et al*., 2003). Several QTL studies have attempted to map genes controlling dormancy in crosses between cultivated rice and Asian weedy rice (Gu *et al.*, 2003; Gu *et al*., 2005a; Gu *et al*., 2005b; Gu *et al.*, 2005c).

In terms of plant growth, a taller weedy plant can shade out cultivated competitors, yet a shorter weedy plant may avoid detection by weed scouts. Thus, most weedy rice are either significantly taller or significantly shorter than domesticated rice (Shivrain *et al.*, 2010). Also, weeds with earlier flowering times can reproduce faster and

may be able to produce more seeds over its lifetime as well as disperse seeds prior to crop harvest (Sahli *et al.,* 2008). A recent study of several growth related traits in rice showed diversity in cultivated and weedy phenotypes. Within cultivated rice, the flowering date and overall height of three major cultivar groups, *aus*, *indica* and *japonica*, were fairly similar (ranges of 75 to 80 cm and 100 to 113 days; Reagon *et al.*, 2011). Within weedy rice, SH weeds tended to be shorter and flower earlier than the BHA weeds and their progenitors *indica*. Also, BHA weeds were a bit taller than their progenitors *aus*, and flowered significantly later. At least one gene for plant height in rice is known (*Sd1*; Monna *et al.*, 2002, Sasaki *et al*., 2002, Spielmeyer *et al.*, 2002) while several genes for flowering time have been identified (*Hd1* (Yano *et al.*, 2000); *Hd3a* (Kojima *et al.*, 2002); *Ehd1* (Doi *et al*., 2004)). Each of these genes is an excellent candidate for investigating weedy alleles contributing to growth habit differences.

Lastly, there appears to be two different combinations of hull color and awning present in weedy rice. As mentioned above, the SH population has a straw hull and no awn which allows it to blend in with the cultivated rice in a field and may aid in avoiding removal by farmers. The BHA population has a dark pigmented hull and a long awn, which may aid in avoiding detection once on the soil but may be noticeable to farmers when still on the plant. A study into barley awns has shown that photosynthesis is enriched in this organ, however, the exact function of the rice awn is unknown (Abebe *et al*., 2009). It is possible that rice awns are involved in photosynthesis, seed dispersal, or even defense against herbivores, but genes for awn presence or length have yet to be cloned from QTL studies (Hu *et al.*, 2011). However, two loci for hull color changes, either black to straw (*Bh4*; Zhu *et al.*, 2011) or gold furrowing on a straw hull

background (*ibf*; Cui *et al.*, 2007), have been identified. Nothing is known currently about weedy rice alleles at these loci.

1.5 Questions of Interest

Bollback & Huelsenbeck (2009) posed an important evolutionary question: "How often does parallel evolution occur when independently evolving lines are exposed to the same environmental challenge?" We can address this question using the weedy/cultivated rice system. The two main subpopulations of weedy rice (SH and BHA) both arose in agricultural fields in the US and were under similar selective pressures within the agroecosystem. Additionally, both weedy rice subpopulations appear to have evolved from closely related varieties of cultivated rice (*indica/aus*). The research I conducted helps to answer questions regarding the incidence of parallelism as a contributing factor to the acquisition of weedy traits in the two main populations of US weedy rice. I hypothesized that, due to the fact that both *indica* and *aus* rice subpopulations share recent common ancestors they should harbor similar genetic biases and, as a result, both weedy rice subpopulations should possess similar genetic mechanisms for acquiring weedy traits.

The ultimate goals of my thesis were threefold: 1) to uncover the genetic mechanisms behind a convergent weedy trait (seed shattering); 2) to investigate the genetic mechanisms behind a divergent weedy trait (flowering time); and 3) to determine whether both weedy rice types possess novel alleles at genes known to have been selected upon during rice domestication. My main focus has been on the seed shattering trait, as it is important for weed proliferation. I first focused on exploring the genetic basis of these

traits through the study of candidate genes (*qsh1/sh4*) in chapter 2. After finding convergence of the weed shattering phenotype with wild rice, *O. rufipogon*, but mediated through different genetic mechanisms, I explored the morphological basis of the shattering trait in chapter 3. Concurrently, I also explored the genetic basis of flowering time in weedy rice by investigating candidate genes (*Hd1 and Hd3a*) in chapter 4 and found that the divergent weed phenotypes could not be completely explained by these two genes alone. The lack of well characterized major effect genes contributing to the weed phenotypes prompted me to attempt to identify genomic regions underlying weedy trait evolution through QTL mapping in chapter 5, and assess the degree of parallel genetic evolution in weedy groups. My dissertation research lays the basis for discovery of the genes underlying the evolution of weediness and the occurrence of parallel evolution in closely related agricultural weeds.

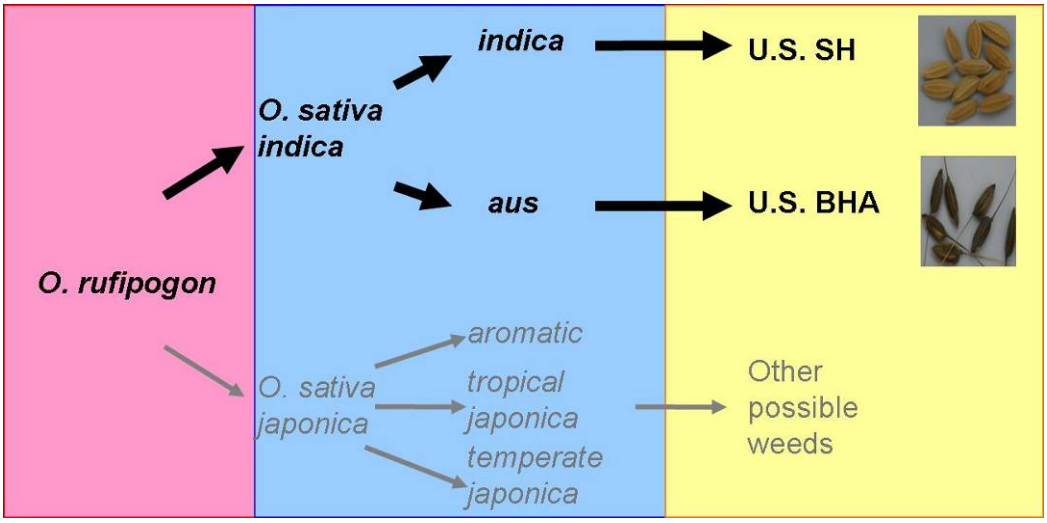


Figure 1.1 Relationships between wild, cultivated and weedy rice.

During the domestication of wild *Oryza rufipogon* (pink), two major cultivated divisions of *Oryza sativa* (*indica/japonica*) arose and subsequently diversified into different varietal groups (blue). Weedy rice (yellow) from the Southern US likely arose from cultivated ancestors *indica/aus*, while some weedy rice found outside the US is more closely related to cultivars of the *japonica* domestication lineage.

CHAPTER 2

MOLECULAR EVOLUTION OF SHATTERING LOCI IN UNITED STATES WEEDY RICE

2.1 Introduction

Invasive weeds that colonize agricultural fields cost millions of dollars in crop losses and weed control measures every year. Many of these agricultural weeds share similar fitness-related traits that make them highly competitive with crop species. For example, rapid growth, deep roots, high seed production and increased seed dispersal allow weeds to acquire more resources, as well as to produce more offspring (Basu *et al.*, 2004). Efficient seed dispersal, in particular, may be a trait crucial to weed fitness. By increasing seed dispersal via 'shattering' or scattering their seeds, weeds can increase their presence in the seed bank and spread into new areas (Harlan & DeWet, 1965). Plants that shatter their seeds within agricultural fields can often avoid collection by farmers, and subsequent seed consumption/destruction, thus persisting within fields. Additionally, shattering at maturity is sometimes necessary to retain sufficient seed moisture for dormancy, a trait favored in agricultural weeds for winter survival and germination during the cropping season (Delouche *et al.*, 2007; Gu *et al.*, 2005b; Gu *et al.*, 2005a).

Most wild cereals, including wild relatives of rice, wheat and barley, have brittle, easily shed (shattering) seeds. Cultivated cereals, however, have undergone selection for reduction of shattering during the domestication process, to increase the amount of seed harvested by humans (Harlan & DeWet, 1965). Reduced seed shattering is thought to be among the earliest and most important traits selected upon during grain domestication

(Fuller *et al.*, 2009; Harlan, 1992). A reduction in seed shattering may have been favored over complete non-shattering to minimize labor during harvest (Li *et al.*, 2006b; Sang & Ge, 2007a). The shattering trait is thus under strong opposing selection in agricultural environments, with high levels of shattering favored in invasive weeds and reduced shattering in cultivated crops.

Weedy or red rice is a weedy type of rice (*Oryza sativa* L.) that invades cultivated rice fields and costs United States farmers millions of dollars each year (Burgos *et al.*, 2008). Weedy rice is an aggressive competitor, decreasing yields and contaminating rice harvests with off-color, brittle grains (Burgos *et al.*, 2006; Cao *et al.*, 2006). The appearance of weedy rice has been associated with a transition to direct seeding, and it is present worldwide, wherever rice is cultivated (Bres-Patry *et al.*, 2001; Olsen *et al.*, 2007). Although morphologically diverse, a suite of possible weediness-enhancing traits tends to characterize weedy rice in the field; these include the presence of red pericarps (bran), high levels of dormancy, and high levels of seed shattering (Delouche *et al.*, 2007; Vaughan *et al.*, 2001; Gealy *et al.*, 2003). Several of these traits are also found in the wild ancestor of cultivated rice, *O. rufipogon*, and other wild *Oryza* relatives, but weedy rice differs from truly wild species in its adaptation to the agroecosystem and presence of some traits characterizing cultivated rice (e.g. high selfing rate (Delouche *et al.*, 2007)).

There are multiple efforts underway to understand the worldwide origins of weedy rice groups. Hypotheses range from invasion of wild *Oryza* relatives, to hybridization among wild and cultivated groups, or de-domestication of cultivated rice varieties (Bres-Patry *et al.*, 2001; Gealy, 2005). In the United States, weedy rice is prevalent in the rice growing regions of the southern Mississippi basin (Gealy, 2005). No

Oryza species is native to the US, and the evolutionary origin of US weedy rice has been a source of debate since it was first documented in the 1840s (Delouche *et al.*, 2007). Previous assessments of genetic diversity have determined that several populations of morphologically divergent weedy rice are present in the US (Gealy *et al.*, 2002; Londo & Schaal, 2007; Reagon *et al*., 2010). The main populations of US weedy rice, designated after their most common grain morphology, include the straw-hulled (SH) group, characterized by straw-colored hulls, high yielding panicles and lack of awns, and the black-hulled awned (BHA) group, characterized by its greater height, black hulls and long awns (Gealy *et al.*, 2002). The BHA group is subdivided into two genetically distinct subpopulations, BHA1 and BHA2 (Reagon *et al.*, 2010). A third group (BRH), characterized by brown hulls, is most likely a result of hybridization between the SH and BHA groups (Reagon *et al*., 2010).

Studies have shown that US weedy rice shares most of its genome with Asian cultivated rice (Londo & Schaal, 2007; Reagon *et al*., 2010). Interestingly, US weedy rice does not share a recent evolutionary origin with cultivars grown in the US, which belong to the *tropical japonica* variety group, though there is evidence for limited hybridization (Reagon *et al*., 2010; Gealy *et al.*, 2009). Instead, studies suggest that SH weeds are most closely related to *indica*, a cultivated rice variety typical of lowland tropical regions, while the BHA groups share a closer relationship with *aus*, a rapidly maturing, photoperiod insensitive rice variety from Bangladesh and Northeastern India. However, neither of these crop varieties has been cultivated in the southern US. Moreover, though patterns of genome-wide variation suggest that weedy rice is not directly descended from

wild rice (Reagon *et al.* 2010; Gealy *et al.*, 2009), questions about possible contributions of wild rice to US weedy rice evolution remain.

Recently, candidate genes underlying some domestication-related traits have begun to be identified in cultivated rice (e.g.: Fan *et al.*, 2006; Gu *et al.*, 2008; Xing *et al.*, 2008). Because these traits often differ between cultivated rice and wild/weedy relatives, candidate genes have opened up new sources of potential information about the evolution of weediness-enhancing traits. Combined with information about genome-wide patterns of polymorphism, candidate genes may help provide a complete picture of the evolutionary origin of weedy rice groups. A recent investigation into a pericarp color candidate gene, *Rc*, revealed that US weedy rice groups carry alleles distinct from those in sampled cultivated or wild rice groups (Gross *et al*., 2010). Although genomic data suggests that US weedy rice originated from cultivated rice varieties, *Rc* data suggests that weeds are not direct descendants of cultivated rice (Gross *et al.*, 2010; Reagon *et al*., 2010). However, because different key traits may have been selected at different stages of the domestication process (Purugganan & Fuller, 2009), weedy rice alleles at important domestication loci may tell complementary stories about the origins of weedy rice.

As a trait crucial to modern cultivation and harvesting practices, there has been great interest in discerning the genetic basis of seed shattering in rice. To date, two quantitative trait loci (QTL) of large effect have been cloned, *qsh1* and *sh4/SHA1*, each explaining over 70% of the variation in their respective crosses. The *qsh1* locus is a homeodomain gene, similar to *Arabidopsis thaliana* REPLUMLESS, which was isolated in a cross between two *O. sativa* varieties, *aus* and *temperate japonica*, that differ in their

shattering propensity (Konishi *et al.*, 2006). A single nucleotide substitution in the regulatory region of the gene decreases the shattering ability in a subset of cultivated *temperate japonica* rice (Konishi *et al.*, 2006; Zhang *et al.*, 2009).

The *sh4* gene, encoding a nuclear transcription factor, was isolated from a cross between cultivated *O. sativa indica* and a wild species, *O. nivara*, and is involved in the degradation of the abscission layer between the grain and the pedicel (Li *et al.*, 2006b; Lin *et al.*, 2007). Highly shattering *O. nivara* possess very defined abscission layers, while non-shattering cultivated rice groups possess discontinuous abscission layers (Ji *et al.*, 2006; Li *et al.*, 2006b). A single nonsynonymous substitution (G/T) in the second exon of *sh4* has been shown to lead to diminished DNA binding with the SH4 protein and incomplete development of the abscission layer in non-shattering rice (Li *et al.*, 2006b). Transgenic *japonica* plants expressing the wild *O. nivara* allele show a significantly increased ability to shatter (Li *et al.*, 2006b). Shattering QTL in the *sh4* genomic region have been consistently identified in studies involving other crosses between cultivated varieties and wild rice (Cai & Morishima, 2000; Xiong *et al.*, 1999).

Sh4 is considered the most significant shattering gene to have been selected upon during domestication (Li *et al.*, 2006b; Purugganan & Fuller, 2009). Examination of *sh4* alleles has shown that all cultivated rice sampled to date shares the non-shattering T mutation, and most rice individuals share a common *sh4* haplotype, despite the fact that at least two separate domestication events gave rise to cultivated Asian rice (Li *et al.*, 2006b; Zhang *et al.*, 2009). The sharing of a common *sh4* haplotype across divergent rice varieties has been attributed to a combination of introgression and strong positive selection (selective sweep) favoring a reduction in shattering in the crop during both

domestication processes (Sang & Ge, 2007a; Sang & Ge, 2007b; Zhang *et al.*, 2009; Li *et al.*, 2006).

Here we assess patterns of polymorphism in weedy rice groups at the identified shattering genes and targeted flanking genomic regions, to determine the possible origin of the shattering phenotype in the US weed and contribute to understanding of US weedy rice evolution. The goals of the present study were to 1) assess levels of shattering in US weedy rice groups, 2) determine the origin of US weedy rice alleles at *qsh1* and *sh4*, and 3) determine the role each locus may play in the shattering phenotype of weedy rice. We find that the shattering associated single nucleotide polymorphism (SNP) at *qsh1* has not played a role in the evolution of weedy rice, as all weeds, wild rice, and most cultivars share the ancestral allele at this locus. Moreover, although cultivated and weedy rice groups differ greatly in their shattering ability, all sampled weedy and domesticated accessions possess similar or identical alleles at the *sh4* locus, suggesting that the domestication-associated T substitution at *sh4* is not sufficient for loss of shattering. Our data supports a direct origin of US weedy rice groups from domesticated ancestors, and implies that genetic changes at other loci must be responsible for the re-acquisition of the shattering trait during the weed's evolution.

2.2 Methods

2.2.1 Plant Material

A phenotypically diverse sample of 58 weedy rice accessions, collected in the Southern US rice belt, was generously supplied by David Gealy (USDA) (Table 2.2). An additional 87 samples of diverse *Oryza* species were included in the study as potential

sources of weedy rice alleles. Cultivated rice accessions belong to five variety groups of Asian *O. sativa*: *indica* (9 samples), *aus* (7), *tropical japonica* (8), *temperate japonica* (4), and *aromatic* (3). Thirteen additional accessions of *tropical japonica* cultivars grown in the U.S were included. Other *Oryza* included geographically diverse samples of *O. rufipogon* (30), the wild ancestor of cultivated Asian rice, *O. nivara* (2), an annual plant that some consider an ecotype of *O. rufipogon* (Zhu & Ge, 2005), *O. glumaepatula* (2), a wild rice from South America, *O. glaberrima* (4), cultivated African rice, and *O. barthii* (2), the wild ancestor of domesticated African rice. *O. meridionalis*, a species native to Oceania, was included as an outgroup. All plants were grown for DNA extraction as described in Reagon *et al.* (2010).

2.2.2 Measurement of the Shattering Phenotype

A subset of 90 *Oryza* accessions, representing selfed progeny of plants grown for DNA extraction, was grown for phenotyping in a completely randomized block design in two Conviron PGW36 growth chambers at the University of Massachusetts Amherst (Table 2.2). Two seeds per accession, one per chamber (block), were planted in 4-inch pots and randomly assigned locations within a chamber. Watering and fertilizer schedules were the same in both chambers and plants were exposed to 12-hour light/ dark cycles. Upon heading, typically two to three months after germination, panicles were bagged to prevent pollen flow and loss of seeds. At 30 days after heading, panicles were tested for shattering using a digital force gauge (Imada, Northbrook, IL). Shattering was measured as Breaking Tensile Strength (BTS) (Konishi *et al.*, 2006; Li *et al.*, 2006b), which is the amount of weight a seed can bear before releasing from the pedicel at the abscission layer. Briefly, panicles were suspended from a ring stand and an individual

seed clipped with a small (~1 g) binder clip. Seeds that released at or prior to this point were recorded as zeros and considered highly shattering. For seeds remaining on the panicle, the force gauge was hooked onto the binder clip and the peak measurement upon grain removal was recorded. Preliminary trials revealed that considerable variation could occur within panicles of cultivated varieties; thus, 25 randomly chosen seeds per plant were measured across two panicles and averages were calculated for each individual. Chamber effects on shattering were non-significant ($P > 0.15$), as determined by a Kruskal-Wallis non-parametric rank test, and were not considered in subsequent analyses.

2.2.3 DNA Extraction, Genotyping, and Sequencing

DNA was extracted as described in Reagon et al (2010). CAPs markers (Neff *et al.*, 2002) were used to determine the *qsh1* allele in all individuals (Table 2.3). Variation at *sh4* was determined by DNA sequencing of the entire open reading frame, the promoter and a downstream region of the gene (Table 2.3). Additionally, six ~500 base pair (bp) regions of genes increasingly distant from the *sh4* locus (several kilobase pairs (Kb) to several megabase pairs (Mb)) were sequenced spanning a region of 5.6 Mb (Table 2.3). Primers were generated using Primer3 (Rozen & Skaletsky, 2000) based on the *O. sativa japonica* (var. Nipponbare) genome (TIGR v. 5 January, 2008). Initial PCR amplification and DNA sequencing was performed by Cogenics (Houston, TX) as described previously (Caicedo *et al.*, 2007; Olsen *et al.*, 2006). Additional PCR amplification was performed on a 500 bp region surrounding the loss-of-shattering associated SNP using LA Taq and GC rich buffer (TaKara) with added glycerol and DMSO. Sequence alignment, including base pair calls, quality score assignment and construction of contigs, was performed as described previously (Caicedo *et al.*, 2007)

using BioLign Version 2.09.1 (Tom Hall, NC State Univ.). DNA sequences obtained for this study have been deposited in GenBank under accession numbers GU220907-GU221904.

2.2.4 Data Analysis

Summary statistics for the *sh4* locus and flanking genes for each population of interest were calculated as described previously (Caicedo *et al.*, 2007). Statistics include Watterson's estimator nucleotide variation (θ_W), the average pairwise nucleotide diversity (θ_π) (Nei & Li, 1979), and Tajima's D (Tajima, 1989) for silent, synonymous, nonsynonymous, and total sites (Table 1). Site type determination was based on annotations of the *O. sativa* genome (TIGR v. 5 January, 2008). Significance of Tajima's D values was tested using DNAsp (Rozas *et al.*, 2003). Genealogical relationships among *sh4* alleles and flanking fragment alleles were determined with Maximum Parsimony (MP) and Neighbor Joining (NJ) analyses as implemented in MEGA 4 (Tamura *et al.*, 2007). Both analyses considered pairwise deletion of gaps/missing data. Distances were calculated using the Kimura 2-parameter model; branch bootstrap estimates were obtained from 1000 replicates. Heterozygotes were rare in our dataset, occurring occasionally only in *O. rufipogon*. When present, heterozygotes were phased using PHASE 2.1 prior to phylogenetic analyses (Stephens *et al.*, 2001; Stephens & Scheet, 2005), and no ambiguity was observed. For all loci, both NJ and MP trees produced similar results, so only the NJ trees are shown. Extended Haplotype Homozygosity (EHH) across the sampled genomic region containing *sh4* was calculated as described by (Sabeti *et al.*, 2002), to test for extended linkage disequilibrium around the putatively selected mutation and assess the possibility of a selective sweep.

2.3 Results

2.3.1 The Shattering Phenotype in Weedy, Wild and Cultivated Rice

We recorded the degree of seed shattering for 90 accessions representing multiple groups of weedy, wild, and cultivated *Oryza* (Table 2.2; Figure 2.1). Degree of shattering is a quantitative and highly variable trait (Ji *et al.*, 2006). Our measurements revealed that some cultivated rice individuals show high variability in shattering within a single panicle (Table 2.2), with BTS for individual seeds occasionally varying by 10 to 200 grams (g); however, extreme differences in BTS values, when present, occur for very few seeds within a panicle. In contrast, variation in shattering levels within panicles is much lower in weedy and wild rice accessions (Table 2.2). For all samples, mean and median shattering values are typically within 10 g.

Mean shattering differences among all measured *Oryza* accessions ranged widely, with values close to 0 g corresponding to a highly shattering phenotype, and values close to 100 g corresponding to complete non-shattering (Figure 2.1; Figure 2.5). In practice, BTS values of 5 g or less are considered shattering, as these seeds can be easily brushed off during measuring device attachment. Broad differences were observed across *Oryza* groups, and a Kruskal-Wallis test confirmed that variety has a significant effect on shattering levels ($P = 0.0013$).

Although lack of shattering is a hallmark of rice domestication, cultivated Asian rice varieties display a range of seed shattering phenotypes, with BTS values ranging from nearly zero to 140 g (Figure 2.1; Table 2.2). The *aus* group, in particular, shows a much narrower range of values (0-50 g), compared to *indica* (5-140 g) and *tropical*

japonica (10-120 g) (Figure 2.1). Additionally, one *indica* and one *aus* accession in our sample have average BTS values less than 5 g and may be considered shattering.

In contrast to cultivated rice, almost all of the wild Asian rice, *O. rufipogon* and *O. nivara* (Figure 2.1), show BTS values of zero, indicating that the species are highly shattering. All weedy rice accessions, with the exception of a single individual (1B06, Table 2.2), show a propensity to shatter, registering BTS values very close to zero. Non-shattering weedy accession 1B06 has been shown to possibly have mixed ancestry (MIX) (Reagon *et al*., 2010), and may have acquired additional non-shattering alleles through hybridization with cultivated rice. A single observed non-shattering *O. rufipogon* accession (2C04), on the other hand, does not resemble cultivated rice phenotypically or genetically (Reagon *et al.*, 2010), suggesting that the non-shattering phenotype is not due to introgression from the crop.

2.3.2 Diversity at the *qsh1* Locus

We genotyped *Oryza* accessions at the *qsh1* locus, to determine whether the previously identified mutation (Konishi *et al.*, 2006) might play a role in the shattering phenotype of weedy rice. All weeds and the majority of rice cultivars were found to have the ancestral SNP, which also characterizes *O. rufipogon* and wild rice species, and is associated with higher levels of shattering (Table 2.2). Consistent with results from other research groups, we find that the non-shattering mutation is limited to two of our accessions belonging to the *temperate japonica* group (Table 2.2), and that the SNP is most likely not involved in variation in shattering levels outside of a small group within this cultivated variety (Konishi *et al.*, 2006; Zhang *et al.*, 2009).

2.3.3 The Genealogy of *sh4*

To determine if the shattering locus, *sh4*, may underlie variation in shattering levels among cultivated and weedy rice, we sequenced the gene in a panel of 144 samples from weedy, cultivated and wild rice groups. The 3.9 kb of aligned sequence data includes the intron and both exons, plus 1040 bp of the promoter region upstream, and 550 bp downstream of the *sh4* gene.

Relationships among haplotypes at the *sh4* locus (Figure 2.2; Figure 2.5) reveal a highly supported clade defined by the derived T mutation. As observed in previous research (Li *et al.*, 2006b; Zhang *et al.*, 2009), all cultivated rice accessions sampled carry this mutation, which is associated with loss of shattering. Moreover, the majority of cultivated rice accessions share an identical haplotype across the 3.9 Kb *sh4* region that we characterized. Three cultivars in our sample, one *aromatic*, one *tropical japonica* and one *aus*, differ from the common cultivated *sh4* haplotype by two, one and one nucleotide substitutions, respectively (Figure 2.2; Figure 2.5). These four SNPs have not been reported in other studies of the *sh4* locus to date, despite the detection of at least seven other low-frequency cultivated *sh4* haplotypes not detected here (Zhang *et al.*, 2009). The two *aromatic* SNPs were the only ones found to occur in coding regions; one substitution alters amino acid 104 from a polar Serine to non-polar Tryptophan, possibly resulting in the shattering phenotype in this individual (Figure 2.6).

Eighteen *sh4* haplotypes were observed within wild *O. rufipogon* accessions. While the majority of the detected haplotypes are divergent from cultivated *sh4* alleles, six accessions carry an identical haplotype as the majority of cultivated rice, and two accessions carry haplotypes that differ by only one and three SNPs from this cultivated

haplotype (Figure 2.2; Figure 2.5). Additionally, both *O. nivara* accessions sampled in this study have the same haplotype as the majority of cultivated rice (Figure 2.5). These wild accessions were all found to shatter their seeds (BTS ~0 g, Table 2.2). The existence of shattering rice with the non-shattering T allele at *sh4* has not been previously reported (Li *et al.*, 2006b; Zhang *et al.*, 2009), and indicates that the presence of this mutation alone is not sufficient to confer a reduction in shattering. Surprisingly, the single non-shattering *O. rufipogon* individual in our sample (2C04) does not carry the T mutation in *sh4*.

Contrary to our expectations, given their high propensity to shatter, all weedy rice accessions sampled carry the non-shattering associated T nucleotide in *sh4*. Moreover, the majority of weedy rice accessions, ~70%, have *sh4* haplotypes identical to the most common haplotype in cultivated rice. Four additional novel *sh4* haplotypes were detected in the 18 remaining accessions of weedy rice. Each of the four unique haplotypes differs from the main cultivated haplotype by a single SNP and is not shared with any cultivated or wild rice groups (Figure 2.5; Figure 2.6). Additionally, three of these SNPs are predicted to cause amino acid replacements and may have functional consequences.

2.3.4 Genealogy of the *sh4* Genomic Region

To further elucidate the possible origin of *sh4* alleles in weedy rice, we examined phylogenetic relationships at loci increasingly distant from *sh4* in both the 5' and 3' directions in the genome. Six ~500 bp loci were chosen for analysis, spaced 7.9 kb, 600 kb, and 1.2 Mb from *sh4* on the 5' side of the gene and 300 kb, 1.1 Mb and 2.4 Mb from *sh4* on the 3' side of the gene (Table 2.3). Further exploration on the 5' side of *sh4* was not carried out, as the final fragment is within 50 kb of the telomere and only one other

gene exists within this region. Two additional loci downstream of *sh4*, sts_040 and sts_021, examined in a previous study (Reagon *et al*., 2010), were also included in our analyses. The furthest locus, sts_021, is 7.9 Mb away from the centromere; thus, our sampling encompasses over two-thirds of the chromosome arm containing *sh4* (~16 Mb). Phylogenies of the eight selected loci surrounding *sh4* were produced to visualize changes in relationship of weedy, wild and cultivated alleles with distance from the *sh4* locus (Figure 2.2). Because of their likely hybrid origin and rarity in US rice fields (Reagon *et al*., 2010), BRH and MX groups were excluded from these phylogenetic analyses.

The resolution of relationships among *Oryza* groups varies greatly according to the diversity at each locus (Figure 2.2). Because multiple sources of evidence support a minimum of two separate rice domestication events (e.g. Sang & Ge, 2007a; Vaughan *et al*., 2008), we examined the *sh4* genomic region to determine at what point cultivated groups began to separate into distinct clades. Similar to *sh4*, most cultivated rice individuals share a single haplotype in the two closest flanking fragments sampled (sh4f_003 and sh4f_004; Figure 2.2). This is consistent with hitchhiking of linked regions during selection on *sh4*; however, these loci are also highly conserved within all *Oryza* (Table 2.3). In the region upstream of *sh4*, multiple clades of domesticated rice appear ~600 kb (fragment sh4f_002), primarily due to diverse haplotypes in the *aus* and *japonica* groups. This trend continues 1.1 Mb upstream (fragment sh4f_001), but a clear separation into the two domesticated clades (*japonica* vs. *aus* and *indica*) is not seen. Downstream of *sh4*, greater haplotype diversity among cultivars is evident in fragment sh4f_005, ~1.1 Mb away and the remaining fragments. However, unlike many STS

fragments previously examined (Caicedo *et al*. 2007), strong divergence of the two domesticated clades is not observed, with haplotype sharing evident among cultivated varieties in the sampled regions. This suggests that the effect of positive selection on *sh4* during rice domestication is evident throughout the genomic region sampled (see below).

In most fragments flanking *sh4*, weedy rice groups share haplotypes with cultivated rice varieties (Figure 2.2). As expected, weedy groups tend to share haplotypes with their putative ancestors; thus, the majority of SH weeds group with *indica* cultivars (e.g. fragment sh4f_001), and the majority of BHA1 and BHA2 weeds group with *aus* cultivars. However, novel weed haplotypes were also observed in some fragments sampled; for example, some BHA1 and BHA2 weeds (11 accessions) share an identical haplotype in fragment sh4f_002 not seen in any other *Oryza* group. Moreover, in nearly every clade containing both weeds and cultivars, some wild *Oryza*, principally *O. rufipogon* or *O. nivara*, is also present (Figure 2.2).

Because a simple look at genealogical relationships within individual fragments in the *sh4* genomic region does not immediately reveal the source of weedy *sh4* alleles, we examined concatenated SNP haplotypes across the region (Figure 2.3). Within 6.2 Mb (up to sts_040) surrounding *sh4*, 13 SH weed accessions are identical to a single *indica* accession (2B02), and seven SH weeds and a single BHA1 weed are identical to three *indica* cultivars. Additionally, two BHA1 and four BHA2 accessions are identical to a single *aus* accession (3A06). When the region 14 Mb away from *sh4* is included (sts_021) only the weeds identical to the *aus* accession remain grouped, indicating a breakdown of the other associations due to recombination. The lack of extended

haplotype sharing between weeds and *tropical japonica*, suggests that weeds cannot have acquired *sh4* alleles through introgression with the local crop.

We also examined concatenated SNP haplotypes for *O. nivara* or *O. rufipogon* accessions sharing the common domesticated *sh4* haplotype. The seven SH and single BHA1 accessions that share extended haplotypes with the three *indica* cultivars, are identical to a single *O. nivara* (2F01) and a single *O. rufipogon* (2C09) across a 6.2 Mb region (Figure 2.3). Once the region 14 Mb away is added, these two wild accessions no longer group with the weeds yet still group with two *indica*s. Of the remaining wild accessions, a single *O. rufipogon* (2D06) is identical to a single *indica* (3A11) accession, but none possess haplotypes identical to weeds or cultivars across the *sh4* genomic region. The greater sharing of extended haplotypes between weeds and cultivars than between weeds and wild rice strongly suggest that weedy rice populations have inherited the derived *sh4* T substitution from domesticated ancestors.

2.3.5 The Impact of a Selective Sweep in the *sh4* Genomic Region

The ubiquity of the derived *sh4* T substitution among cultivated rice accessions, stemming from multiple domestication events, suggests that *sh4* has been subjected to strong positive selection during the domestication process (Vaughan *et al.*, 2008; Zhang *et al.*, 2009). To determine how positive selection on *sh4* in cultivated rice has affected *sh4* diversity in weedy rice groups, we assessed levels of genetic diversity at the sampled regions. As expected, silent site nucleotide diversity at *sh4* is very low in cultivated rice (Figure 2.7, Table 2.1). Values for *indica*, *aus* and *tropical japonica*, the three rice varieties most likely to have contributed to weedy rice, are all over an order of magnitude smaller than genome-wide averages estimated from a set of 111 STS loci (1.9, 1.9, and

1.6 per kb, respectively) (Caicedo *et al.*, 2007). A recent study reported higher levels of *sh4* variation in cultivated groups, but still well below genome averages (Zhang *et al.*, 2009). Conversely, *sh4* nucleotide diversity in *O. rufipogon* (Table 2.1) is close to the genomic average (~5.2 per kb) (Caicedo *et al.*, 2007) and in line with the diversity seen in Zhang *et al.* (2009).

The three main groups of US weedy rice also show a reduction in nucleotide diversity at *sh4*, but the level of reduction differs among groups. Silent site nucleotide diversity values for SH, BHA1 and BHA2 range from 0 to 0.2 per kb (Table 2.1), while their genome wide averages based on 48 STS loci are 0.692, 0.829 and 0.657, respectively (Reagon *et al*., 2010). In general, the reduction in diversity at *sh4* compared to genomic values in weedy rice groups is less drastic than in cultivated rice, perhaps due to the genome-wide low levels of diversity associated with the bottlenecks giving rise to weedy groups (Reagon *et al*., 2010). Surprisingly, the BHA2 group showed only a mild decrease in diversity at *sh4* and a positive Tajima's D (Table 2.1), consistent with the presence of two moderate frequency haplotypes.

In cultivated and most weedy rice groups, there is also a decrease in diversity, to differing degrees, in genes flanking *sh4* (Table 2.1). The majority of loci sampled show diversity below the genome average within all cultivars. The *indica*, *aus* and *tropical japonica* groups have negligible amounts of diversity in fragments sampled up to 1.2 Mb on 5' side (sh4f_002) and 1.1 Mb on 3' side (sh4f_005) (Table 2.1), consistent with a selective sweep in the region. However, these fragments also show low levels of diversity in *O. rufipogon*, in line with overall reduced diversity previously reported on this arm of chromosome 4 (Mather *et al.*, 2007). Remarkably high levels of diversity are evident in

the furthest locus sampled from the *sh4* gene, sts_021, which shows a particularly drastic increase in diversity in *indica* and *tropical japonica* varieties.

Most loci sampled in the *sh4* genomic region show no diversity in the three major weedy rice populations, consistent with the proposed bottlenecks at founding (Reagon *et al*., 2010). Remarkably, however, some fragments in the *sh4* genomic region display higher levels of diversity in weedy groups than their putative progenitors (Table 2.1; Figure 2.7). In particular, the BHA2 group is highly diverse at *sh4* and locus sh4f_002; because some BHA2 haplotypes at these loci are not found in other cultivated or wild *Oryza* groups sampled (Figure 2.2; Figure 2.5), high diversity levels may be due to inheritance from diverse unidentified ancestors, or new mutations since the origin of the weedy group.

To further assess the genomic impact of selection on the *sh4* T substitution in cultivated rice, and subsequent inheritance in weedy rice, we determined the breakdown of linkage disequilibrium (LD) across the *sh4* region using the Extended Haplotype Homozygosity (EHH) analysis (Sabeti *et al.*, 2002). As expected, homozygosity breaks down most quickly for the *O. rufipogon* group possessing the ancestral G substitution in *sh4*, within 100 bases of the SNP (Figure 2.4A). For *O. rufipogon* accessions containing the derived T substitution, breakdown occurs more slowly, consistent with its derived status. For both groups homozygosity is at or near zero within 1.1 Mb downstream of the mutation.

In contrast to wild rice, and indicative of strong positive selection on *sh4*, cultivated rice groups all have more extensive haplotype homozygosity throughout the examined genomic region (Figure 2.4B). Particularly noteworthy is the fact that *indica*

shows no breakdown of homozygosity within *sh4*, although the *aus* and *tropical japonica* groups do. No group reached an EHH value of zero upstream of *sh4* within the region sampled; however, downstream of the gene, *tropical japonica* is the first group to reach a homozygosity value of zero. These patterns of LD suggest that *sh4* originated in the ancestors of *tropical japonica* and subsequently introgressed into *indica*, where there may have been less time for recombination to lead to breakdown of LD.

Homozygosity patterns for weed groups in the *sh4* genomic region are similar to those of the cultivars above but show a much slower breakdown of LD overall (Figure 2.4C). Unlike cultivated rice, however, all weedy groups possess unique SNPs within *sh4*. This accounts for the initial breakdown of homozygosity within the gene. The high levels of homozygosity observed for weedy groups are consistent with inheritance of *sh4* alleles from ancestors with low levels of diversity and high levels of LD within the *sh4* genomic region.

2.4 Discussion

The loss of shattering as a seed dispersal mechanism is a key domestication trait, distinguishing cultivated cereals from their wild relatives. Seed shattering is also a trait associated with weed fitness, with increased levels of seed dispersal likely favored in weeds infesting agricultural systems (Harlan & DeWet, 1965). Recent advances dissecting the genetic basis of seed shattering variation in cultivated and wild rice (Konishi *et al.*, 2006; Li *et al.*, 2006b; Lin *et al.*, 2007) offer a unique opportunity to assess the evolution of this fitness-related trait in populations of weedy rice.

Multiple populations of weedy rice with independent origins exist in the US (Londo & Schaal, 2007; Reagon *et al.*, 2010; Gealy *et al.*, 2009). Surveys of

polymorphism have shown that the main populations of US weedy rice share genetic backgrounds with, and are possibly descendants of, *indica* and *aus* cultivated rice varieties (Londo & Schaal, 2007; Reagon *et al*., 2010). We have confirmed that all US weedy rice populations are highly shattering (Figure 2.1). The near complete lack of variability in this trait across weedy rice groups contrasts with the variance in shattering levels in cultivated rice varieties. The fact that all weedy rice shatters, despite separate origins of major weedy rice groups, suggests that shattering is a trait strongly selected for during weedy rice evolution. Coupled with genomic data indicating weedy rice origins from non-shattering ancestors, this pattern gives rise to questions about how weeds have acquired the shattering trait.

Environmental variation is known to affect the seed shattering trait in cultivated rice (Ji *et al.*, 2006), and thus our shattering measurements could differ from those obtained under field conditions. However, extensive qualitative assessments of US weedy rice in single and multiple US rice fields report the US weed as highly shattering (e.g. Gealy *et al.*, 2003; Noldin *et al.*, 1999; Delouche *et al.*, 2007; Oard *et al.*, 2000). Thus, our growth-chamber measurements of shattering levels in weedy rice seem consistent with observations in the weed's native environment. Likewise, multiple studies report wild rice as highly shattering in field conditions examined outside of the US (e.g. Cai & Morishima, 2000; Xiao *et al.*, 1998), consistent with our results. Lastly, our measurements of US cultivated *tropical japonica* varieties are consistent with low shattering levels of the crop in US rice fields (Table 2.2). Thus, our measurements of shattering under growth chamber conditions seem to accurately reflect known phenotypes of weedy and cultivated rice in US fields.

To date, two loci of large effect have been shown to underlie the seed shattering trait in cultivated rice: *qsh1* and *sh4* (Konishi *et al.*, 2006; Li *et al.*, 2006b; Lin *et al.*, 2007), As reported by others (Konishi *et al.*, 2006; Zhang *et al.*, 2009), we found that the *qsh1* shattering associated SNP is only relevant to shattering variation within the cultivated *temperate japonica* group, where some individuals possess a derived mutation associated with extreme loss of shattering. All US weedy rice individuals possess the ancestral allele that is common in all non-*temperate japonica* cultivated and wild rice groups (Table 2.3).

In contrast to *qsh1*, *sh4* is considered to be a key gene under strong selection during rice domestication (Zhang *et al.*, 2009). We found that all cultivated rice individuals examined are fixed for a T substitution in exon 1 of *sh4* (Figure 2.5), which is associated with loss of shattering (Li *et al.*, 2006b). Moreover, consistent with prior observations (Li *et al.*, 2006b; Lin *et al.*, 2007; Zhang *et al.*, 2009), the majority of rice cultivars share an identical haplotype at *sh4*, suggesting a single origin of the non-shattering allele in domesticated rice. Surprisingly, despite their ability to shatter, our survey revealed that all US weedy rice accessions carry the T substitution associated with non-shattering at *sh4*, and that most weeds share the common cultivated *sh4* haplotype (Figure 2.2; Figure 2.5). This demonstrates that the T substitution characteristic of cultivated *sh4* alleles is not sufficient for reduction of shattering in all genetic backgrounds.

Unequivocal determination of the ancestry of weedy rice from *sh4* sequence data is complicated by detection of the common cultivated *sh4* haplotype at low frequencies in wild rice accessions (six out of 30 *O. rufipogon*). Three other surveys of *sh4* diversity,

which have included *O. rufipogon* samples complementary to our own (> 50), have not detected the common cultivated *sh4* haplotype in any *O. rufipogon* (Li *et al.*, 2006; Zhang *et al.*, 2009; Lin *et al.*, 2007), which supports our conclusions regarding the rarity of this allele within wild rice. Interestingly, the wild rice accessions possessing the common cultivated *sh4* haplotype share at least 50% genomic identity with cultivated rice (Reagon *et al*., 2010), suggesting they may have acquired these alleles through introgression; however, intermediate crop-wild morphologies have not been observed for these accessions (e.g. height, tillering, hull color, awns, etc.), and an ancestral existence of these alleles in wild rice cannot be completely ruled out.

We consider weedy inheritance of *sh4* alleles from wild ancestors unlikely for several reasons: 1) inheritance of the common cultivated *sh4* haplotype in the independently evolved SH, BHA1, and BHA2 weedy rice groups is more likely to have occurred from a group where the haplotype is nearly fixed (cultivated rice), than from one where it is rare; 2) for loci sampled across a 15.2 Mb genomic region surrounding *sh4*, clades containing SH weeds tend to contain at least one *indica* cultivar and clades containing BHA weeds tend to contain at least one *aus*, as expected from their genomic-inferred ancestry; 3) three distinct extended haplotypes across a 6.2 Mb genomic region containing *sh4* are shared among cultivated and weedy rice accessions, whereas a single extended haplotype is shared with wild rice (Figure 2.2; Figure 2.3).

Our identification of the main "cultivated" *sh4* haplotype in all US weedy rice groups constitutes the strongest evidence to date of an origin of these weeds from domesticated ancestors. If weeds inherited their *sh4* alleles from domesticated rice, two mechanisms could account for the novel SNPs carried by some weedy accessions at *sh4*

and other sampled loci. The SNPs could have accumulated through mutation since divergence from cultivated ancestors, possibly aided by release from selection for non-shattering at *sh4*. Novel SNPs could also have been acquired through introgression with un-sampled wild and/ or cultivated individuals.

The single origin of the *sh4* allele in cultivated rice is striking because a preponderance of evidence supports a minimum of two rice domestication events in different areas of Asia, one giving rise to the *indica* and *aus*, and another to the *japonicas* and *aromatic* group (see Sweeney & McCouch, 2007). Several models have been proposed to account for this discrepancy (Lin *et al*., 2007; Sang & Ge, 2007a; Sang & Ge, 2007b). Recent evidence suggests that the *sh4* T mutation was first fixed in one set of cultivars, and quickly spread to independently domesticated rice groups via gene flow and selection (Zhang, 2009). The cultivated rice group in which the T substitution was initially fixed has not been identified, though some studies have suggested an origin in rice outside of China (Zhang *et al*, 2009). Haplotypes favored by positive directional selection are expected to manifest an extended block of LD around the favored mutation, and our survey of polymorphism in the *sh4* genomic region is consistent with strong selection on *sh4* in all cultivated rice groups prior to the evolution of weedy rice. Patterns of extended homozygosity in the region are also consistent with an origin of the *sh4* T mutation in ancestors of the *tropical japonica* group, with subsequent introgression into ancestors of *indica* (Figure 2.4). Finer scale characterization of the *sh4* genomic region may be needed to rule out the effects of sampling stochasticity on the observed patterns.

The presence of “non-shattering” *sh4* alleles in US weedy rice despite their propensity to shatter (Figure 2.1; Figure 2.5), implies that weedy groups must have re-

acquired the shattering trait through the involvement of other, unidentified loci. These could be major loci that have not yet been identified within *Oryza*, or numerous loci with small effect that are thus difficult to detect. The ability to shatter despite having the T substitution is also present in some *O. rufipogon* and one *aus* cultivar. Alleles at genes facilitating shattering may have been acquired by weedy groups through *de novo* mutation, introgression from wild rice, or perhaps inherited from the few domesticated backgrounds that are able to produce BTS values at the lower end of the scale (Figure 2.1). Whether divergent weedy rice groups have acquired the shattering traits through similar genetic mechanisms remains an open question. Ongoing fine scale characterization of the shattering trait via microscopy and BTS time-course evaluations across *Oryza* groups may help determine the likelihood of a shared genetic basis for shattering between wild and weedy rice. Ultimate identification of loci contributing to shattering in weedy rice may be facilitated by numerous QTL studies of this trait (Gu, 2005a; Onishi, 2007; Cai, 2000; Ji, 2006; Xiong, 1999; Thomson, 2003), including some involving crosses between Asian weeds and cultivated rice (Gu, 2005a; Bres-Patry, 2001). To shed further light on the genetic basis of shattering in US weeds, we are currently generating mapping populations from US weedy rice parents and their putative progenitors.

Assessments of genomic patterns of polymorphism have supported origin of US weedy rice groups from two domesticated rice varieties, *indica* and *aus* (Londo & Schaal, 2007; Reagon *et al*., 2010). In contrast, assessments of polymorphism at a candidate locus for pericarp color, *Rc*, have revealed that alleles in weedy groups, which are exclusively red-pigmented, are not derived from alleles carried by the more common

white pericarp cultivars (Gross *et al*., 2010). However, red pericarp cultivated rice varieties exist, implying that selection on *Rc* is likely to have been a feature during the development of modern cultivated varieties, rather than the early stages of rice domestication; thus, polymorphism at *Rc* suggests that US weedy rice groups arose prior to the emergence of white-pericarp cultivated rice, perhaps from primitive red-pericarp domesticates (Gross *et al*., 2010). The *sh4* polymorphism data reported here further refines our understanding of the origin of US weedy rice. All weed groups must have originated after the fixation of the non-shattering *sh4* allele in all cultivated rice groups. Thus, the progenitors of weedy rice must have been "domesticated enough" to have undergone selection for reduced shattering. Future investigation of additional candidate domestication and weedy loci are likely to further contribute to our understanding of the evolutionary origins of this noxious weed.

Table 2.1: Silent site nucleotide diversity per kb. Watterson's estimator nucleotide variation (θ_W), the average pairwise nucleotide diversity (θ_π) and Tajima's D for wild *O. rufipogon*, cultivated *O. sativa* and weedy *O. sativa*.

			Cultivated *Oryza sativa*						US Weedy Rice	
		Oryza rufipogon	*indica*	*aus*	*trp jap*	*tmp jap*	*arom*	SH	BHA 1	BHA 2
***sh4* locus**	θ_π	4	0	0.094	0.03	0	0.2	0	0.04	0.2
	θ_W	5	0	0.12	0.92	0	0.2	0	0.99	0.1
	Tajima's D	-0.85	N/A	-1.01	-1.16	N/A	0	N/A	-1.16	1.44
Flanking Fragments										
sh4f_001	θ_π	2.2	2.4	0.79	0.26	0	0	0	0	0
	θ_W	2.4	2.1	1.1	0.77	0	0	0	0	0
	Tajima's D	-0.19	0.41	1.18	-1.16	N/A	N/A	N/A	N/A	N/A
sh4f_002	θ_π	7.3	0	1.7	0	0	1.5	0	2.3	2.6
	θ_W	1.2	0	1.8	0	0	1.5	0	1.4	1.7
	Tajima's D	-1.26	N/A	-1	N/A	N/A	0	N/A	1.8	1.79
sh4f_003	θ_π	2.2	0	0	0	0	0	0	0	0
	θ_W	4.4	0	0	0	0	0	0	0	0
	Tajima's D	-1.2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
sh4f_004	θ_π	1.8	0	0	0	4.1	0	0	0	0
	θ_W	3	0	0	0	3.3	0	0	0	0
	Tajima's D	-0.89	N/A	N/A	N/A	1.46	N/A	N/A	N/A	N/A
sh4f_005	θ_π	2.4	2.1	1.8	2	1.6	2.6	1.9	2	0
	θ_W	2.5	1.5	1.6	1.1	1.9	2.6	1.1	1.2	0
	Tajima's D	-0.13	1.17	0.56	1.57	-0.82	0	1.43	1.47	N/A
sh4f_006	θ_π	1.6	2	0	0.48	2.1	0	0	0	0
	θ_W	3.5	3	0	0.73	2.6	0	0	0	0
	Tajima's D	-1.34	-1.45	N/A	-0.62	-0.97	N/A	N/A	N/A	N/A

Table 2.2. List of accessions used for this study. Accessions are grouped by type (weed, wild or cultivar). Identification numbers as well as genotypes at *qsh1* and *sh4* are listed along with phenotypic values for seed shattering.

Group	Study ID a	USDA ID/ Common Name c	IRGC/RA/GRIN d	Origin b	Mean BTS (grams) e	Std Dev	*sh4*	*qsh1*
Weedy rice	SH_1A04	**1091-01**	**x**	AR	0.3	0.5	T	G
	SH_1A07	**1098-01**	**x**	MO	0.2	0.4	T	G
	SH_1A08	**1134-01**	**x**	AR	0	0	T	G
	SH_1A09	**1135-01**	**x**	AR	0.3	0.5	T	G
	SH_1A10	**1141-01**	**x**	AR	1.2	3.1	T	G
	SH_1A11	**1160-01**	**x**	AR	0	0	T	G
	SH_1A12	**1179-01**	**x**	LA	1.7	2.8	T	G
	SH_1B05	1995-15	**x**	AR	0	0	T	G
	SH_1B03	16B	**x**	AR	0	0	T	G
	SH_1B07	1996-05	**x**	MS	0.4	0.2	T	G
	SH_1C02	**1001-01**	**x**	AR	1	2	T	G
	SH_1C03	**1002-02**	**x**	AR	2.3	2.7	T	G
	SH_1C06	**1047-01**	**x**	LA	0.1	0.4	T	G
	SH_1C07	**1073-02**	**x**	MO	0.3	0.4	T	G
	SH_1C10	**1190-01**	**x**	LA	0.6	0.9	T	G
	SH_1C11	**1199-01**	**x**	MO	1	2.8	T	G
	SH_1D01	**1344-02**	**x**	MO	0.7	1.6	T	G
	SH_1D06	1995-12	**x**	LA	0	0	T	G
	SH_1D09	1996-08	**x**	MS	1.5	2	T	G
	SH_1E03	**1210-02**	**x**	MO	2.3	5	T	G
	SH_1E07	**1333-02**	**x**	MO	0.2	0.4	T	G
	SH_1A01	**1004-01**	**x**	MO	1.5	1.8	T	G
	SH_1E05	**1163-01**	**x**	LA	0.1	0.4	T	G
	SH_1A06	**1196-01**	**x**	AR	1.2	2.9	T	G
	MXSH_1B06	1996-01	**x**	AR	35.6	17.9	T	G
	MXSH_1D10	2002-51	**x**	AR	0	0	T	G
	MXBH_1E10	**2002-2-p21**	**x**	AR	0	0	T	G
	MXBH_1D11	2004-1-A	**x**	AR	0	0	T	G

Table 2.2, cont.

MX_1B10	MS4R788_93	**x**	MS	0	0	T	G
BRH_1C12	**1300-02**	**x**	MO	0	0	T	G
BRH_1D12	**1183-01**	**x**	AR	0.6	2.6	T	G
BRH_1C09	**1111-01**	**x**	AR	0	0	T	G
BRH_1C08	**1092-02**	**x**	MS	0	0	T	G
BRH_1E06	**1120-02**	**x**	MO	0.1	0.7	T	G
BHA2_1E04	**1214-02**	**x**	LA	0	0	T	G
BHA2_1B01	**1188-01**	**x**	MS	0	0	T	G
BHA2_1C01	TX4	**x**	TX	0	0	T	G
BHA2_1A02	**1025-01**	**x**	AR	0	0	T	G
BHA2_1A03	**1081-01**	**x**	AR	0	0	T	G
BHA2_1E02	**1107-01**	**x**	AR	0	0	T	G
BHA2_1C05	**1042-01**	**x**	AR	0	0	T	G
BHA2_1E08	**1202-02**	**x**	AR	0	0	T	G
BHA1_1B11	StgB	**x**	AR	0	0	T	G
BHA1_1D07	1995-13	**x**	LA	0	0	T	G
BHA1_1B02	10A	**x**	AR	0	0	T	G
BHA1_1B08	1996-09	**x**	MS	7.2	21.6	T	G
BHA1_1E09	**2002-2-p1**	**x**	AR	0	0	T	G
BHA1_1D08	1995-14	**x**	LA	0	0	T	G
BHA1_1B09	LA3	**x**	LA	0	0	T	G
BHA1_1B12	StgS	**x**	AR	4.6	5	T	G
BHA1_1D05	PrCoTall_3	**x**	AR	0	0	T	G
BHA1_1E01	**1166-02**	**x**	MS	0	0	T	G
BHA1_1A05	**1096-01**	**x**	AR	0	0	T	G
BHA1_1B04	18A	**x**	AR	0	0	T	G
BHA1_1C04	**1005-02**	**x**	AR	0	0	T	G
BHA1_1D02	PrCoSrt_1	**x**	AR	0	0	T	G
BHA1_1D03	PrCoTall_1	**x**	AR	0	0	T	G
BHA1_1D04	PrCoTall_2	**x**	AR	0	0	T	G

Table 2.2, cont.

Cultivated rice								
aromatic	3A01	DOM-SOFID	RA4929/12880/PI584607	Iran	x	x	T	G
	3A02	ARC-13829	RA48947/42469	India (N. Kakhimpur)	x	x	T	G
	2B01	DA 13	x	Bangladesh	3	1	T	G
aus	2B03	Aus 196	29016	Bangladesh	12.3	9.8	T	G
	3A03	JHONA-349	RA4979/6307	India	43.1	25	T	G
	3A04	KASALATH	RA5339	India	12.8	11.6	T	G
	3A05	DV85	RA5323/8839	Bangladesh	0.3	0.5	T	G
	3A06	BJ-1	RA5345/45195	India	18.3	3.1	T	G
	3A07	DHALA_SHAITTA	RA5361/PI180060	Bangladesh	x	x	T	G
indica	2B02	Bei Khe	22739	Cambodia	30.1	17.5	T	G
	3A08	RATHUWEE	RA4911/8952/PI584605	Sri Lanka	72.3	47.8	T	G
	3A09	Khao Dawk Mali -105	RA4878/27748	Thailand	80.7	42.6	T	G
	3A10	LalAman	RA4956/46202	India	110.6	16.5	T	G
	3A11	Dholi Boro	RA4984/27513	Bangladesh	137.4	11.8	T	G
	3A12	Ai-chiao-hong	RA4967/51250/PI584576	China	4	3.3	T	G
	3B01	CHAU	RA4974/56036	Vietnam	x	x	T	G
	3B02	CHHOTE_DHAN	RA4978/58930	Nepal	17	14.4	T	G
	3C05	DEE_GEO_WOO_GEN	RA5344/PI279131	Taiwan	60.9	25.3	T	G
	3B03	POPOT-165	RA4987/43545	Indonesia (E. Kalimantan)	7.5	4.5	T	G
temperate japonica	3B04	Ta hung ku	RA5029/1107	China	37	29.5	T	T
	3B05	Kamenoo	RA4990/PI403629	Japan	x	x	T	G
	3B06	Nep-Hoa-Vang	RA4945/40748	Vietnam	x	x	T	T
	3B07	SHOEMED	RA5123/PI392539	Philippines	x	x	T	G
tropical japonica	3B08	Khao Hawm	RA5030/24225	Thailand	46.8	8.5	T	G
	3B09	Mirti	RA4970/25901/PI584553	Bangladesh	12	22.9	T	G
	3B10	KU115	RA5294/PI597044	Thailand	8.7	8.2	T	G
	3B11	CICIH_BETON	RA4955/43372	Indonesia (Bali)	85.4	5.37	T	G
	3B12	GOTAK_GATIK	RA4959/43397/PI584572	Indonesia (C. Java)	104.5	67.7	T	G
	3C01	ASSE_Y_PUNG	RA5333/CIor461	Philippines	115.9	41.3	T	G

Table 2.2, cont.

	3C02	Kotobuki Mochi	RA4882/2545	Japan	35.7	22	T	G
	3C03	TREMBESE	RA4988/43675	Indonesia (East Java)	59	53.5	T	G
US cultivars	3C04	LEMONT	RA4998/66756	US	112	35.5	T	G
	2A01	BENGAL	PI561735	US	x	x	T	G
	2A02	BLUE_ROSE	CIor1962	US	204	16.6	T	G
	2A03	CAROLINA_GOLD	CIor1645	US	208	8.2	T	G
	2A04	CL121	n/a	US	x	x	T	G
	2A05	CL161	n/a	US	x	x	T	G
	2A06	CYPRESS	PI561734	US	x	x	T	G
	2A07	DELITUS	CIor1206	US	x	x	T	G
	2A08	DREW	n/a	US	x	x	T	G
	2A09	EDITH	CIor2127	US	207	14	T	G
	2A10	PALMYRA	CIor9463	US	x	x	T	G
	2A11	REXORO	CIor1779	US	x	x	T	G
	2A12	ZENITH	CIor7787	US	x	x	T	G
Wild Asian rice and outgroups								
O. rufipogon	3C06	x	106163	Laos	x	x	T	G
	2C01	x	81990	Myanmar	x	x	X	G
	2C02	x	100588	Taiwan	0	0	T	G
	2C03	x	100904	Thailand	x	x	G	G
	2C04	x	100916	China	77.2	62	G	G
	2C05	x	104501	India	x	x	G	G
	2C06	x	104599	Sri Lanka	x	x	G/T	G
	2C07	x	104624	China	0	0	G	G
	2C08	x	104714	Thailand	x	x	T	G
	2C09	x	104833	Thailand	0	0	T	G
	2C10	x	104871	Thailand	0	0	G	G
	2C11	x	105388	Thailand	x	x	G	G
	2C12	x	105491	Malasia	0	0	X	G
	2D01	x	105568	Philippines	0	0	G	G
	2D02	x	105711	India	0	0	T	G
	2D03	x	105720	Cambodia	x	x	G	G

Table 2.2, cont.

	2D04	x	105855	Thailand	x	x	X	G
	2D05	x	105888	Bangladesh	0	0	G	G
	2D06	x	106086	India	0	0	T	G
	2D07	x	106103	India	x	x	G	G
	2D08	x	106122	India	x	x	G	G
	2D09	x	106134	India	x	x	G	G
	2D10	x	106150	Laos	x	x	G	G
	2D11	x	106168	Vietnam	x	x	G	G
	2D12	x	106169	Vietnam	x	x	T	G
	2E01	x	106321	Cambodia	0	0	T	G
	2E02	x	106346	Myanmar	x	x	G	G
	2E03	x	106453	Indonesia	x	x	G	G
	2E04	x	106518	Vietnam	x	x	G	G
	2E05	x	106523	Papua New Guinea	x	x	G	G
O. nivara	2F01	x	86662	Thailand	0	0	T	G
	2F02	x	103821	China	0	0	T	G
O .barthii	2F03	x	101226	Mali	x	x	G	G
	2F04	x	104081	Nigeria	x	x	X	G
O. glaberrima	2F05	x	86779	Liberia	x	x	G	G
	2F06	x	100983	Nigeria	x	x	X	G
	2F07	x	101855	Burkina Faso	26	11.9	G	G
	2F08	x	102410	Mali	50	29.3	G	G
O. glumaepatula	2F09	x	105561	Colombia	0	0	G	G
	2F10	x	105670	Brazil	0	0	G	G
O. meridionalis	2F11	x	93261	Indonesia	0	0	G	G
	3C07	x	101148	Australia	0	0	G	G

a based on STRUCTURE and identity from Reagon et al 2010

b Origin for weeds is a US state abbreviation, Origins for cultivated and wild rice is country

c accessions with RA numbers were acquired from Susan McCouch while all others were acquired from IRRI

d Accessions in bold were selfed 4 generations at the USDA stock center

e BTS stands for Breaking Tensile Strength and is the maximum weight a seed can hold before releasing

x-- no data available

Table 2.3. List of primer sequences and their location. Primers are grouped by gene (*sh4* and *qsh1*) as well as genetic versus flanking region. Additionally, the enzyme used in the *qsh1* CAPS study is identified and the cut site is listed

Primer ID	**Forward**	**Reverse**	**Region**
sh4_001	AAGCTCCCAAGTGTCAAAGC	TGTTACCCTATGCGTCTACTCC	5' UTR
sh4_002	TGCATATATACATACTCCCTCTCG	GCTGACTCTTGTTTGTTTCG	5' UTR
sh4_003	AAAGCCTGTGATTAAGGTAAAGCAG	TGATGAGGATGAGCGTCTCG	5'UTR/Exon 1
sh4_004	ACGGGCACCTGACTGCTACG	GAGGTGGGTGGTGGTGATGG	Exon 1
sh4_005	AGGTCTACGACGCGCTCTCC	AAATATCCGGATTCGCATCACC	Exon 1/Intron
sh4_006	TTTACCTTTTTACCGTGGGGTTCG	ACTCCGCCCCAAAACAATACAG	Intron
sh4_007	GTAGATATCCTTTGACCGTTCTATC	GCCATTCCAAACAATAATCC	Intron/Exon 2
sh4_008	GCAATCATATGCAATGCAGCAAGG	TCTGATCTGATTCGGCCAACACC	3' UTR
sh4_009	CTCGTGCTGAGCTATACAAGC	GTACAATGCACGTCAATTCC	3' UTR
sh4_010	TATATAGGGACGGTTCCACCAAGC	TCTTCTTCCCGGTCTTCTTTTCAG	5' UTR
sh4_011	ACCCGAACAGAGTTTTGATGAGC	ATTGGCGGTGATGAGGATGAG	5' UTR/Exon 1
sh4_012	ATCGCTCCCCGAACACCAAAC	GGGAGAGCACCTCGGAGAGC	Exon 1
sh4_013	AGGAGGAGCGCACCGAGGTC	TCTGATCTGATTCGGCCAACACC	3' UTR
sh4_014	ACTACCGCAAGGGGAACTGGAC	TAGCTGTTACATCCCCTCCCCTC	Exon 1

***sh4* Flanking Region Primers**

Primer ID	**Forward**	**Reverse**	**LOC ID (Gramene)**	**Distance from *sh4* (Mb)**
sh43_001	CCATGGTCTGATGGTGAACG	CAGGAAGTACCCACATGCTTTAGG	Os04g57550	0.008
sh43_004	GAGGAAGAATGCCGGTGTCC	CCAGGCCATTTTCAGAGAAGC	Os04g58580	0.6
sh43_007	CCCTTGGAGTGCATCATCAGG	AGGGCATATGCAACGGATGG	Os04g59570	1.2
sh45_002	GATGTGCAAGAAGGCCAAGG	ACTTAGGCAAGAGCGGAAGAAGG	Os04g53530	2.4
sh45_008	GCAGCAGGAAATGCACATCAGG	AAAGTAGTTCACGAAGAGCCGAGTCC	Os04g55690	1.1
sh45_012	CGTTTTCTTCAGGGTGTTCAGG	AGCACATACAGGACCACCATACG	Os04g56920	0.3
sts_040	GCAGTGGATTTTCCAGCTCTCC	CACGACTACATCAGGGTTGAATGG	Os04g44560	5
sts_021	GTAGCCAAGATTGGGCTGTGG	GATACCAAAAGCGTCCACGTACC	Os04g30040	14

***qsh1* CAPS Primers**				
			Enzyme:	BspH1
qsh1	GGCTTTTTAGTTCGTGCATGTAGC	GGTTACACAAGTTCCCCGCAC	**Cut Site:**	TCATGA

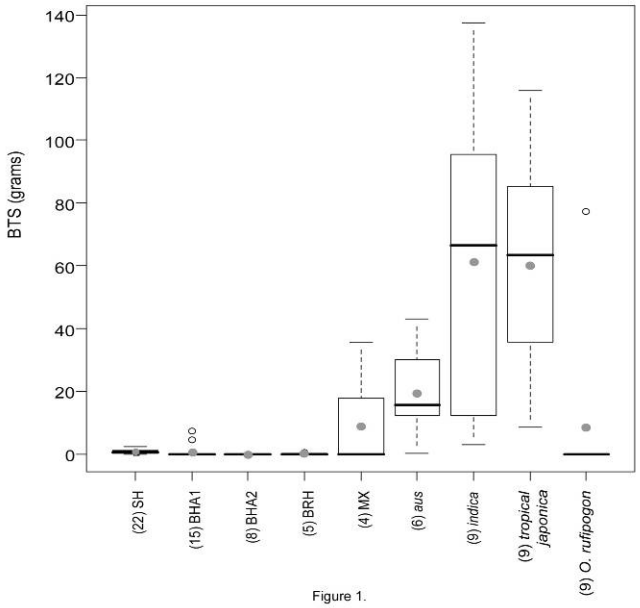


Figure 2.1. Seed shattering phenotype in weedy, wild and cultivated *Oryza*.
Distributions are of average accession BTS values for each *Oryza* group. The black line represents the median of each distribution, and the grey dot the mean; white dots represent outliers. Numbers in parenthesis correspond to sample sizes. Weedy rice groups are as follows: SH (straw-hulled), BHA1 and BHA2 (black hulled and awned), BRH (brown hulled) and MIX (mixed origin). Both *O. rufipogon* and *O. nivara* accessions have been grouped together under the heading *O. rufipogon*.

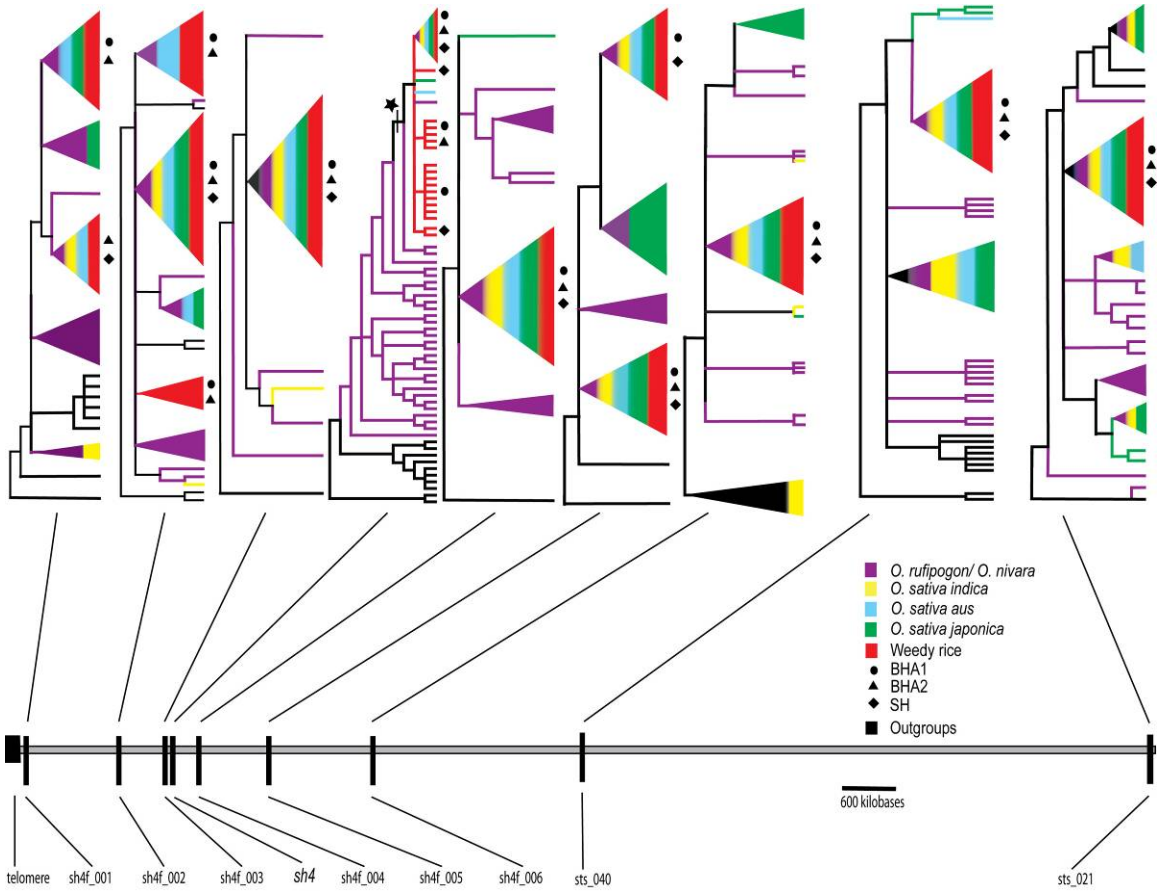


Figure 2.2. Phylogenies of flanking regions surrounding *sh4*.
Neighbor Joining trees for each of eight ~500 bp regions at varying distances from the *sh4* locus. Diagram is to scale. Only branches with bootstrap values over 50% are shown. The star on the *sh4* locus tree denotes the T substitution associated with loss of shattering. For clarity, all *tropical japonica*, *temperate japonica* and *aromatic* rice have been grouped under the *japonica* heading and colored green. Additionally, all weed groups have been colored red, but the main groups are distinguishable via icons placed to the right of each tree.

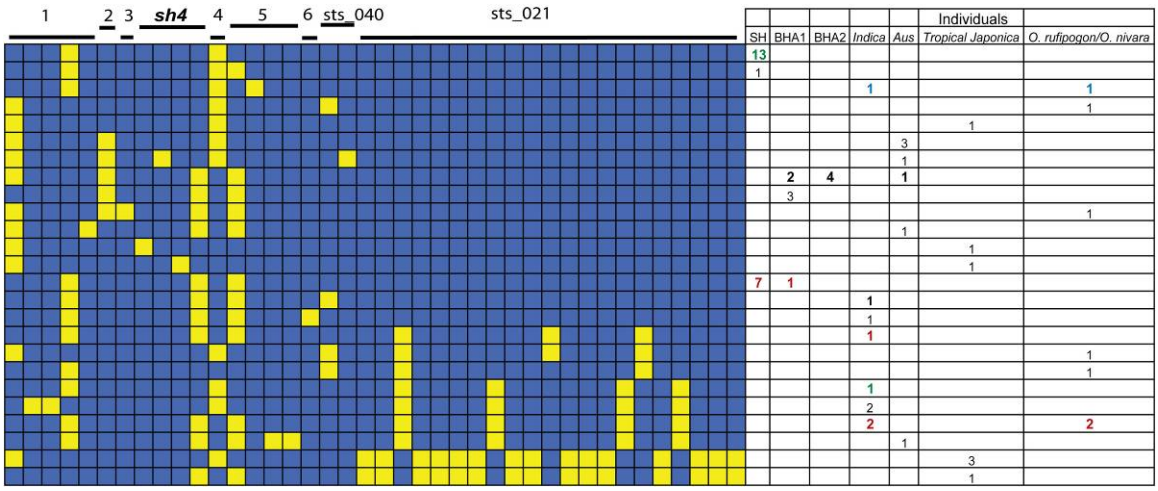


Figure 2.3. Graphical view of concatenated *sh4* haplotypes.
Haplotypes across the genomic region surrounding *sh4* are shown for the 90 individuals (wild, weedy, and cultivated) that share the common *sh4* haplotype containing the T SNP. The numbers across the top represent flanking regions (1- 6 = sh4f_001- _006). Yellow squares represent SNPs found in at least one haplotype. A tally of individuals from each cultivated, weedy, or wild group is shown to the right. Colors of accession counts indicate haplotypes that are identical across a 6.2 Mb region (up to sts_040) containing *sh4*.

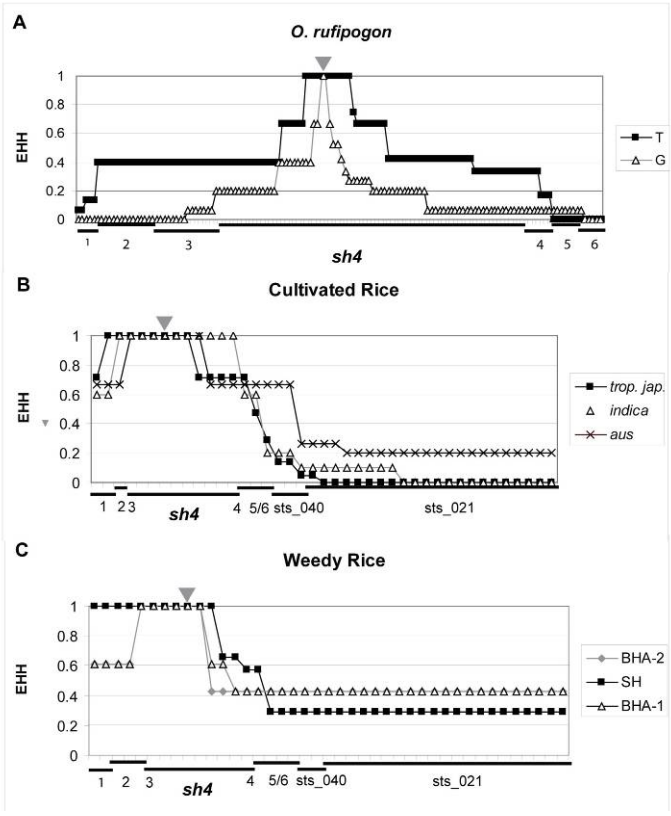


Figure 2.4. Extended haplotype homozygosity surrounding *sh4*.
EHH was performed on concatenated alignments containing the *sh4* gene and all eight flanking regions in order as they appear on the chromosome. Sts_040 and sts_021 were not included for *O. rufipogon* as haplotype homozygosity had already reached zero. The grey triangle atop each panel represents the location of the T mutation associated with loss of shattering in *sh4*. Numbers under black bars represent flanking regions (1- 6 = sh4f_001- _006). **A.** EHH for *O. rufipogon* groups possessing a T or a G at the SNP associated with shattering variation. **B.** EHH results for three cultivated rice groups. **C**. EHH results for the main US weedy rice groups.

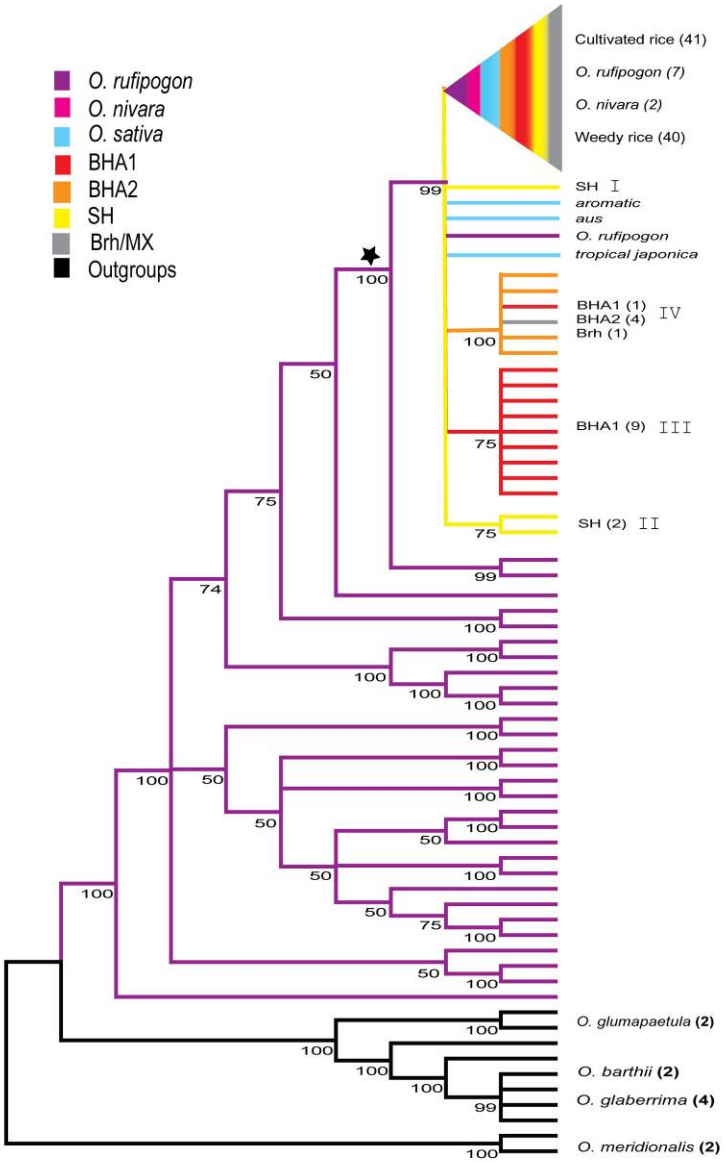


Figure 2.5. Neighbor joining tree of *sh4* haplotypes.
Numbers below branches represent bootstrap support in percentages; only clades with over 50% support are labeled. The black star denotes the G to T substitution associated with loss of shattering in domesticated rice. Color key at left of the tree identifies *Oryza* groups represented by the observed haplotypes. The *O. sativa* group contains accessions of the five recognized domesticated rice populations: *aus*, *indica*, *aromatic*, *tropical japonica* and *temperate japonica*. Labels on the right side the tree identify the number of individuals sharing a haplotype. A triangle is placed anywhere more than ten individuals share an identical haplotype. Four haplotypes unique to weedy rice are numbered (I, II, III, and IV) while haplotypes unique to *O. rufipogon* are not labeled or numbered. Three of the unique weedy rice haplotypes contain mutations that alter amino acids: Glutamic Acid to Lysine in exon 1 in haplotype II, Arginine to Leucine in exon 2 in haplotype III, and Arginine to Tryptophan in exon 2 in haplotype I.

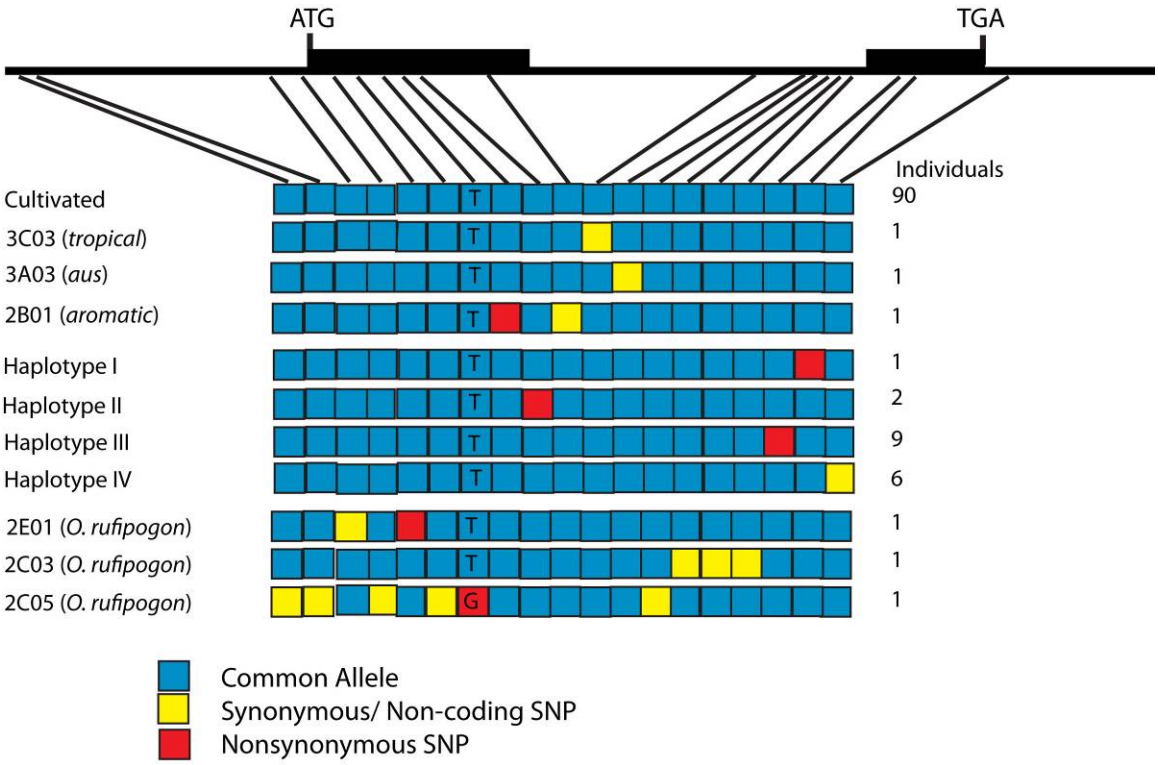


Figure 2.6. Graphical view of unique *sh4* haplotypes.
The top haplotype represents the common shared cultivated haplotype found in 90 individuals from cultivated, weedy, and wild groups. Of the three unique cultivar haplotypes, only the *aromatic* individual (2B01) contains a non-synonymous SNP. Four unique weedy haplotypes (I - IV) are displayed where three of the four contain nonsynonymous SNPs. Haplotype numbers match those of Figure 2. Additionally, three wild individuals are shown. 2E01 and 2C03 contain the non-shattering T nucleotide plus additional coding and non-coding SNPs. 2C05 was chosen to represent one of the many wild haplotypes containing a shattering G nucleotide for comparison.

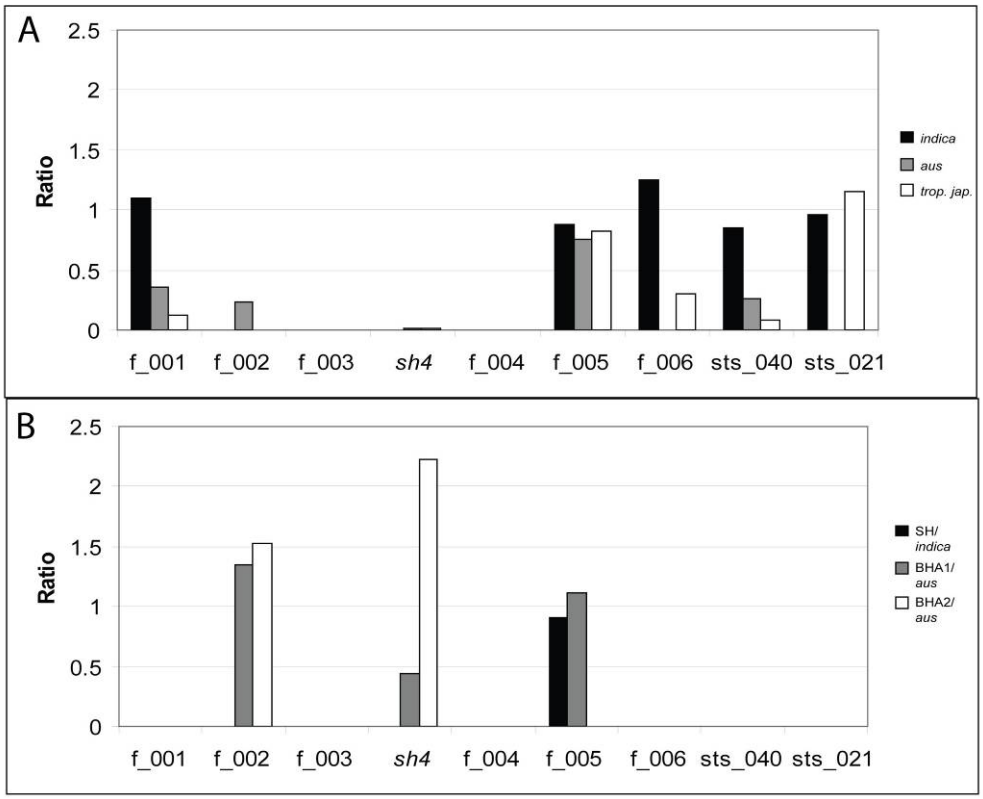


Figure 2.7. Ratios of silent site nucleotide diversity at *sh4* and surrounding loci. The ratio of the average pairwise nucleotide diversity (θ_π) per kb is shown **A)** for three cultivated groups (*indica*, *aus* and *tropical japonica*) against wild *O. rufipogon*, and **B)** for the three major weed groups (SH, BHA1 and BHA2) against their putative progenitors (*indica* or a*us*). Overall diversity is low across the entire region in cultivated and weedy rice groups.

CHAPTER 3

TIMING IS EVERYTHING: EARLY DEGRADATION OF ABSCISSION LAYER IS ASSOCIATED WITH INCREASED SEED SHEDDING IN US WEEDY RICE

3.1 Background

Abscission is the process by which plants shed unwanted organs, such as those that have been damaged or diseased, or release ripe seeds and fruits (Patterson, 2001). Seed abscission is an important mechanism for seed dispersal in many wild cereals (Harlan & DeWet, 1965). During domestication of grass species (e.g. wheat, rye, barley, and rice), a critical shift occurred towards reductions in seed-shedding ability, facilitating the harvesting of grains (Harlan & DeWet, 1965; Fuller *et al*., 2009; Sang, 2009; Zhang & Mergoum, 2007). Seed shattering is costly to farmers, as crop yield is diminished, and lost seeds may lead to persistence of crop volunteers in cultivated fields (Zhang & Mergoum, 2007; Roberts *et al*., 2000). However, seeds that require intense labor to harvest are also undesirable, along with those that remain on the plant and germinate (i.e. preharvest sprouting). A balance between ease of shattering and difficult threshing is maintained in crop species, determined in part by specific demands of the harvesting system (e.g. hand vs. machine threshing) (Sang & Ge, 2007a; Li *et al*., 2006b). In contrast, in agricultural weeds — plants that invade cultivated fields — increased seed dispersal is believed to be favored, much as it is in wild species (Harlan & DeWet, 1965). Seed shattering is a commonly observed trait in agricultural weedy plants that are related to domesticated species (Harlan & DeWet, 1965). Seed shattering is thus under opposing selection in crops and weeds inhabiting agricultural complexes.

Domesticated Asian rice (*Oryza sativa* L.) is one of the world's most important

crop species, providing about 20% of the world's caloric intake (FAOSTAT, 2008). Cultivated rice fields worldwide are invaded by a weedy relative of rice known as weedy or red rice (*O. sativa*) (Burgos *et al*., 2008). Weedy rice is costly to farmers in terms of yield losses and removal efforts, as it competes aggressively with cultivated rice and can contaminate harvests (Burgos *et al*., 2008; Cao *et al*., 2006). The ability of weedy rice to survive and spread in cultivated rice fields has been attributed in part to its reported capacity to shatter seeds (e.g. Delouche *et al*., 2007; Gealy *et al*., 2003; Noldin *et al*., 1999; Oard *et al*., 2000). High levels of seed shattering are also prevalent in the wild ancestor of cultivated rice, *O. rufipogon*, which is native to tropical wetlands of South Asia (Oka, 1998). Cultivated Asian rice, in contrast, shows a wide range of seed threshability levels, from nearly shattering to difficult to thresh, but is generally less shattering than wild and weedy species (Ji *et al*., 2006; Thurber *et al*., 2010).

Organ abscission in plants depends on the formation of abscission zones, which are morphologically distinct structures generally consisting of one to multiple layers of cells dense with cytoplasm (Patterson, 2001; Roberts *et al*., 2000). Swelling and dissolving of the middle lamella between adjacent cell walls in the abscission layer allows for organ release (Patterson, 2001; Ayeh *et al*., 2009). In many plants, the abscission layer is formed long before the activation of cell separation and breakage occur (Ayeh *et al*., 2009; Cho *et al*., 2008). Seed shattering in *Oryza* is dependent on the proper formation and subsequent degradation of an abscission layer between the flower and the pedicel. QTL (quantitative trait loci) associated with loss of shattering have been identified on nearly every rice chromosome, and three loci have been cloned to date: *sh4/SHA1*, *qsh1* and *OsCPL1* (Li *et al*., 2006b; Konishi *et al*., 2006; Ji *et al*., 2010). Of

these loci, *sh4*, which encodes a nuclear transcription factor, is considered the most important contributor to reduced shattering during rice domestication (Purugganan & Fuller, 2009). A single nonsynonymous substitution (G to T) in the first exon of *sh4* leads to reduced function of SH4 and incomplete development of the abscission layer in non-shattering cultivated rice (Li *et al*., 2006b). This non-shattering mutation is fixed in all cultivated rice varieties examined to date (Li *et al*., 2006b; Thurber *et al*., 2010; Zhang *et al*., 2009; Lin *et al*., 2007), spanning the highly differentiated *japonica* and *indica* cultivar groups. There is still some controversy whether Asian rice was independently domesticated at least twice from *O. rufipogon* populations (Garris *et al*., 2005; Caicedo *et al*., 2007; Londo *et al*., 2006), or only once (Fuller *et al*., 2009; Fuller *et al*., 2010). Regardless of the domestication scenario, the ubiquity of the T substitution in cultivated rice suggests very strong selection for loss of shattering (perhaps in combination with introgression) during domestication (Li *et al*., 2006b; Zhang *et al*., 2009; Lin *et al*., 2007).

Recently, we examined the seed shattering phenotype and the *sh4* shattering locus in populations of US weedy rice (Thurber *et al*., 2010). Several genetically differentiated populations of weedy rice occur in the US, and these can be distinguished by their predominant hull morphology (Reagon *et al*., 2010). Main populations include the straw-hulled (SH) group, early flowering weeds characterized by straw-colored hulls and lack of awns, and the black-hulled awned (BHA) group, later flowering weeds with seeds that have predominantly black hulls and long awns (Reagon *et al*., 2010; Gealy *et al*., 2002; Londo *et al*., 2007). Genome-wide data indicate that SH and BHA weedy rice groups share genomic identity with Asian domesticated rice from the *indica* and *aus* variety

groups, respectively, suggesting weedy origins within these cultivated groups (Reagon *et al*., 2010; Gealy *et al*., 2002; Londo *et al*., 2007). Minor US weedy rice groups include the brown-hulled (BRH) group, which are putative hybrids between SH and BHA weeds, and the mixed groups (MX), containing individuals likely to be hybrids between weeds and local *tropical japonica* cultivars (Reagon *et al*., 2010). We have found that nearly all US weedy rice readily shatters its seeds to a similar degree as wild rice (Thurber *et al*., 2010). However, all populations of US weedy rice share the "non-shattering" *sh4* substitution common to cultivated rice, regardless of their propensity to shatter (Thurber *et al*., 2010). These results support the evolution of US weedy rice from cultivated ancestors and, since wild and major weedy groups have separate origins, the parallel evolution of the shattering trait among these *Oryza* groups. Our results further imply that weedy rice re-acquired the shattering trait through the involvement of unidentified loci other than *sh4* (Thurber *et al*., 2010).

In an effort to understand how weedy rice may have re-evolved the shattering trait after its loss in domesticated ancestors, we investigate here the morphological basis of shattering in US weedy rice groups. Given that wild and weedy rice do not share the ancestral *sh4* shattering substitution characteristic of *O. rufipogon*, it is possible that wild and weedy groups do not share the same morphological shattering mechanism. Moreover, despite sharing the same "non-shattering" mutation at the *sh4* locus (Thurber *et al*., 2010), the two major US weedy rice populations — SH and BHA — have separate origins, and may have acquired the shattering phenotype in mechanistically different ways, representing a separate instance of parallel evolution. To our knowledge, no study to date has investigated the morphological basis of the shattering trait in weedy rice. We

examine the abscission layer at the flower-pedicel junction in weedy rice prior to, at and shortly after flowering to determine morphology and level of degradation of this layer in relation to seed shattering ability, and compare these results to those of wild and cultivated *Oryza*, to gain insight into how traits important to weed fitness can evolve.

3.2 Results and Discussion

3.2.1 Abscission Layer Formation Differs in Wild and Cultivated *Oryza*.

We observed the abscission layer at the flower-pedicel junction at flowering in six wild *Oryza* (Table 3.1, denoted with asterisk): four *O. rufipogon*, the wild ancestor of cultivated Asian rice, and two *O. nivara*, an annual ecotype of *O. rufipogon* (Zheng *et al*., 2010). All six wild *Oryza* show clear abscission layers between the flower and the pedicel at flowering (Figure 3.1A-F, and data not shown). The layer is slightly curved and occurs on both sides of the vascular bundle. Further magnification (60x) of the abscission layer shows very dark staining of cells at the center of the layer with some cells beginning to swell. This dark staining is most likely due to high lignification of these cells' walls, as abscission layer cells have been shown previously to be highly lignified (Tabuchi *et al*., 2001). Cells surrounding the layer are highly organized into rows and perpendicular to the plane of abscission. (Figure 3.1B, D, F). No degradation of the abscission layer is yet observed at this stage. The occurrence of well-developed abscission layers upon flowering suggests that all six wild *Oryza* accessions will shatter their seeds readily, an observation that is consistent with our previous measurement of shattering levels of ripe seeds in these accessions (average Breaking Tensile Strength (BTS) = 0 g, Table 3.1; also see (Thurber *et al*., 2010)).

We also observed the flower-pedicel junction at flowering in four cultivated rice samples (Figure 3.1 G-L and data not shown) belonging to the *aus* and *indica* cultivar groups, the putative ancestors of US weedy rice. None of the spikelets (i.e. rice flowers with attached glumes) sampled shows formation of a clear abscission layer upon flowering, although two *indica* accessions (3A09 and 3A11; Figure 3.1G, H, K, L) show weak staining in the region of the abscission layer. In these accessions, further magnification shows diffuse staining of cells in the abscission zone, although cellular organization is not as defined as in the wild tissue samples at this stage (Figure 3.1H, J, L). This further supports the absence of an abscission layer, and, in all cultivated samples, the pedicel blends in easily with the floral tissue at flowering. The lack of an abscission layer at flowering in all three *indica* cultivated accessions is consistent with their lack of shattering (average BTS= 70 to 137 g, Table 3.1). The single *aus* sampled is considered a very easy seed releasing variety (average BTS= 18 g, Table 3.1), yet it also appears to not possess an abscission layer at flowering (Figure 3.1G, H), suggesting that formation of this layer may be delayed and incomplete.

Our overall observations of clear abscission layers upon flowering in shattering wild *Oryza* individuals and lack of abscission layers at this stage in non-shattering cultivated rice are consistent with previous studies (see Li *et al*., 2006b; Ji *et al*., 2006; Konishi *et al*., 2006; Lin *et al*., 2007), and serve as a baseline for comparison to weedy rice. Because our observations do not differ from those published previously for other cultivated and wild rice samples, we concluded that abscission layer traits are robust under our growth conditions, and we did not sample additional time points of abscission layer development. Studies have documented that the abscission layer begins to form at

least one week prior to flowering in wild *O. rufipogon* (and some exceptionally easy threshing *indica* and *aus* cultivars), and by flowering is prominent and clearly visible with staining (Lin *et al*., 2007; Oba *et al*., 1995; Jin, 1986; Jin & Inouye, 1982; Jin & Inouye, 1985). The abscission layer in *O. rufipogon* begins to degrade at or within a week of pollination, about two weeks after flowering, and continues degradation as the seed begins to form and mature, until the seed is released (Jin, 1986; Jin & Inouye, 1982; Jin & Inouye, 1985). In contrast, in cultivated rice varieties, the abscission layer (if present) remains intact for at least 12 days after pollination (Lin *et al*., 2007). Both previous studies and ours show that there are dramatic differences in abscission layer formation and degradation between wild and cultivated rice, likely due to selection against shattering during the domestication process.

3.2.2 Degradation of the Abscission Layer is Accelerated in Weedy Rice.

To determine the role of abscission layer formation and degradation in the shattering phenotype of weedy rice, we sampled six weedy rice accessions from three separate groups (SH (3), BHA (2), MX (1); Table 3.1, denoted with asterisk) at each of three time points: prior to, at and after flowering. With the exception of the non-shattering MX accession (MXSH_1B06, average BTS= 35 g, Table 3.1), all other weedy rice shatter easily, regardless of population identity (average BTS < 8g, Table 3.1). We chose the single MX individual, as it was the only accession found in (Thurber *et al*., 2010) that did not shatter extensively, and was one of the few accessions identified as a putative hybrid between SH weeds and US *tropical japonica* (Reagon *et al*., 2010). We hypothesized that abscission layer formation and degradation in shattering weedy

samples would resemble that observed for *O. rufipogon* and *O. nivara*, while the non-shattering weed individual would resemble cultivated rice.

One week prior to flowering, all five shattering weedy rice accessions, including the two shown in Figure 3.2 (SH_1A08 and BHA_1A05) possess well-defined abscission layers (Figure 3.2A, G). Inspection with a higher magnification 60x lens shows that the BHA and SH weedy rice abscission layers prior to flowering (Figure 3.2B, H) are similar in staining and organization to the wild rice at flowering stage (Figure 3.1B, D, F); the highly lignified cells are darkly stained and starting to swell slightly, while the cells around the region are parallel to the plane of abscission. In contrast, the non-shattering MX weed shows only unbalanced, diffuse staining in the abscission zone with no clear organization of cells surrounding the zone (Figure 3.2M, N).

At flowering, the abscission layers for all the BHA and SH shattering weeds already show mild to moderate degradation and swollen cells at the abscission zone (Figure 3.2C, I; Figure 3.4). Further magnified images show very swollen cells at the abscission layer with the darkest staining seen on the edges that are now exposed due to breakage (Figure 3.2D, J). All five shattering weeds already show degradation that is not observed in their shattering wild relatives at the flowering stage, yet there is some variation in the degree of degradation between weed accessions (Figure 3.1; Figure 3.4). In contrast, the non shattering MX still shows only diffuse, weak staining, yet is beginning to form an abscission layer to one side of the vascular bundle (Figure 3.2O, P). Interestingly, when compared to wild and cultivated spikelets at this developmental stage, MX looks very similar to the non-shattering *indica* cultivars (Figure 3.1G, I, K).

A week after flowering has occurred, which is roughly one to two weeks prior to seed set in weedy rice, all SH and BHA shattering weeds sampled show moderate to near complete separation at the abscission layer and are only held together at the tips of the layer and the vascular bundle (Figure 3.2E, K, and data not shown). Cells that are still attached at the layer are swollen and darkly stained along the plane of breakage. Cells that have already been separated are losing their dark staining, possibly due to rearrangement of cell wall components (Figure 3.2F, L). A week after flowering, the non-shattering MX individual has developed a complete abscission layer, yet the cells at this layer have not begun to swell or degrade (Figure 3.2Q). When examined more closely, the cells of the non-shattering weed look very similar to wild abscission layer cells at flowering and to the shattering weeds prior to flowering: the cells are darkly stained and show a clear abscission layer with organized cells in the abscission zone (Figure 3.2R).

Taken together, our microscopy results demonstrate that shattering weeds display abscission layer developmental differences compared to wild and cultivated rice. Both wild and weedy individuals develop similar looking abscission layers in the same location of the floral-pedicel junction; this similar cellular morphology is consistent with the shared shattering trait of wild and weedy individuals. Moreover, abscission layer formation in shattering weedy rice occurs at least one week prior to flowering, if not earlier, similar to what has been reported for shattering wild rice (Lin *et al*., 2007; Oba *et al*., 1995). However, at flowering, the abscission layer in weedy rice has already begun to degrade, in some cases severely, which is not the case in shattering wild rice or easy threshing varieties of cultivated rice (Ji *et al*., 2006) (Figure 3.1; Figure 3.2; Figure 3.4).

This suggests that timing of abscission layer degradation, rather than morphological differences, distinguishes the shattering trait in weedy and wild rice groups. Surprisingly, despite their independent origins from separate cultivar groups (*aus* and *indica*, respectively), both BHA and SH weeds show similar abscission layer traits and timing. This suggests that both US weedy rice groups may have re-acquired the shattering trait in a similar mechanistic manner, opening the question of whether common genetic elements are involved.

Further investigation of additional developmental stages and a finer scale of developmental series may help identify more precisely when the abscission layer forms in weedy rice and how rapidly after formation it degrades. It is unclear from previous studies how the abscission layer degradation process is activated in rice, yet it is possible that the degradation repertoire is activated only after a certain stage of abscission layer development is complete. While further research is needed, our results indicate that weedy rice may reach this formative stage earlier than wild shattering relatives, and as a result, show earlier degradation. It is also possible that the formation of the abscission layer progresses at the same rate in both weedy and wild rice, with weedy rice abscission activating their degradation repertoire earlier in abscission layer formation than in wild rice.

3.2.3 Seed Shattering Time Course Profiles are Altered in Weedy Rice Compared to the Wild Relatives.

The early degradation of US weedy rice abscission layers may confer an earlier shattering phenotype than reported for wild rice. Earlier degradation of the abscission layer suggests that as soon as the weedy seed is mature, or nearly so, it can more readily

fall to the ground. The timing of seed release is considered important to weed fitness, as it may be beneficial to disperse seeds prior to harvest (Shivrain *et al*., 2010a); earlier shattering could thus be a response to rice cultivation practices. Additionally, or alternatively, earlier release may prevent seeds from drying out and losing dormancy, another trait that enhances weediness (Gu *et al*., 2005); higher moisture content in seeds is known to confer a greater level of dormancy (Delouche *et al*., 2007), but desiccation of rice seeds occurs as they mature. Easy shattering may not necessarily always be an advantage, however. Seeds that shatter before they are mature enough to germinate will lower a plant's fitness (Oba *et al*., 1995).

Phenotypically, little is known about the shattering levels in weedy rice groups across floral/seed development. Previous studies in cultivated and wild rice have shown that shattering level increases dramatically after 15 days post flowering in wild rice and in some cultivated rice samples grown in both field and greenhouse settings (Ji *et al*., 2006; Oba *et al*., 1995). In an effort to determine if shattering levels mirror the observed formation and degradation of the abscission layer in US weedy rice groups, we assessed levels of shattering as the amount of weight a grain can hold prior to release from the panicle (breaking tensile strength; BTS) in eight cultivated, five wild and seven weedy rice individuals, at various time points through seed development (Figure 3.3; Table 3.2).

To date, we have examined eight cultivated rice varieties from the *indica*, *aus* and *tropical japonica* groups (Table 3.2). Four of these samples are shown in Figure 3.3A (3A06, 3A11, 2B03 and 3A09). All cultivated rice accessions show consistent high BTS values between 150 g to 250 g from before flowering through ten days after flowering. By 15 days after flowering, BTS values have dropped close to the level previously seen

in these cultivars at maturity (between 25 g and 125 g), and remain at these levels through 30 days after flowering, consistent with measurements reported in (Thurber *et al*., 2010). The five wild rice individuals surveyed (2F02, 2C12, 2C04, 2C02 and 2C09) show a similar shattering pattern to cultivated rice up through ten days post flowering (Figure 3.3B; Table 3.2). However, at 15 days post flowering, the BTS levels have dropped dramatically to near 0 g and stay at this level through 30 days post flowering (Figure 3.3B; Table 3.2). This is consistent with all reported observations of *O. rufipogon* and *O. nivara* shattering behavior across floral development (Ji *et al*., 2006; Oba *et al*., 1995), and is consistent with the wild rice seed shattering trait at maturity (Table 3.1).

All six shattering weeds examined (SH_1A08, SH_1A09, BHA1_1B08, BHA1_1A05, BHA1_1C04 and BHA1_1B02) registered BTS values above 150 g five days before through five days after flowering (Figure 3.3C; Table 3.2). By ten days after flowering, BTS values for three weeds (SH_1A08, BHA1_1C04 and BHA1_1A05) have dropped to below 60 g, while all other weeds are still registering values around 150 g. By fifteen days after flowering, all shattering weeds shown have dropped their BTS values dramatically to nearly 0 g (Figure 3.3C; Table 3.2). The BTS values thereafter stay at 0 g throughout the remainder of seed maturation for all shattering weeds shown. The single non-shattering weed (MXSH_1B06) shows a different time course as the shattering weeds. The sharpest decreases in BTS values are only seen after 20 days after flowering and instead of dropping to 0 g the BTS values for this individual only go as low as 40 g (Figure 3.3C; Table 3.2).

The variation in timing of the sharp reduction in BTS values across the weeds surveyed indicates that shattering ability is only partly correlated with abscission layer

degradation rates. Though all weedy rice accessions used in our microscopy study displayed earlier degradation of the abscission layer than what is seen in wild rice, a range of degradation severity seems to exist (Figure 3.2; Figure 3.4). Two weed samples that showed reduction in BTS values five days prior to other weeds tested appear to possess the highest degraded abscission layers at flowering (Figure 3.2). Weeds with drastically reduced BTS values at 15 days, a timing consistent with that of wild rice, seem to have somewhat less-degraded layers at flowering (Figure 3.4). Overall the weedy rice individuals that showed the least degradation at flowering have similar shattering time courses to what has been shown previously for wild rice, while those with the most degradation show an earlier drop in BTS values. This indicates that the timing of when shattering is first noticeable in weedy rice is variable, despite the fact that all weeds degrade their abscission layer at an earlier time than wild rice.

3.2.4 Novel Mutations Likely Underlie the Parallel Evolution of Shattering in Weedy and Wild Rice.

Previous studies of the *sh4* locus in wild and domesticated rice have implicated this gene in both the formation and degradation of the abscission layer at the flower-pedicel junction (Li *et al*., 2006b; Lin *et al*., 2007). A mutation in the *sh4* gene, strongly selected upon during rice domestication, is associated with reduction in shattering in cultivated rice varieties due to the formation of a discontinuous abscission layer (Li *et al*., 2006b). Transgenic experiments have further demonstrated that the ancestral *sh4* allele (present in wild *O. rufipogon*) can restore shattering in non-shattering cultivated rice (Li *et al*., 2006b). Our previous work showed that US weedy rice groups carry the derived non-shattering mutation fixed in cultivated rice (Thurber *et al*., 2010), demonstrating that

the functional mutation identified in the *sh4* locus does not result in non-shattering in the weed, and is thus not sufficient for loss of shattering. This suggested that novel loci, perhaps distinct from those acting in wild rice species, are involved in the evolution of shattering in US weedy rice groups.

The distinct developmental profile observed here for weedy rice abscission layers further supports that US weedy rice groups did not acquire the shattering trait through introgression with wild species. Thus, this and our previous work (Thurber *et al*., 2010) suggest that parallel evolution of shattering in weedy and wild rice has occurred through both different loci and different developmental mechanisms. Studies in several other systems have shown that parallel evolution between populations can arise from independent mutations in the same gene, as has been shown for body shape characteristics in two independent populations of freshwater stickleback and for two independently evolved populations of melanic *Peromyscus* rodents (Kingsley *et al*., 2009; Schluter *et al*., 2004). Conversely, studies of independent melanic populations of rock pocket mice have also shown that convergent phenotypes can sometimes be achieved through mutations in different genes (Nachman *et al*., 2003; Hoekstra & Nachman, 2003). The acquisition of the shattering trait in wild and weedy rice groups further supports the possible role of independent loci in parallel evolution.

Interestingly, the similarities in abscission layer traits (development and shattering time course) between two distinct weedy rice groups, SH and BHA, suggest that the gene(s) involved in reacquiring seed shattering may be the same in both populations. This is surprising, as these groups have been shown to have independent evolutionary origins (Reagon *et al*., 2010; Londo *et al*., 2007). The convergence in the mechanistic basis of

seed shattering among these weedy rice groups may indicate certain genetic or morphological constraints inherent to re-evolving the shattering trait after its loss through domestication. Future studies into the genes involved in the progression of abscission layer formation and degradation in both weedy and wild rice will be integral to the study of weed evolution.

3.3 Conclusions

Our results show that the shattering trait in US weedy rice has a distinct mechanistic basis from that of the shattering wild ancestor of rice, consistent with the re-evolution of this trait in weedy groups from domesticated ancestors. Surprisingly, independently evolved weedy groups have converged on this feature of abscission layer development. In some cases, the altered timing of abscission layer degradation appears to lead to earlier shattering in weedy rice compared to wild rice.

3.4 Methods

3.4.1 Plant Materials for Microscopy

All accessions used in this study are a subset of those used in (Thurber *et al*., 2010) for which phenotypic and sequence data are available. Five weedy rice accessions, collected in the Southern US rice belt, were generously supplied by David Gealy (USDA) (Table 3.1). Accessions were chosen to represent the two major weedy rice groups (SH and BHA) based on population structure analysis (Reagon *et al*., 2010) and a group of putative weed-crop hybrids (MX) showing some resistance to seed shattering. Additional samples of wild and cultivated *Oryza* were originally obtained from the International Rice Research Institute (IRRI) (*O. rufipogon* (4) and *O. nivara*, a close relative or annual

ecotype of *O. rufipogon* (2)) and Susan McCouch (*O. sativa* (4)). All plants were grown in a Conviron PGW36 growth chamber at the University of Massachusetts Amherst. One seed per accession was planted in a 4 inch pot and grown as described in (Thurber *et al*., 2010). Panicles from wild and cultivated individuals were collected at flowering, while panicles from weedy individuals were harvested at three time points: one week prior to flowering, at flowering and one week after flowering. For observations prior to flowering, panicles were collected when the boot, or flag leaf sheath, was swollen yet before flowers had begun emerging. At flowering, panicles were collected once 50% of the panicle had emerged from the boot. Panicles to be collected after flowering were bagged upon flowering to prevent pollen flow and loss of seeds. At each collection, approximately eight flower-pedicel tissue samples were excised from the flowers at the topmost end of the panicle using a dissecting scope.

3.4.2 Microscopy

Tissue samples were fixed with glutaraldehyde (100 mM) in a solution containing 100 mM PIPES pH 7.0, 100 mM Glutaraldehyde, 0.5 mM $CaCl_2$, and 5.0 mM $MgCl_2$ for 2 hours. Following fixation samples were dehydrated first in an ethanol series then further dehydrated in acetone. Dehydrated samples were infiltrated and embedded in Epon Araldite resin (Mollenhauer, 1964). Samples were sectioned longitudinally using a diamond knife on a rotary microtome (Porter-Blum JB4) to create 2 micrometer sections. Sections were dried onto rectangular microscope slides and subsequently stained for 3 minutes with Toluidine Blue (0.5% solution in 0.1% sodium carbonate, pH 11.1), a metachromatic dye which stains regions with high lignin dark blue-green and regions of unlignified cell wall reddish purple (see O'Brien *et al*., 1964). Bright field images were

taken at both 10x and 60x using a Nikon TE 300 Inverted Microscope with an attached CCD camera (Quantix CoolSnap HQ; Roper Scientific).

3.4.3 Time Course Shattering Measurements

Five weedy rice accessions, along with five wild rice accessions and eight cultivated *O. sativa* accessions (see above) were analyzed for shattering ability during floral and seed development (Table 3.1). All plants were grown as described above for microscopy. Panicles from each individual were collected ~ 5 days before flowering (swollen boot with top most flower of panicle approaching emergence), at flowering (50% of panicle emerged from boot), as well as 5, 10, 15, 20, 25, and 30 days after flowering. Upon flowering, panicles to be collected were bagged to prevent pollen flow and loss of seeds. The oldest (topmost) 10 flowers per panicle were analyzed for breaking tensile strength (BTS), or shattering level, using a digital force gauge as described in (Thurber *et al*., 2010). BTS is a measure of the maximum amount of weight, in grams, a single flower or grain can hold before releasing; values at or near zero grams (g) are considered highly shattering while values over 100 g represent non-shattering or hard threshing (Li *et al*., 2006b; Thurber *et al*., 2010; Konishi *et al*., 2006). Average BTS values for the ten measurements are reported for each sample.

Table 3.1: List of accessions used for this study. Accessions are identical to those used in a previous study (Thurber *et al*., 2010) and are grouped by type (weed, wild or cultivar). Identification numbers as well as phenotypic values for seed shattering are reported here as well as in (Thurber *et al*., 2010).

Group	Study ID [a]	USDA ID/ Common Name [c]	IRGC/RA/GRIN	Origin [b]	Mean BTS (gram) [d]	Std. Dev
Weedy rice	SH_1A08*	1134-01	x	AR	0	0
	SH_1A09*	1135-01	x	AR	0.3	0.5
	SH_1C02*	1001-01	x	AR	1	2
	MXSH_1B06*	1996-01	x	AR	35.6	17.9
	BHA1_1B08*	1996-09	x	MS	7.2	21.6
	BHA1_1A05*	1096-01	x	AR	0	0
	BHA1_1B02	10A	x	AR	0	0
	BHA1_1C04	1005-02	x	AR	0	0
Cultivated rice						
aus	3A06*	BJ-1	RA5345/45195	India	18.3	3.1
	2B03	Aus 196	29016	Bangladesh	12.3	9.8
	3C05	Dee_Geo_Woo_Gen	RA5344/PI279131	Taiwan	60.9	25.3
indica	3A11*	Dholi Boro	RA4984/27513	Bangladesh	137.4	11.8
	3A08*	Rathuwee	RA4911/8952/PI584605	Sri Lanka	72.3	47.8
	2B02	Bei Khe	22739	Cambodia	30.1	17.5
	3A09*	Khao Dawk Mali - 105	RA4878/27748	Thailand	80.7	42.6
tropical japonica	3B09	Mirti	RA4970/25901/PI584553	Bangladesh	12	22.9
	3B12	Gotak_Gatik	RA4959/43397/PI584572	Indonesia	104.5	67.7
Wild Asian rice						
O. rufipogon	2C02*	N/A	100588	Taiwan	0	0
	2C09	N/A	104833	Thailand	0	0
	2C04	N/A	100916	China	0	0
	2C12	N/A	105491	Malaysia	0	0
	2D06*	N/A	106086	India	0	0
	2D12*	N/A	106169	Vietnam	0	0
	2E01*	N/A	106321	Cambodia	0	0
O. nivara	2F01*	N/A	86662	Thailand	0	0
	2F02*	N/A	103821	China	0	0

a Based on STRUCTURE and identity from Reagon *et al*, 2010

b Origin for weeds is a US state abbreviation, origins for cultivated and wild rice is country

c Accessions with RA numbers were acquired from Susan McCouch while all others were acquired from IRRI

d BTS (Breaking Tensile Strength) corresponds to the maximum weight a seed can hold before releasing; from data reported in Thurber *et al*, 2010

*-- Individuals used for Microscopy; all others used only for shattering time course

x-- no data available

Table 3.2.: Average BTS values across floral and grain development. Average BTS values for each individual at -5, 0, 5, 10, 15, 20, 25, and 30 days after flowering, recorded in grams.

			Days After Flowering						
Group	**Study ID**	**-5**	**0**	**5**	**10**	**15**	**20**	**25**	**30**
Weedy rice									
	SH_1A09 *	230.6	221	195.6	152.7	2.3	0	0	0
	SH_1A08 *	204.4	187.5	173.4	35.2	5.8	0	0	0
	BHA1_1B08 *	221.1	212.8	192.6	153.2	4.5	0	0	0
	BHA1_1B02	182.5	167.9	135.4	103.3	0	0	0	0
	BHA1_1A05 *	217.1	207.3	185.9	57.2	16.9	0	0	0
	BHA1_1C04	190.7	178.1	135.8	6.7	0	0	0	0
	MXSH_1B06 *	244.2	236.9	239.4	186.1	194.9	146.6	71.8	40.3
Wild Asian rice									
	2F02 *	228	195.4	172	144.4	14.5	0	0	0
	2C12	188.3	168.5	151.7	147.9	6.28	0	0	0
	2C04	195.4	176.3	162.2	162.8	4.93	0	0	0
	2C02	150.2	136.8	124.4	121	76.7	13.2	0	0
	2C09	128.4	127.7	116.7	104.1	6.37	0	0	0
Cultivated rice									
tropical japonica	3B09	195.2	187.6	175.5	185.3	22.3	19.1	18.1	18.9
	3B12	198.2	179.6	147.5	166	112.9	66.1	83.8	68.4
indica	3C05	234.3	230.7	203.3	223.8	75.8	50.4	58.6	44.8
	3A09 *	201.2	184.3	148.9	172.8	53.6	63.5	44.6	56.3
	3A11 *	224.3	225.4	212	184.5	136.5	136.7	119.3	123.2
	2B02	215.8	197.2	181.2	161.3	155.1	61.7	34.7	30.2
aus	2B03	231.1	213	207.4	134.3	35.6	24.9	13.3	14.6
	3A06 *	226.7	220.8	197.6	123.6	14.6	14.2	12.2	14.6

*-- Individuals used for Microscopy

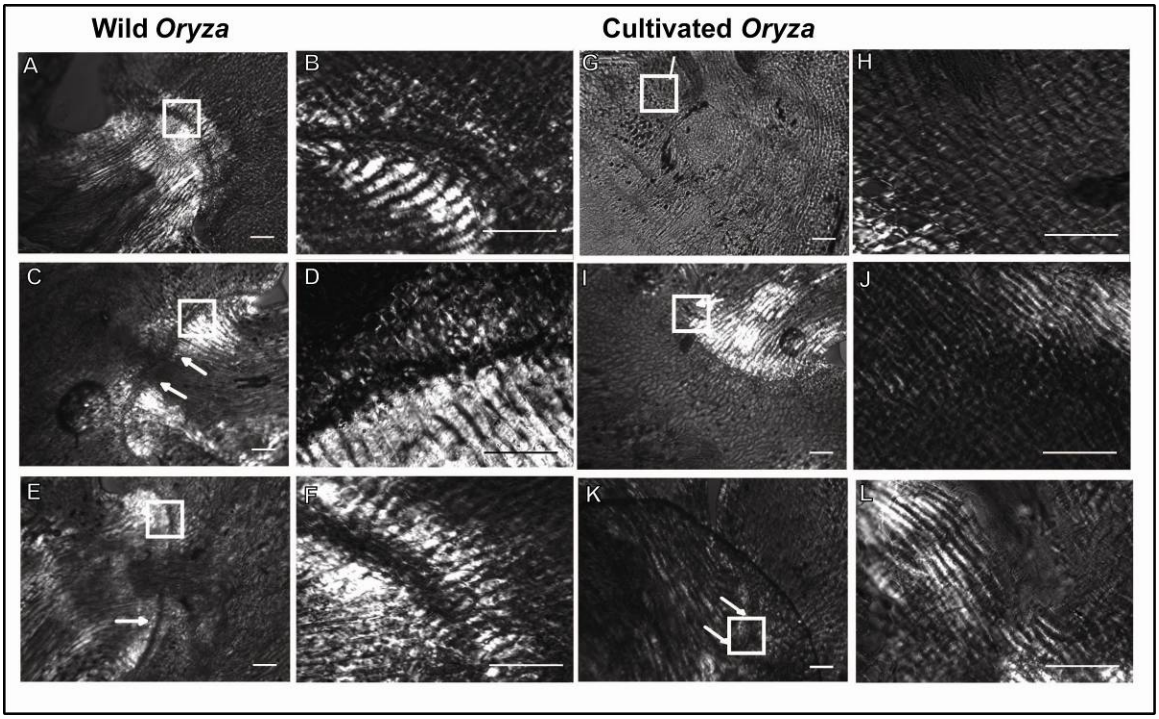


Figure 3.1. Comparison of wild and cultivated *Oryza* flower-pedicel junctions.
Panels A-F are wild *Oryza* (A/B- 2F02 (*O. nivara*), C/D- 2F01 (*O. nivara*), E/F- 2C02 (*O. rufipogon*)). Panels G-L are cultivated *O. sativa* varieties (G/H- 3A11 (*indica*), I/J- 3A06 (*aus*), K/L- 3A08 (*indica*)). Arrows point to the region of the abscission zone, while white boxes show the region magnified further at right. Abscission layers can be seen as darkly stained bands. All samples shown here were taken at flowering for their respective accession and are all magnified at 10 x on the left and 60 x on the right. Scale bars on bottom right represent 100 μm for 10x images and 50 μm for 60x images.

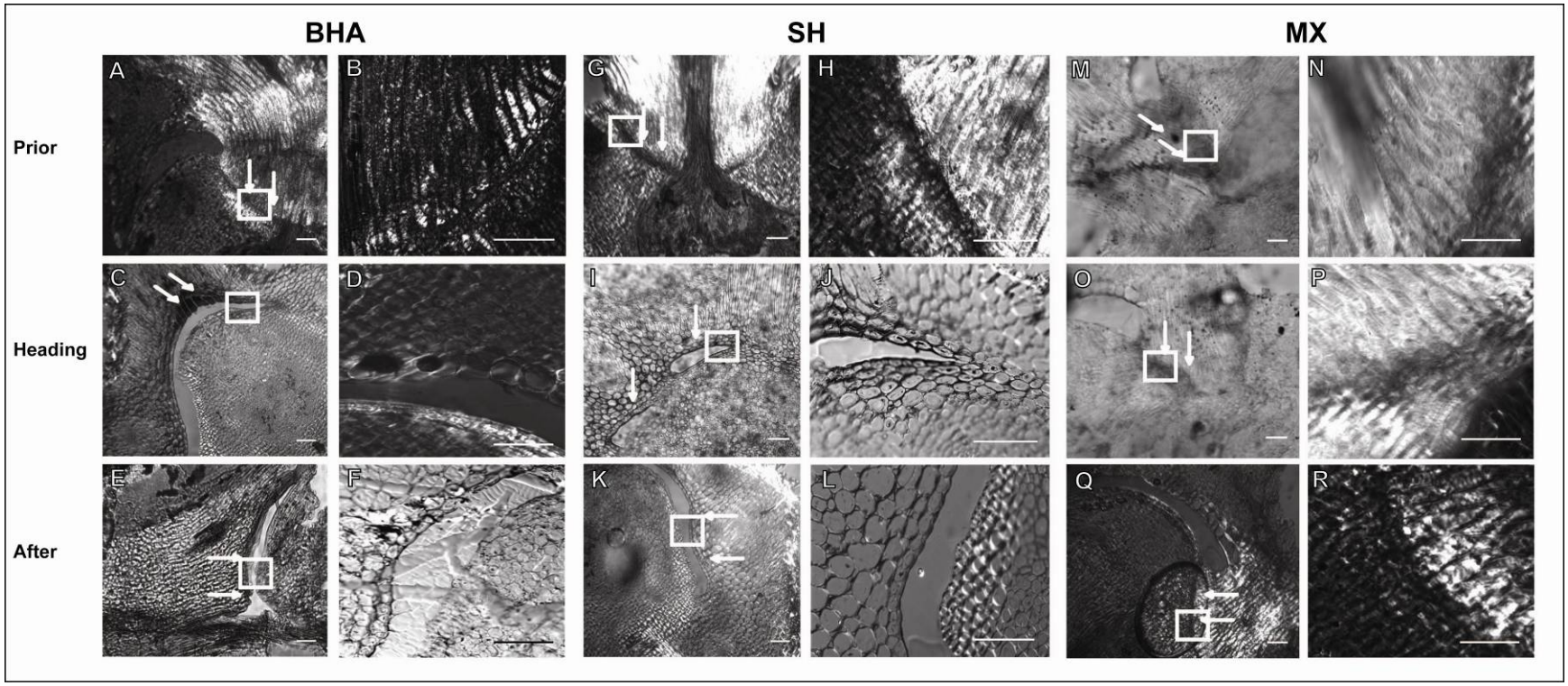


Figure 3.2. Comparison of abscission layers across weedy *Oryza* populations.
Panels A-F are shattering BHA_1A05, Panels G-L are shattering SH_1A08, Panels M-R are non-shattering MXSH_1B06. Each individual was collected 1 week prior to flowering (Prior), at flowering (Flowering) and 1 week after flowering (After). Arrows point to the region of the abscission zone while white boxes outline the region magnified further. Abscission layers can be seen as darkly stained bands. Images at left were taken at 10 x magnification while those at right are 60 x magnification. Scale bars on bottom right represent 100 μm for 10x images and 50 μm for 60x images.

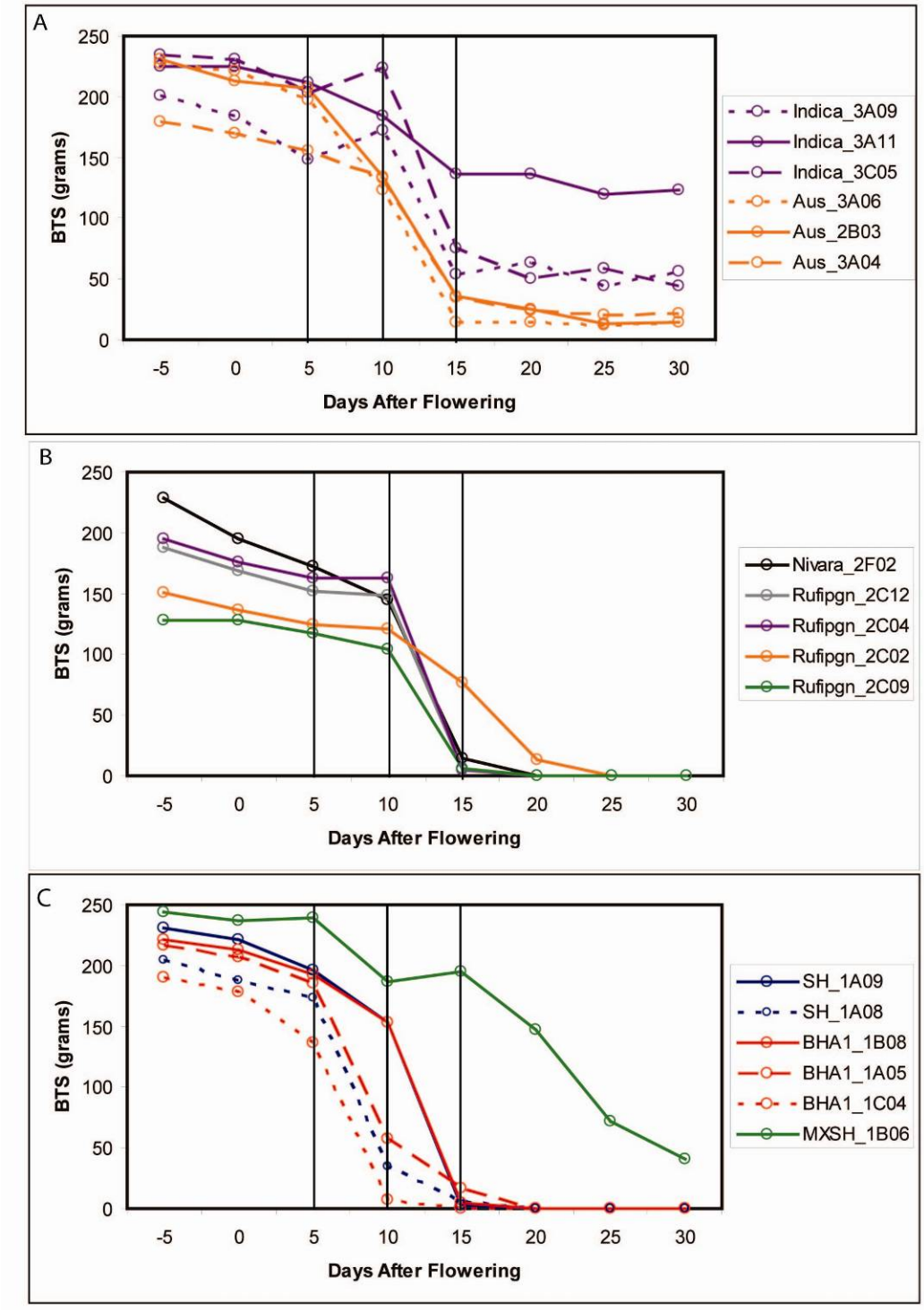


Figure 3.3. Shattering across floral and grain development.
Shattering levels for cultivated (4), wild (5) and weedy (5) individuals were recorded every five days from 5 days prior to flowering (-5) through 30 days after flowering (30). Panel A shows shattering levels for cultivated rice, Panel B shows shattering levels for wild rice, and Panel C shows shattering levels for weedy rice.

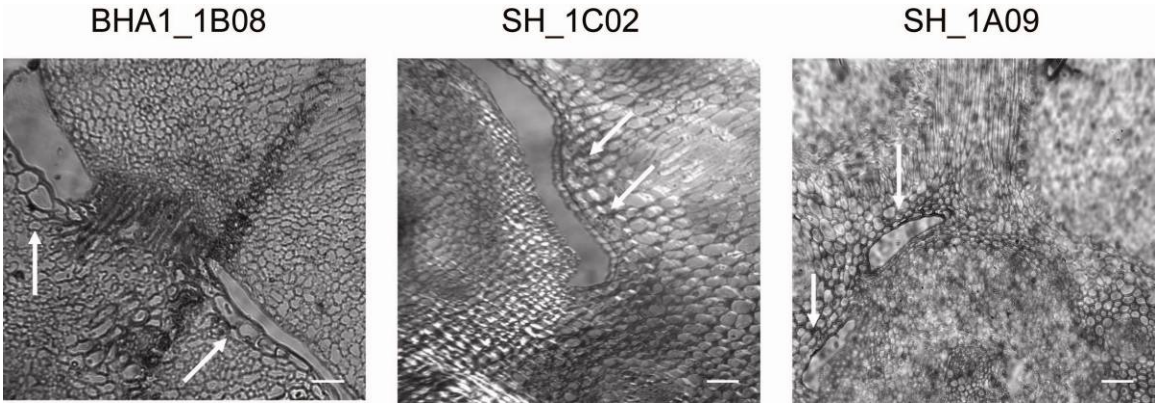


Figure 3.4. Additional weedy rice abscission layer images at flowering.
Samples shown here were taken at flowering for their respective accession and are all magnified at 10 x with scale bars on bottom right representing 100 μm. Arrows point to the breakdown of the abscission layer.

CHAPTER 4

EVOLUTION OF FLOWERING TIME IN US WEEDY RICE

4.1 Introduction

Flowering time variation in plants is important for local adaptation. Multiple environmental variables, such as temperature and latitude, with different day lengths depending on season, can act as cues for flowering time in different species (Weber & Schmid, 1998; Franks *et al.*, 2007). It has been observed that tropical plants often flower during cooler seasons, as day length decreases, to avoid extreme heat, while temperate plants flower during warmer seasons, as day length increases, to avoid freezing temperatures in winter (Greenup *et al.*, 2009). This is true for both wild and weedy plants as well as for domesticated crops. As such, flowering time has been selected upon in multiple plant species and to different degrees and directions.

In crops, humans have most commonly selected to reduce or eliminate photoperiod sensitivity so that crops, especially cereals, can be grown in locations outside of their native range (Sawers *et al.*, 2005). Flowering time is thought to play an important role in the ability of agricultural weeds to compete with crops in the field (Ellstrand *et al.*, 2010). Some weeds may benefit from flowering simultaneously with their local crop, as this decreases conspicuousness and its seed may be collected and replanted. Weeds can also benefit from earlier flowering and seed dispersal before crop harvest, thus escaping into the seed bank. For conspecific weeds (weeds related to the crop they invade), many species show differences in flowering phenotype when compared to their cultivated relatives (Ellstrand *et al.*, 2010).

As described in previous chapters, the weed my research focuses on is weedy rice found in the Southern US. Genome-wide data suggest that in the United States weedy rice forms two major genetically differentiated groups; the SH group tends to have straw colored hulls with no awns and the BHA group tends to have black hulls with long black awns (Londo & Schaal, 2007; Reagon *et al.,* 2010). Genome-wide data also suggest that each weed group independently arose from Asian cultivated rice ancestors, SH from *indica* and BHA from *aus*. Both of these cultivated ancestors are varieties of rice that are thought to stem from a common domestication event from *O. rufipogon*, wild Asian rice (Sweeney & McCouch, 2007; Londo *et al.,* 2006).

Despite the likely descent of weedy rice from domesticated rice, weedy rice possesses many traits different from cultivated rice. As is shown in chapters 2 and 3, weedy rice shatters and disperses its seed, unlike the crop in which shattering has been selected against to facilitate harvest (Thurber *et al.,* 2010; Thurber *et al.,* 2011). In Southern US rice fields, straw hulled weedy rice typically flowers earlier than weedy rice with black hulls and also shows some photoperiod sensitivity in relation to region of origin (Shivrain *et al.,* 2010b; Shivrain *et al,.* 2009). Flowering time in weedy rice may also overlap with flowering time in the US crops as there is a lot of variation in this trait within weed ecotypes (Shivrain *et al.,* 2010b).

In rice, flowering time, also referred to as heading date, is known to be affected by photoperiod (day length), but has also been suggested to be regulated by temperature, with photoperiod insensitive plants affected more than sensitive plants (Luan *et al.,* 2009). Flowering time is highly variable within both cultivated and wild rice, although rice is commonly referred to as a short-day (SD) plant (Yano *et al.,* 2000; Dung *et al.,*

1998). Under SD, less than 12 hours of daylight, flowering is promoted, while under long-day (LD) conditions, day lengths longer than 14 hours, flowering time is delayed (Yano *et al.*, 2000). The flowering time regulatory network has been well worked out in rice, and several genes are known. For a complete review of flowering time regulation see Tsuji *et al.* (2010). Briefly, under SD, *OsGI*, ortholog of *Arabidopsis GIGANTEA*, activates both *Hd1*, which encodes a B-box zinc finger protein and is the ortholog of *Arabidopsis CONSTANS*, and *Ehd1*, a B-type response regulator of which there is no *Arabidopsis* ortholog (Figure 4.1). Both genes go on to activate *Hd3a*, which encodes a Phosphatidylethanolamine-binding protein and is the ortholog of *Arabidopsis FT* and the mobile florigen, whose protein originates in the leaves and moves to the shoot apical meristem, initiating the changeover from vegetative to reproductive growth (Tamaki *et al.*, 2007; Yang *et al.*, 2007). Under long days, several MADS-box transcription factors act on *Ehd1* to promote flowering through activation of *Hd3a* and *RFT1*, a close paralog of *Hd3a* (Tsuji *et al.*, 2010). However, concurrent with this, *Hd1* acts negatively on *Hd3a* to repress flowering (Figure 4.1). Flowering under long day conditions is very complex and it is still unclear exactly what the cut off is between a long and short day. Importantly, a recent study (Takahashi *et al.*, 2009) has shown that the three major determinants of flowering time diversity in cultivated rice are variation in *Hd1* coding sequence affecting protein function, *Ehd1* expression levels and *Hd3a* promoter sequence.

Since flowering time is an important agronomic trait and has been implicated as a major difference between weedy and cultivated rice we were interested in finding out how flowering time has evolved in US weedy rice. Our two main questions were: 1) How

does flowering time differ between weedy rice groups and between weedy rice and their cultivated progenitors? and 2) Do the genes that control flowering time variation in cultivated rice also play a role in flowering time evolution in US weedy rice? We found that weedy rice populations have diverged in their flowering phenotype and that these phenotypes are different from the weeds' cultivated ancestors. Additionally, two major determinants of flowering time in cultivated rice appear to be only partially responsible for these differences in flowering time in weedy rice.

4.2 Methods

4.2.1 Plant Materials

All plant material for this study was previously described in (Thurber *et al.,* 2010) and include 58 US weedy rice accessions and 87 samples of AA genome *Oryza* species including cultivated Asian *O. sativa* from the *indica, aus, aromatic, tropical japonica* and *temperate japonica* groups as well as wild *O. rufipogon* and *O. nivara*. Additional outgroup species of non-Asian cultivars and wild relatives (*O. glaberrima, O. barthii, O. glumaepatula* and *O. meridionalis*) were also included. For further details of origin and collection see Table 4.1.

4.2.2 DNA Extraction and Sequencing

Plants used for DNA extraction were grown in Conviron PGW36 growth chambers at the University of Massachusetts Amherst and DNA was extracted and quantified as described in (Reagon *et al.,* 2010). Primers for the *Hd1* gene open reading frame and *Hd3a* promoter were designed using Primer3 (Rozen & Skaletsky, 2000) based on the *O.*

sativa genome (TIGR v. 5 January, 2008). Initial PCR amplification and sequencing was performed by Cogenics (Houston, TX), while DNA sequencing of in-house PCR products was performed by either Cogenics (Houston, TX) or Beckman Coulter Genomics (Danvers, MA) as described in (Caicedo *et al.*, 2007; Olsen *et al.*, 2006). Sequence aligning including base pair calls, quality score assignment and construction of contigs was done as in (Caicedo *et al.*, 2007) using BioLign Version 2.09.1 (Tom Hall, NC State Univ.). Approximately 1.25 kilobases (kb) of exonic sequence data was generated for each individual at the *Hd1* gene. An additional ~1 kb of data was generated for the *Hd3a* promoter region. Due to high conservation of the *Ehd1* promoter and coding sequences in rice this gene was not investigated despite its expression level being highly correlated with flowering time (Takahashi *et al.* ,2009). Primers used in this study can be found in Table 4.2.

4.2.3 Data Analysis

Maximum Parsimony and Neighbor Joining phylogenetic trees were made using Mega 4 (Tamura *et al.*, 2007) under default conditions. As both trees produced similar results, only the consensus Neighbor Joining trees resulting from 1000 bootstraps, with bootstraps reported at branches over the 50% cut off value are shown (Figures 4.3 and 4.5).

4.2.4 Measurement of Heading Date

As described in (Thurber *et al.*, 2010), a subset of 90 *Oryza* accessions was grown in growth chambers under day neutral conditions and phenotyped for heading date. Rice's

evolutionary origin is in tropical latitudes where day neutral conditions are most common (Khush, 1997). The range of changing day lengths experienced in a typical southern US rice field can not be captured in a growth chamber, but flowering of both weedy and cultivated rice likely happens after the summer solstice but before the autumnal equinox (D. Gealy, personal communication). This puts the day length experienced by these plants below 14 hours (LD) and closer to day neutral conditions. Additionally, we also believe that the SD path is the one active in the field and growth chamber, yet we can not be sure given that the boundary between what is considered a long and short day has not been well-explored (Figure 4.1). For further descriptions of plant growth conditions please see Thurber *et al.* (2010). Heading date was measured as the approximate 50% heading time, between when the first few florets began emerging from the boot until anthesis of those first florets, on the first emerging panicle for each plant (as described by (Counce *et al.*, 2000)). Dates were then transformed into number of days relative to the germination date, which was fairly uniform across all individuals. Averages were calculated for each individual and for each major group using Excel (Table 4.1). Boxplots of the flowering date data were made using R (Figure 4.2).

4.3 Results and Discussion

4.3.1 Heading Date Phenotype

In the field in the US, weedy rice has been documented as showing range in flowering dates in relation to the flowering date of the local US crop, typically *tropical japonica* (Shivrain *et al.*, 2010b). Our common garden experiment mirrors what has been observed in the field; SH weeds flower significantly earlier than *tropical japonica,* while

BHA weeds flower concurrently or after (Figure 4.2 Kruskal Wallis test; P <0.001, see also (Reagon *et al.*, 2011)). However, flowering time in the weeds differs from that observed in their cultivated ancestors. We found that SH weeds flower significantly earlier than their progenitor *indica* and most wild accessions (Figure 4.2; Mann-Whitney (P <0.05) (Reagon *et al.*, 2011)). BHA weeds flower significantly later than their progenitor *aus*, but more in line with what we see in wild accessions (Figure 4.2; Mann-Whitney (P <0.05) (Reagon *et al.*, 2011)). Among cultivars, the *aus* cultivars flowered slightly earlier than either *indica* or *tropical japonica*, however this difference is not statistically significant (Figure 4.2). The *O. rufipogon* wild rice showed a wide range in heading dates, from early to late. Additionally, out of a variety of traits related to how weeds grow (height, tiller number, flowering time, emergence growth rate and average growth rate), flowering time has been shown to be the most divergent trait among crops and weeds under our growth chamber environment, which suggests this trait has been under selection (Reagon *et al.*, 2011). Our phenotypic data suggests that there is no single optimal weed strategy to flowering, indicating that this trait has not evolved in parallel within the two weed groups. Additionally, flowering time differs markedly between weeds and their ancestors, suggesting this trait has evolved rapidly since the weeds arose.

4.3.2 Relationship of Weedy, Cultivated and Wild *Hd1* Alleles

To try to understand the differences in flowering phenotype between the weed groups and their progenitors, we investigated sequence polymorphism at the major flowering time candidate gene *Hd1*. In particular, we wanted to know whether the US

weeds had functional or non-functional *Hd1* alleles. *Hd1* has been implicated in many studies as being the most important and strongly selected locus for photoperiod sensitivity and flowering time variation in cultivated rice (Takahashi *et al.,* 2009; Yano *et al.*, 2000; Fujino *et al.*, 2010). Multiple alleles in cultivated rice harbor deletions or SNPs that render the resulting protein non-functional and cause later flowering under short days. The most common of these alleles, present in ~43% of rice cultivars, is a 2 base pair (bp) deletion in the second of two exons that causes a premature stop codon and is shared between *indica* and *japonica* varieties (Takahashi *et al.*, 2009).

We sequenced both exons of the *Hd1* gene (chromosome 6) in a panel of 144 wild, weedy and cultivated rice accessions and visualized the relationship of *Hd1* haplotypes on a Neighbor Joining tree (Figure 4.3). This is a highly diverse gene with over 50 haplotypes for the coding region alone. A majority of the haplotypes are unique to wild rice, yet there are several haplotypes unique to cultivated rice and even a few unique to weeds (Figure 4.3). The most common haplotypes (Haplotypes 2 and 3) are shared between weeds and cultivars.

The haplotypes containing the well characterized 2 bp deletion, including Haplotype 3, are present in nearly all BHA weeds along with all US cultivars and a subset of wild and Asian domesticates including a majority of *aus* cultivars (84%), the progenitors of BHA weeds. This is consistent with the later flowering observed in these groups and is also consistent with Takahashi's (2009) finding that the deletion is common in cultivated rice from both *indica* (*indica/aus*) and *japonica* (*tropical japonica/temperate japonica/aromatic*) lineages. In contrast, the SH weeds form a separate clade of haplotypes that does not contain the deletion and includes a small subset of Asian

domesticated accessions, three of which are *indica* cultivars, the progenitors of SH weeds. Haplotype 2 represents the majority of the SH weeds (91%) and cultivars, including 20% of *indica* sampled, that do not have the 2 bp deletion in their coding sequence, consistent with early flowering. The *indica* cultivar group we sampled is highly diverse in their *Hd1* alleles, representing seven different haplotypes with no more than 30% of *indica* cultivars containing a single shared haplotype. The fact that there is no majority haplotype in *indica* and that SH weeds share a haplotype with a subset of these cultivars from China and Cambodia, may be useful for narrowing down the origin of this weed group.

Additional non-functional alleles have been identified by Takahashi (2009), yet only one is present in our study panel. This allele, Haplotype 4, is shared by a single *indica* cultivar and a single weed of mixed ancestry and contains a four base pair deletion shown to produce a non-functional *Hd1* allele (Takahashi *et al.*, 2009). Other novel deletions, ranging in size from single bases to 43 bases, which may reduce or eliminate function of this gene, are present in several haplotypes (Haplotypes 1, 7, 27, 37, 40, and 43). Some of these haplotypes (1, 7 and 43) can be found in some cultivars or a few SH weeds, while others contain separate deletions unique to wild rice (27, 37 and 40). Additional SNPs that cause amino acid changes are present, yet the extent to which these mutations cause functional changes in the protein is not known.

Since the functionality of the HD1 protein is integral to rice flowering and the common 2 bp deletion has been shown to increase flowering under SD, we wanted to know if there was a difference between individuals with or without functional alleles. Due to small samples sizes of the weeds and cultivars separately and the close relatedness

of these groups, we chose to pool these individuals. However, this means that our analysis does not account for population structure. When weeds and cultivars are pooled, there is a significant difference in days to flowering between individuals with functional *Hd1* alleles and those with the 2 bp deletion (Figure 4.4). Those with a loss of function allele flower significantly later than those with a functional allele ($P = 5.62e^{-8}$). This suggests that flowering time behavior in weedy rice is partially determined by *Hd1* polymorphism and by alleles that have been inherited from ancestral groups.

The differences between the weed groups (SH/BHA) can be explained by their divergent *Hd1* haplotypes. However, *Hd1* haplotype alone cannot explain the differences in flowering time phenotype between weeds and their progenitors. SH weeds share an allele with *indica* cultivars and BHA weeds share an allele with *aus* cultivars, yet SH weeds flower earlier than *indica* cultivars and BHA weeds flower later than *aus* cultivars. Interestingly, the *aus* cultivars show low levels of both phenotypic variation and numbers of *Hd1* haplotypes while *indica* cultivars show nearly twice as much phenotypic variation and many more *Hd1* haplotypes (Table 4.1; Figure 4.2).

4.3.3 Identification of Weedy Rice *Hd3a* Promoter Haplotypes

As described above, polymorphism at *Hd1* seems to explain flowering time differences between US weedy rice groups but does not explain how weeds can share haplotypes with their ancestors and still show different flowering phenologies. Thus we decided to investigate another major flowering time determinant, *Hd3a*, the mobile florigen that sparks the transition to flowering. This gene is located on chromosome 6 and is about 6.4 Mb upstream of *Hd1* (Kojima *et al.*, 2002). Research has shown that the

Hd3a promoter type is highly associated with flowering time in cultivated rice and that there are two main promoter types of *Hd3a* in cultivated rice (Takahashi *et al.*, 2009). Type A promoters lead to lower expression of *Hd3a* and a later flowering phenotype, while Type B promoters have higher *Hd3a* expression and earlier flowering under SD. These promoter types differ by eleven SNPs and a twelve bp indel, any of which may be responsible for the differences in gene expression. However, none of these mutations alters known regulatory sequences. Both types of promoter occur in *indica* and *japonica* varieties.

We sequenced ~1kb of promoter region in 84 accessions of wild, weedy, and cultivated rice. In our data set, there is a moderate amount of diversity in *Hd3a* promoter haplotypes (Figure 4.5). The type A promoters (Haplotypes 1, 4 and 5) are distinct from the Type B promoters (Haplotypes 2, 3, 6, 8, 10, 21 and 26). The inclusion of weedy rice brings in new recombinant haplotypes, which cannot be classified as A or B types based on sequence polymorphism, and might even be intermediate in expression level. About 42% of BHA and 11% of SH weeds have these unique recombinant promoter types. However, the majority of SH weeds (75%) group with *indica* (50%), sharing B type promoters, consistent with early flowering seen in both the growth chamber and field. Interestingly, most BHA weeds (42%) group with *aus* (60%), also sharing a B type promoter, which is unexpected given their later flowering. It is important to note that there is one Type A haplotype that is shared between a subset of both SH (11%) and BHA (17%) weeds and all of the local cultivars sampled, which is suggestive of hybridization between crops and weeds in the field.

As with *Hd1*, we were interested in seeing if there was a difference in phenotype between individuals with Type A and Type B *Hd3a* promoter types. Again, due to small samples sizes and the removal of intermediate promoter types, we pooled weeds and cultivars. When weeds and cultivars are pooled, there is no significant difference in days to flowering between individuals with Type A *Hd3a* promoter and those with Type B promoters (Figure 4.4; P = 0.05426). However, the Type A do show a trend towards later flowering as has been shown in cultivated rice. Given that both weedy rice groups share *Hd1* and *Hd3a* haplotypes with their progenitors, other genes must be involved in the drastic differences in phenotype between the weeds and their progenitors.

4.4 Conclusions/Future Directions

Our phenotypic data shows that flowering time differs greatly between weedy rice groups and between weedy rice and their cultivated progenitors (Figure 4.2). Flowering is highly variable in weedy rice and multiple flowering strategies might contribute to the success of weeds as a whole. Flowering time also seems to have evolved rapidly since weedy rice's divergence from cultivated ancestors. In the case of SH weeds, little overlap of flowering with the crop may occur, thus reducing the chance for hybridization. BHA weeds, however, may overlap more with the flowering time of the local *tropical japonica* crop, permitting hybridization and the potential escape of transgenes if GM crops are grown (Shivrain *et al.*, 2010a).

Within weedy rice there is polymorphism in the genes known to affect flowering time diversity in cultivated rice. SH weeds flower earlier and do not posses the 2 bp deletion in *Hd1*. BHA weeds flower later and do posses the 2 bp deletion in *Hd1* (Figures

4.3 and 4.4). There is diversity in *Hd3a* promoters in US weedy rice yet most SH and BHA weeds share haplotypes with progenitors that may be weakly associated with flowering time (Figures 4.4 and 4.5). Since both weed groups share similar haplotypes at two major flowering time loci with progenitors yet flower at significantly different times, other genes must be involved in flowering time divergence. One of the most promising candidates is *Ehd1*, a B-type response regulator, whose expression level has been found to be an important flowering time regulator in cultivated rice (Takahashi *et al.*, 2009). Since the expression level of *Ehd1*, not sequence variation in either the promoter or the coding regions of this gene, correlates the best with flowering time differences we did not pursue this gene for this study. Future work quantifying the expression levels of *Ehd1* and other flowering pathway genes in weedy rice under SD, LD and day neutral conditions might be useful for understanding how each of these genes works to regulate flowering in the weeds. Additionally, a QTL study conducted in our lab identified at least one region on chromosome 8 that is involved in flowering time differences between SH and BHA weeds and an *indica* cultivar (see Chapter 5). Within this region is a very promising candidate gene (*Ghd8/DTH8/qHY8*), encoding for a putative histone-like CCAAT-box binding transcription factor (Wei *et al.*, 2010; Yan *et al.*, 2011; Cai *et al.*, 2011). This gene may function as a regulator of *Ehd1* and *Hd3a* downstream of *Hd1*, with nonfunctional alleles conferring weaker photoperiod sensitivity (Wei *et al.*, 2010; Yan *et al.*, 2011; Cai *et al.*, 2011).

Table 4.1. List of Accessions used for this study. Accessions are grouped by type (weed, wild or cultivar). Identification numbers as well as genotypes at *Hd1* and *Hd3a* are listed along with phenotypic values for heading date.

Group	Study ID [a]	USDA ID/ Common Name [b]	IRGC/RA/ GRIN [c]	Origin [d]	Hd1 Haplotype	Functionality [e]	Hd3a Haplotype	Type (A/B) [f]	Heading date (days) [g]
Weedy rice	SH_1A04	**1091-01**	**x**	AR	2	f	10	B	66.7
	SH_1A07	**1098-01**	**x**	MO	2	f	6	B	56.5
	SH_1A08	**1134-01**	**x**	AR	2	f	x	x	65
	SH_1A09	**1135-01**	**x**	AR	2	f	x	x	59.5
	SH_1A10	**1141-01**	**x**	AR	2	f	6	B	54.5
	SH_1A11	**1160-01**	**x**	AR	2	f	6	B	61
	SH_1A12	**1179-01**	**x**	LA	2	f	6	B	60
	SH_1B05	1995-15	**x**	AR	2	f	1	A	x
	SH_1B03	16B	**x**	AR	2	f	x	x	111
	SH_1B07	1996-05	**x**	MS	2	f	x	x	68
	SH_1C02	**1001-01**	**x**	AR	2	f	20	?	66.5
	SH_1C03	**1002-02**	**x**	AR	5	f	6	B	63.5
	SH_1C06	**1047-01**	**x**	LA	2	f	x	x	65.5
	SH_1C07	**1073-02**	**x**	MO	2	f	6	B	57.5
	SH_1C10	**1190-01**	**x**	LA	7	nf 36bp	6	B	68
	SH_1C11	**1199-01**	**x**	MO	7	nf 36bp	6	B	73.5
	SH_1D01	**1344-02**	**x**	MO	2	f	6	B	67.5
	SH_1D06	1995-12	**x**	LA	2	f	6	B	67
	SH_1D09	1996-08	**x**	MS	2	f	6	B	68
	SH_1E03	**1210-02**	**x**	MO	2	f	6	B	78
	SH_1E07	**1333-02**	**x**	MO	2	f	6	B	66
	SH_1A01	**1004-01**	**x**	MO	2	f	19	?	61
	SH_1E05	**1163-01**	**x**	LA	2	f	1	A	78.5
	SH_1A06	**1196-01**	**x**	AR	2	f	6	B	58
	MXSH_1B06	1996-01	**x**	AR	4	nf 4bp	14	?	111
	MXSH_1D10	2002-51	**x**	AR	x	x	1	A	x
	MXBH_1E10	**2002-2-p21**	**x**	AR	11	nf 2bp	1	A	85
	MXBH1D11	2004-1-A	**x**	AR	x	x	x	x	110
	MX_1B10	MS4R788_93	**x**	MS	3	nf 2bp	2	B	123

Table 4.1 cont.	BRH_1C12	**1300-02**	**x**	MO	2	f	2	B	95
	BRH_1D12	**1183-01**	**x**	AR	2	f	6	B	56
	BRH_1C09	**1111-01**	**x**	AR	2	f	6	B	68
	BRH_1C08	**1092-02**	**x**	MS	2	f	2	B	95
	BRH_1E06	**1120-02**	**x**	MO	2	f	6	B	70.5
	BHA2_1E04	**1214-02**	**x**	LA	8	nf 2bp	2	B	89
	BHA2_1B01	**1188-01**	**x**	MS	3	nf 2bp	9	?	129
	BHA2_1C01	TX4	**x**	TX	3	nf 2bp	11	?	107
	BHA2_1A02	**1025-01**	**x**	AR	3	nf 2bp	1	A	105
	BHA2_1A03	**1081-01**	**x**	AR	3	nf 2bp	x	x	117
	BHA2_1E02	**1107-01**	**x**	AR	x	x	x	x	112.5
	BHA2_1C05	**1042-01**	**x**	AR	x	x	18	?	117.5
	BHA2_1E08	**1202-02**	**x**	AR	2	f	1	A	58
	BHA1_1B11	StgB	**x**	AR	3	nf 2bp	x	x	125
	BHA1_1D07	1995-13	**x**	LA	3	nf 2bp	x	x	125
	BHA1_1B02	10A	**x**	AR	x	x	x	x	96.5
	BHA1_1B08	1996-09	**x**	MS	3	nf 2bp	17	?	115.5
	BHA1_1E09	**2002-2-p1**	**x**	AR	x	x	1	A	113
	BHA1_1D08	1995-14	**x**	LA	3	nf 2bp	7	?	124
	BHA1_1B09	LA3	**x**	LA	3	nf 2bp	9	?	153
	BHA1_1B12	StgS	**x**	AR	3	nf 2bp	13	?	130.5
	BHA1_1D05	PrCoTall_3	**x**	AR	3	nf 2bp	2	B	159
	BHA1_1E01	**1166-02**	**x**	MS	3	nf 2bp	12	?	113
	BHA1_1A05	**1096-01**	**x**	AR	3	nf 2bp	x	x	101.5
	BHA1_1B04	18A	**x**	AR	3	nf 2bp	x	x	118
	BHA1_1C04	**1005-02**	**x**	AR	6	nf 2bp	2	B	104.5
	BHA1_1D02	PrCoSrt_1	**x**	AR	3	nf 2bp	2	B	118
	BHA1_1D03	PrCoTall_1	**x**	AR	3	nf 2bp	x	x	x
	BHA1_1D04	PrCoTall_2	**x**	AR	3	nf 2bp	2	B	128
Cultivated rice *aromatic*	3A01	DOM-SOFID	RA4929/ 12880/ PI584607	Iran	12	f	x	x	x

Table 4.1 cont.									
	3A02	ARC-13829	RA48947/ 42469	India	12	f	x	x	x
	2B01	DA 13	x	Bangladesh	12	f	x	x	x
aus	2B03	Aus 196	29016	Bangladesh	13	nf 2bp	2	B	81.5
	3A03	JHONA-349	RA4979/ 6307	India	3	nf 2bp	2	B	91
	3A04	KASALATH	RA5339	India	3	nf 2bp	11	?	92.7
	3A05	DV85	RA5323/ 8839	Bangladesh	3	nf 2bp	x	x	89
	3A06	BJ-1	RA5345/ 45195	India	3	nf 2bp	x	x	104
	3A07	DHALA_ SHAITTA	RA5361/ PI180060	Bangladesh	3	nf 2bp	8	B	x
indica	2B02	Bei Khe	22739	Cambodia	2	f	3	B	81
	3A08	RATHUWEE	RA4911/8952/ PI584605	Sri Lanka	3	nf 2bp	6	B	103
	3A09	Khao Dawk Mali -105	RA4878/ 27748	Thailand	42	f	x	x	104.5
	3A10	LalAman	RA4956/ 46202	India	15	f	x	x	123.5
	3A11	Dholi Boro	RA4984/ 27513	Bangladesh	15	f	x	x	125
	3A12	Ai-chiao-hong	RA4967/ 51250/ PI584576	China	2	f	4	A	100
	3B01	CHAU	RA4974/ 56036	Vietnam	35	f	x	x	x
	3B02	CHHOTE_ DHAN	RA4978/ 58930	Nepal	15	f	10	B	87.5
	3C05	DEE_GEO_ WOO_GEN	RA5344/ PI279131	Taiwan	4	nf 4 bp	x	x	108.5
	3B03	POPOT-165	RA4987/ 43545	Indonesia	12	f	x	x	134

Table 4.1 cont.

temperate japonica	3B04	Ta hung ku	RA5029/1107	China	x	x	1	A	x
	3B05	Kamenoo	RA4990/ PI403629	Japan	43	nf 43bp	1	A	x
	3B06	Nep-Hoa-Vang	RA4945/ 40748	Vietnam	12	f	4	A	x
	3B07	SHOEMED	RA5123/ PI392539	Philippines	1	nf 36bp	x	x	x
tropical japonica	3B08	Khao Hawm	RA5030/ 24225	Thailand	3	nf 2bp	1	A	156.5
	3B09	Mirti	RA4970/ 25901/ PI584553	Bangladesh	44	nf 2bp	x	x	83
	3B10	KU115	RA5294/ PI597044	Thailand	3	nf 2bp	x	x	101.5
	3B11	CICIH_BETON	RA4955/ 43372	Indonesia (Bali)	45	f	4	A	x
	3B12	GOTAK_ GATIK	RA4959/ 43397/ PI584572	Indonesia (C. Java)	2	f	21	B	93
	3C01	ASSE_Y_ PUNG	RA5333/ CIor461	Philippines	46	nf 2bp	1	A	127
	3C02	Kotobuki Mochi	RA4882/2545	Japan	3	nf 2bp	5	A	75
	3C03	TREMBESE	RA4988/ 43675	Indonesia (East Java)	9	nf 2bp	1	A	x
US cultivars	3C04	LEMONT	RA4998/ 66756	US	10	nf 2bp	1	A	x
	2A01	BENGAL	PI561735	US	3	nf 2bp	4	A	x
	2A02	BLUE_ROSE	CIor1962	US	3	nf 2bp	x	x	96.5
	2A03	CAROLINA_ GOLD	CIor1645	US	3	nf 2bp	x	x	x
	2A04	CL121	n/a	US	3	nf 2bp	x	x	x

Table 4.1 cont.	2A05	CL161	n/a	US	3	nf 2bp	4	A	x
	2A06	CYPRESS	PI561734	US	x	x	4	A	x
	2A07	DELITUS	CIor1206	US	x	x	x	x	x
	2A08	DREW	n/a	US	3	nf 2bp	1	A	x
	2A09	EDITH	CIor2127	US	3	nf 2bp	4	A	100.5
	2A10	PALMYRA	CIor9463	US	3	nf 2bp	1	A	x
	2A11	REXORO	CIor1779	US	3	nf 2bp	4	A	x
	2A12	ZENITH	CIor7787	US	3	nf 2bp	1	A	x
Wild Asian rice and outgroups									
O. rufipogon	3C06	x	106163	Laos	47	f	23	?	x
	2C01	x	81990	Myanmar	14	f	x	x	x
	2C02	x	100588	Taiwan	15	f	x	x	x
	2C03	x	100904	Thailand	16	f	x	x	x
	2C04	x	100916	China	17	nf 20bp	x	x	64.5
	2C05	x	104501	India	19	f	x	x	x
	2C06	x	104599	Sri Lanka	20	f	x	x	x
	2C07	x	104624	China	21	f	x	x	x
	2C08	x	104714	Thailand	22	f	16	?	x
	2C09	x	104833	Thailand	23	f	x	x	98
	2C10	x	104871	Thailand	24	f	x	x	123.5
	2C11	x	105388	Thailand	25	f	x	x	x
	2C12	x	105491	Malasia	3	nf 2bp	x	x	174
	2D01	x	105568	Philippines	26	f	x	x	106.5
	2D02	x	105711	India	27	nf 20bp	x	x	154
	2D03	x	105720	Cambodia	28	f	x	x	x
	2D04	x	105855	Thailand	29	f	x	x	x
	2D05	x	105888	Bangladesh	30	f	x	x	141
	2D06	x	106086	India	2	f	x	x	x
	2D07	x	106103	India	31	f	x	x	x
	2D08	x	106122	India	32	f	6	B	x
	2D09	x	106134	India	33	f	x	x	x
	2D10	x	106150	Laos	x	x	x	x	x

Table 4.1 cont.	2D11	x	106168	Vietnam	18	nf 20bp	x	x	x
	2D12	x	106169	Vietnam	3	nf 2bp	6	B	x
	2E01	x	106321	Cambodia	34	f	x	x	x
	2E02	x	106346	Myanmar	35	f	24	?	x
	2E03	x	106453	Indonesia	16	f	17	?	x
	2E04	x	106518	Vietnam	16	f	22	?	x
	2E05	x	106523	Papua New Guinea	36	f	x	x	x
O. nivara	2F01	x	86662	Thailand	37	nf 1bp	x	x	x
	2F02	x	103821	China	3	nf 2bp	25	?	x
O. barthii	2F03	x	101226	Mali	38	f	x	x	x
	2F04	x	104081	Nigeria	39	f	x	x	x
O. glaberrima	2F05	x	86779	Liberia	39	f	x	x	x
	2F06	x	100983	Nigeria	39	f	x	x	x
	2F07	x	101855	Burkina Faso	39	f	x	x	x
	2F08	x	102410	Mali	39	f	x	x	x
O. glumaepatula	2F09	x	105561	Colombia	40	nf 7bp	x	x	x
	2F10	x	105670	Brazil	41	f	x	x	x
O. meridionalis	2F11	x	93261	Indonesia	48	f	x	x	x
	3C07	x	101148	Australia	49	f	x	x	x

a based on STRUCTURE and identity from Reagon et al 2010

b accessions with RA numbers were acquired from Susan McCouch while all others were acquired from IRRI

c Accessions in bold were selfed 4 generations at the USDA stock center

d Origin for weeds is a US state abbreviation, Origins for cultivated and wild rice is country

e Functionality is based on previous characterization of Hd1 alleles by CITE.

f Type is based on previous characterization of hd3a promoters by CITE, ? represent haplotypes sharing characteristics of both A and B promoters.

g Heading date was measured in days from date of germination to 50% emergence of first panicle

x-- no data available

Table 4.2. List of Primer Sequences and their location. Primers are grouped by gene (*Hd1* and *Hd3a*) as well as genetic versus flanking region.

***Hd1* Gene Primers**			
Primer ID	**Forward**	**Reverse**	**Region**
Hd1_003	CGCGATTGCTTTGGCTTAAAAGG	CTCCTCCTTCGCCTCCAACTCC	5'UTR/Exon1
Hd1_004	CTACTACCACAAGCAAGGCTACTTCG	CCCGACAAGATCAAAGTACTCATCG	Exon1
Hd1_005	GGTGGACTCTTGGCTTCTCCTCTCC	ATCAGCCTAAAGATCGCAGCTTGG	Exon1/Intron
Hd1_006	CTCCAAACAAAAGCTACTGTCTAAC	GCATATCTATCACCGTGCTGTC	Intron/Exon2
Hd1_007	GGAGGCGGGTATAGTACCAGACAG	CATTTCATCTCATCACTGCTCTTTGC	Exon2/3'UTR
Hd1_012	AGGGACAGTCCTAAGTTAAAATGG	CACATGGACAAGTCTATCGATCC	Exon1
Hd1_013	AGGTGGACTCTTGGCTTCTCCTC	TGGAGAATGTTTTTCCACTCAAAAG	Exon1/Intron
Hd1_014	CACAATTGATCTTTAGGCAGACCAG	CTGTCCATGGAGCTGAAGTGAAG	Intron/Exon2

***Hd3a* Promoter Primers**		
Primer ID	**Forward**	**Reverse**
Hd3ap_001	TCATTAATTGCCTTACCTCAAC	TTGTTGGTTTTTCCGCAACT
Hd3ap_002	GCATCCACAAAATTTCTAGGC	CGATCTTGCAAAAAACCCTG
Hd3ap_003	CGCCGACATAGAAAGGAAAG	GTTAGGGTCACTTGGGCTTG

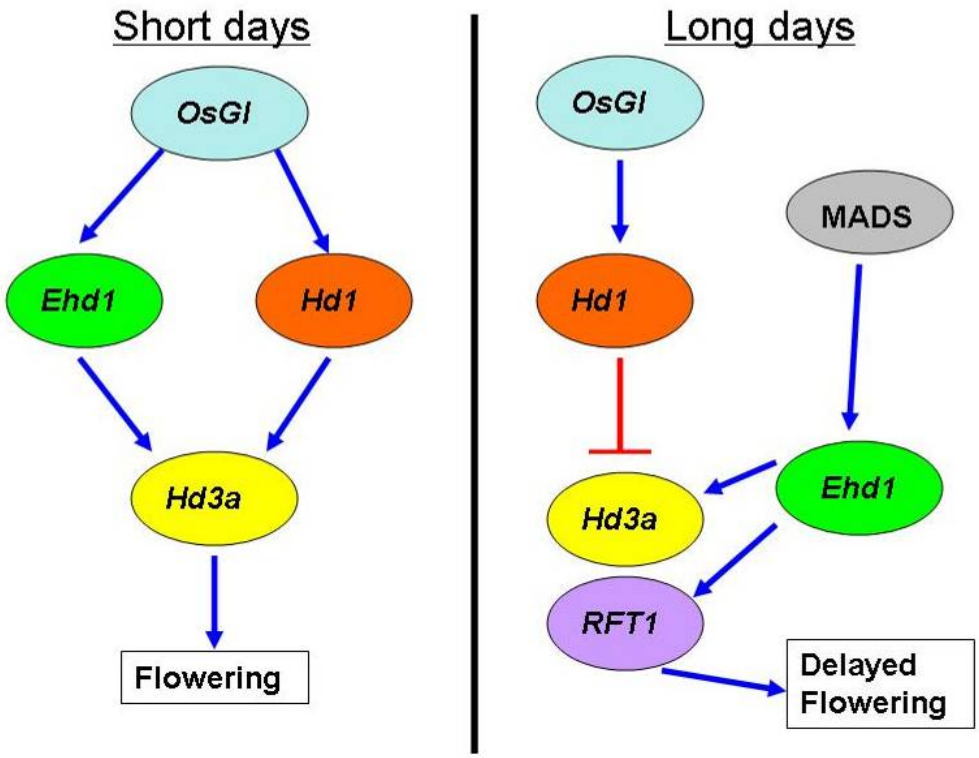


Figure 4.1. Brief overview of flowering time gene regulation in rice.
This figure is adapted from Tsuji, Taoka and Shimamoto (2010). Under short days *OsGI*, ortholog of *Arabidopsis GIGANTEA*, activates both *Hd1*, the ortholog of *Arabidopsis CONSTANS*, and *Ehd1*, for which there is no *Arabidopsis* ortholog. Both genes go on to activate *Hd3a*, the ortholog of *Arabidopsis FT*, promoting flowering. Under long days, several MADS-box transcription factors act on *Ehd1* to promote flowering through activation of *Hd3a* and *RFT1*, a close paralog of *Hd3a*. However, concurrent with this *OsGI* activates *Hd1* which then acts negatively on *Hd3a* to delay flowering.

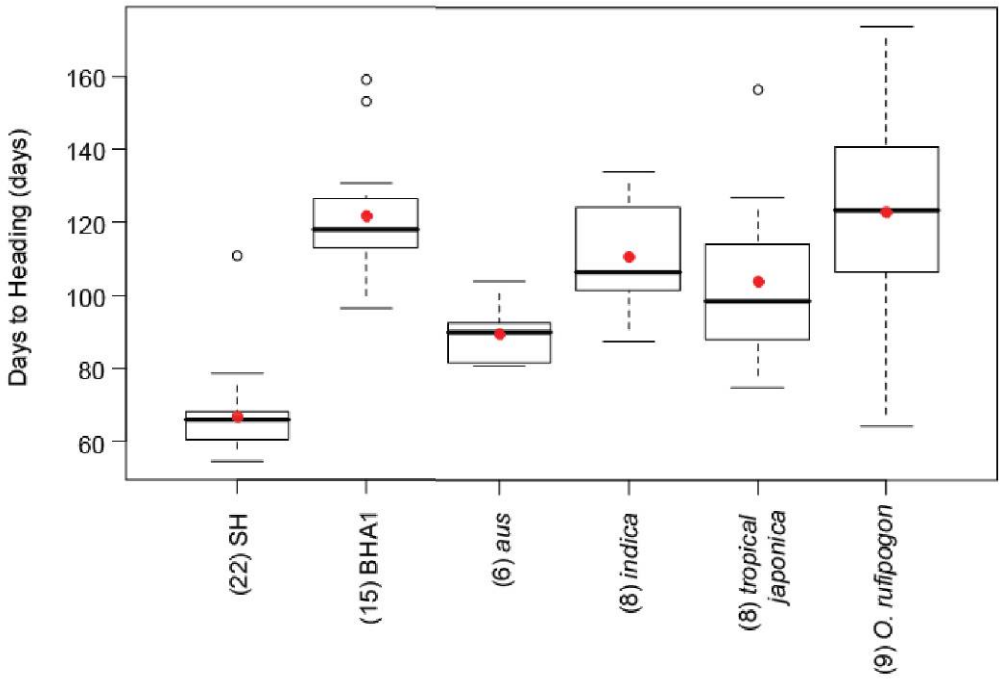


Figure 4.2. Flowering time phenotype in weedy, cultivated and wild *Oryza*.
Flowering time, also referred to as "Days to Heading" or "Heading Date", was averaged across two individuals per accession and a boxplot distribution of those averages is shown here. Black line is median, red dot is mean and white dots represent outliers. Numbers in parenthesis represent sample sizes. Weedy rice groups are as follows: SH (straw-hulled), BHA1 (black hulled and awned).

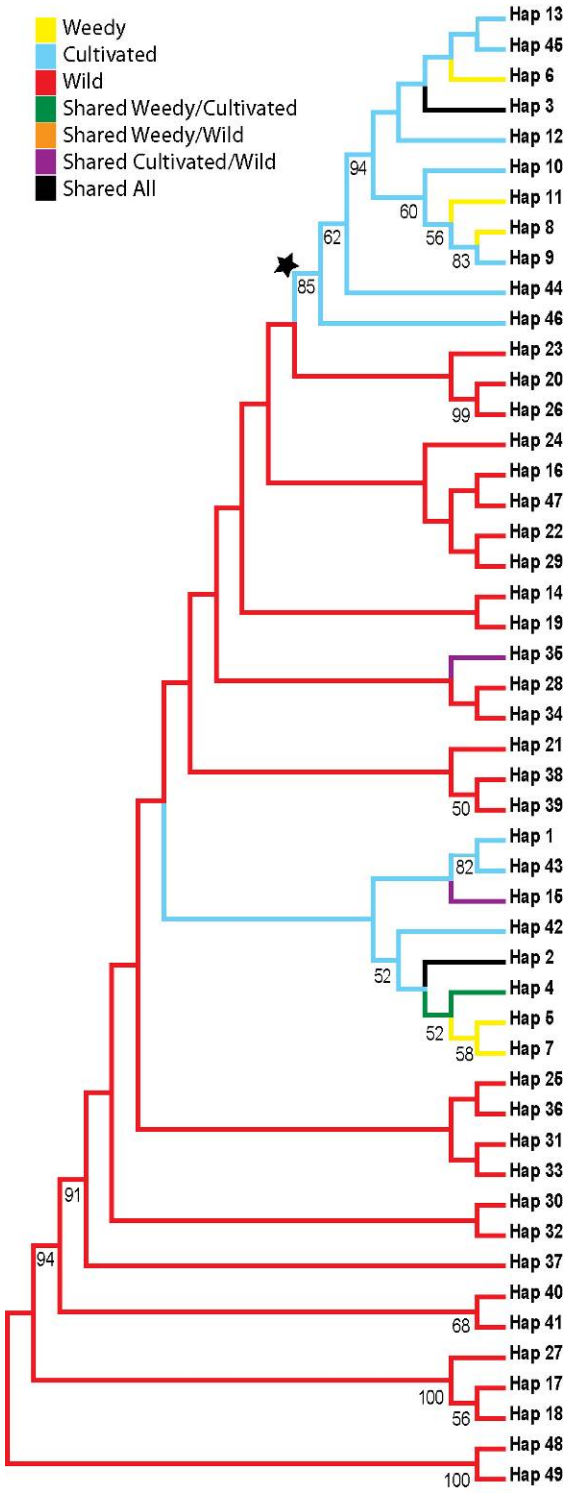
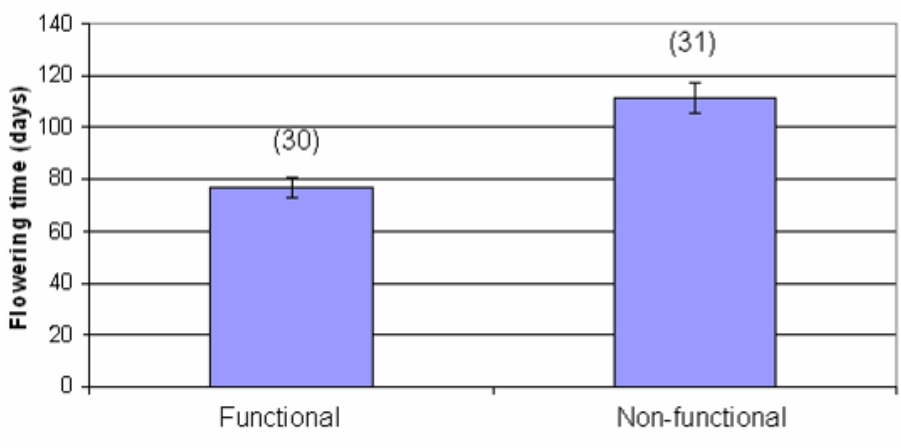


Figure 4.3. Neighbor joining tree for *Hd1* coding region haplotypes.
Numbers below branches correspond to bootstraps using 1000 replicates. The clade marked with a star contains haplotypes that share a suite of ~14 SNPs and a causative two base pair deletion in exon two that truncates the HD1 protein at the C terminal end, leaving the protein non-functional. Haplotypes are color coded by the key on the top left.

A



B

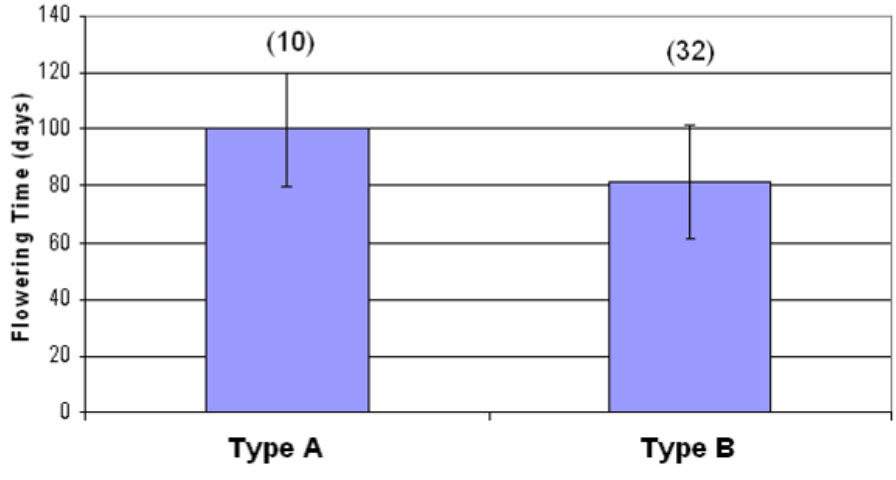


Figure 4.4. Phenotypic differences between major haplotype classes.
A) Differences in heading dates between functional and nonfunctional *Hd1* haplotypes in cultivated and weedy rice. Differences are statistically significant (P = 5.62 e -8). B) Differences in heading dates between Type A and Type B *Hd3a* promoters in cultivated and weedy rice. Differences are not statistically significant (P = 0.05426). In both panels, numbers in parenthesis denote sample size of combined weedy and cultivated individuals.

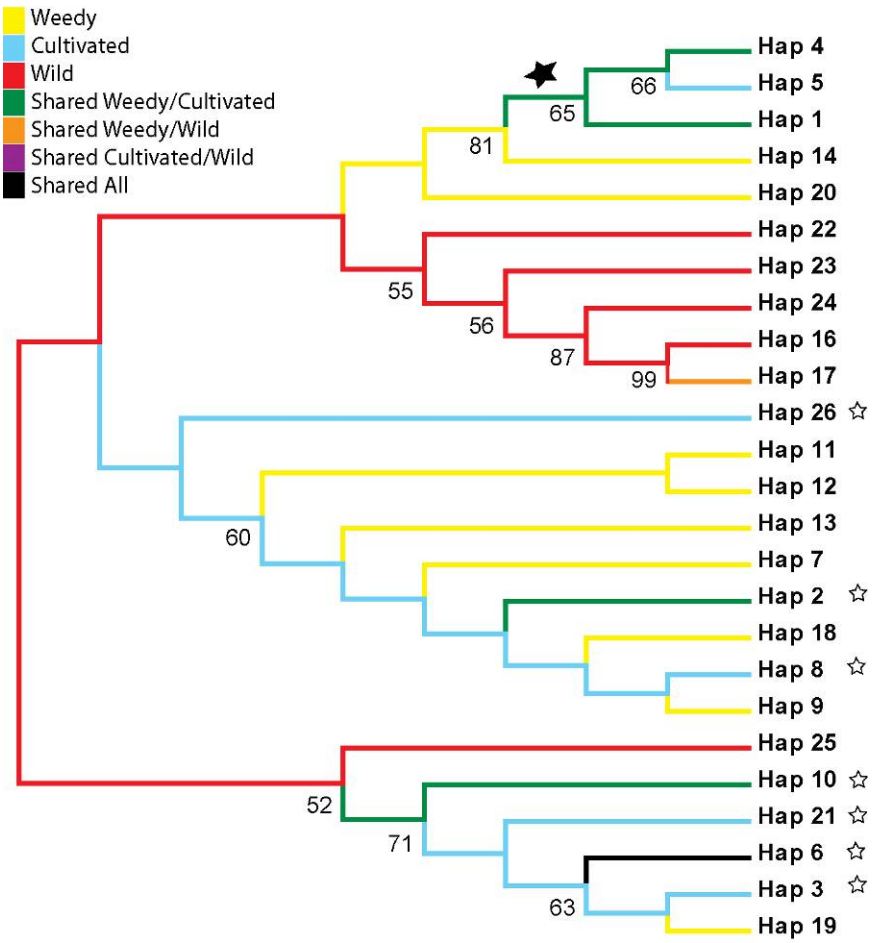


Figure 4.5. Neighbor joining tree for *Hd3a* promoter haplotypes.
We sequenced ~1kb of promoter region in 84 accessions of wild, weedy, and cultivated rice. The Type A promoters (black star) are distinct from the Type B promoters (white stars) as classified by (Takahashi *et al.*, 2009). Type A promoters typically show lower expression of the *Hd3a* gene compared to Type B promoters. The inclusion of weedy rice and additional wild samples brings in new recombinant haplotypes which have yet to be classified as A or B types and might even be intermediate in expression level.

CHAPTER 5

SIMILAR TRAITS, DIFFERENT GENES/ DIFFERENT TRAITS, SIMILAR GENES: EXAMINING PARALLEL EVOLUTION IN RELATED WEEDY RICE POPULATIONS

5.1 Introduction

The repeatability of adaptive evolution is an outstanding question in biology. The presence of similar traits in independently evolved lineages has often been documented (e.g. Schluter *et al.*, 2004), and it has recently become possible to determine the extent to which this is a result of similar changes in shared genetic systems (Stinchcombe & Hoekstra, 2008). Shared genetic biases among taxa that could result in disproportionate use of the same genes are often invoked to explain the occurrence of trait convergence (e.g. Hodin, 2000; Schluter *et al.*, 2004). These biases have been traditionally believed to be more likely among closely related species, suggesting that convergent phenotypic evolution among relatives is more likely attributable to shared genetic mechanisms (e.g. Arendt & Reznick, 2007). To date, however, studies have revealed that the genetic bases of convergent phenotypes can range from similar to different genetic changes, both in closely and distantly related taxa (e.g. Kingsley *et al.*, 2009; Rompler *et al.*, 2006; Steiner *et al.*, 2008; Yoon & Baum, 2004). Because patterns have been slow to emerge, the extent to and circumstances under which convergent phenotypic evolution is due to shared genetic mechanisms is currently an active area of inquiry.

Plants evolving in the agricultural environment offer many examples of convergent phenotypic evolution. For example, although domesticated in different world regions, many cultivated grasses have experienced similar selective pressures by humans;

crop grasses have been selected for alterations in seed traits, annual life cycles, increased selfing and decreased seed shedding (Purugganan & Fuller, 2009). Similarly, trait convergence is often also evident in agricultural weeds - highly competitive plants that repeatedly invade the disturbed cropland soils (Basu *et al.*, 2004) and belong to a wide variety of genera. Despite sometimes being unrelated, agricultural weeds often converge on similar adaptive traits such as rapid growth, high seed production, increased seed dispersal and deep roots (Ellstrand *et al.*, 2010; Harlan & DeWet, 1965). Little is currently known about the genetics underlying the evolution of these so-called 'weedy' traits, but the preponderance and diversity of agricultural weeds makes these ideal systems for studies of the genetic basis of convergent evolution.

Rice fields worldwide are often invaded by a weedy type of rice known as weedy or red rice (*Oryza sativa* L) (Olsen *et al.*, 2007). Weedy rice is a major agricultural pest, as it is an aggressive competitor that spreads rapidly and drives down the quality of the rice harvest. Moreover, because it is closely related to the crop it invades, weedy rice is difficult to detect in rice fields in its early growth stages and hard to control with herbicides (Vaughan *et al.*, 2001). While limited, studies of weedy rice in various world regions have indicated that weedy rice populations often resemble the local predominant crop variety (see Olsen *et al.*, 2007) suggesting repeated independent origins of weedy rice populations or contributions of local groups to the genetic make-up of weedy rice. The presence of multiple populations of weedy rice around the world and their convergence on some typical weed-adaptive traits offer a unique opportunity to the study of parallel evolution at various geographic scales.

In the United States, where over 30% rice fields are infested with weedy rice (Shivrain *et al.,* 2009), our work and that of others has shown that two major independently derived and morphologically and genetically differentiated populations of weedy rice co-occur. The Straw Hulled (SH) group most closely resembles cultivated rice with straw-colored hulls and slightly larger grains; the Black Hulled & Awned (BHA) group often resembles the wild rice, with black or brown colored hulls, small grains and long awns (Londo & Schaal, 2007; Reagon *et al.,* 2010; Vaughan *et al.,* 2001). Genome-wide assessments of polymorphism indicate that SH and BHA weedy populations are more closely related to *indica* and *aus* variety groups of domesticated Asian rice, respectively, than to other major cultivated or wild *Oryzas* (Londo & Schaal, 2007; Reagon *et al.,* 2010) (Figure 5.1). Although there is debate over exactly how many times Asian rice was domesticated, it is well accepted that cultivated rice was domesticated from Asian wild rice (*Oryza rufipogon/Oryza nivara*), with subsequent diversification of variety groups. Cultivated rice varieties are thus genetically differentiated (Caicedo *et al.,* 2007; Garris *et al.,* 2005), and the *aus* and *indica* putative ancestors of US weedy rice groups are distinct from the *japonica* cultivars grown in the US (Figure 5.1). The origins of US weedy rice from crop ancestors suggests that the evolution of weedy traits in these groups could be a process of "de-domestication," whereby selection favors reversions of domestication traits to forms characterizing wild species. This in turn suggests a different level at which parallel genetic evolution in weeds could be occurring: convergence of weedy and wild traits could be acquired through mutations in the same genes that were targeted during domestication.

US weedy rice exhibits many traits that are associated with the persistence of weeds, such as increased seed dormancy and seed shattering, faster growth, taller plants, and modified flowering times (Delouche *et al.,* 2007; Shivrain *et al.,* 2010a). We have previously characterized some of these traits in US weedy rice populations relative to their putative cultivated progenitors and have noted different degrees of phenotypic convergence. For example, weedy rice from both SH and BHA are highly prone to shattering both in the field and under controlled environmental conditions, a trait that is absent in the domesticated progenitors (Noldin *et al.,* 1999; Thurber *et al.,* 2010) (Table 5.4). Likewise, higher growth rates have been observed for SH and subpopulation of BHA compared to their ancestors (Reagon *et al.,* 2010). In contrast, flowering time (i.e. heading date) is strongly in differentiated in both weed groups compared to their cultivated progenitors, but the shifts are in opposite directions: under day neutral conditions, SH flowers significantly earlier than *indica*, whereas BHA groups flower significantly later than *aus* (Reagon *et al.,* 2011) (Table 5.4). In field conditions, blackhull weeds also typically flower later than strawhull weeds (Shivrain *et al.,* 2009). Thus, although the same trait has been affected in the course of weed evolution, there has not been convergence on a single phenotypic value. A similar situation is seen for plant height. Weedy rice shows a range of plant heights, but under growth chamber conditions SH weeds are generally shorter than their *indica* progenitors, and BHA weeds are generally taller than the *aus* (Reagon *et al.,* 2011) (Table 5.4). In the field, both weed groups tend to be taller than the local *japonica* crop (Shivrain *et al.,* 2009), likely driven by the recent selection for semi-dwarf high yielding rice plants since the green revolution (see citations in Asano *et al.,* 2007). Remarkably, these divergent weedy phenotypes have

evolved under near identical selective pressures, as weedy rice from both populations are often found in the same rice field during a single growing season (Shivrain *et al.,* 2010a). Given the convergence of phenotypic values for some traits and divergent evolution for others, we are interested in determining to what extent common genes underlie weedy trait evolution in US weedy rice groups. We hypothesized that, given the weeds' origins from cultivated ancestors sharing a domestication origin (Figure 5.1), US weedy rice groups are likely to have shared biases leading to mutations in the same genes underlying convergent weedy traits. We also hypothesized that shared ancestral pathways could lead to different mutations in the similar genes underlying divergent weedy traits. To test for parallel genetic evolution, we attempted crosses between US weeds and their putative progenitors to capture the genetic differences that have accumulated since each weed group diverged from a cultivated background (Figure 5.1). Using F_2 populations, we carried out QTL mapping of four quantitative traits that have either converged or diverged between weedy rice groups. We also carried out mapping of two qualitative traits specific to the BHA weed group, to see if underlying genomic regions overlapped with loci involved in these phenotypes in wild rice. Our goal was not to identify causal genes, as this cannot be done with an F_2 population, but to begin assessment of the degree to which shared genomic regions underlie weedy traits in both groups. We find that, in most cases, parallel genomic regions are not involved in traits characterizing weedy rice groups; the exception to this is flowering time, which, though divergent among groups, may involve modification of alternative alleles at a single locus involved in the *Oryza* flowering time pathway.

5.2 Methods

5.2.1 Plant Materials

We created two mapping populations (S and B) by crossing two weedy rice individuals from distinct populations (SH-RR09 and BHA-RR20) with a single *indica* cultivar Dee Geo Woo Gen (DGWG), which in both cases was the pollen donor. The weedy rice parents are representatives of the SH and BHA populations of US weedy rice as determined by population structure assessments (Reagon *et al.,* 2010). The *indica* cultivar group was chosen as a parent because this group is putatively ancestral to the SH weed group and is closely related to the BHA ancestor, *aus* (Caicedo & Purugganan, 2005; Garris *et al.,* 2005; Reagon *et al.,* 2010). Multiple attempts to cross BHA weeds with *aus* cultivars failed, thus the closely related *indica* parent was selected. Weed and crop parents were selected to maximize phenotypic differences in potential weed-adaptive traits based on previous growth chamber data (Table 5.1). The resulting F_1 plants largely showed phenotypes intermediate between the two parents; the single exception was for the B population F_1, which produced seeds with black hulls and awns suggesting that these traits are controlled by few genes in which the BHA allele is dominant. F_1 plants were confirmed to be the result of crosses and were allowed to self fertilize to create the F_2 seeds used for mapping.

Approximately 250 F_2 seeds per population, offspring from a single F_1 for each cross, were sown in a greenhouse in Amherst, MA on April 1st 2010 in four-inch pots set in two-inch trays of ten pots each. Approximately twenty-five trays per population were distributed randomly throughout the greenhouse. Seeds were heat treated for twelve hours at 37^{o} Celsius and the hulls were partially or totally removed prior to planting to

eliminate dormancy. Three replicate pots of each parental line were also sown in a single tray in the greenhouse to serve as phenotypic controls. Water was maintained at a height of approximately one to two inches in trays to keep soil moist, and fertilizer was applied as described in (Reagon *et al.*, 2010). The emergence date of F_2 seedlings was not uniform within each population despite dormancy releasing treatment. Due to inadequate F_2 germination, two additional waves of planting were performed with ~100 seeds on April 22nd 2010 and May 13th 2010. Since different planting dates likely put seedlings under different light and temperature environments and could affect trait values, we compared trait distributions and averages for heading date, plant height and seed shattering across planting waves. No differences were observed for any trait (data not shown). Additionally, no differences were observed in the QTL detected using only the first wave individuals and the full dataset; thus we decided to use all three planting waves to increase our statistical power. In the end, 184 S population and 159 B population individuals were usable for QTL mapping, having phenotypes for all traits evaluated and genotypes at all markers (see below). We considered these sizes sufficient, as a minimum of 150 individuals has been found a good baseline when not carrying out fine mapping (Bernardo, 2008), and QTL that explain as little as 5% of the variance can be detected with samples of a few hundred F_2 (Flint & Mott, 2001).

5.2.2 Trait Evaluation

Four quantitative traits were evaluated in each F_2 population. These included Heading Date (HD), Plant Height (PH), Growth Rate (GR) and Seed Shattering (SS). Additionally, two qualitative traits, Hull Color (HC) and Awn Presence (AW), were evaluated in the B population, as these traits did not differ between S population parents.

HD was measured in days from the date of seedling emergence until the first panicle had emerged halfway from the boot. Panicles were bagged at this stage to ensure selfing and prevent loss of seeds. PH was measured in centimeters, at heading, from the base of the plant at the soil to the tip of the tallest panicle excluding any awn. GR was calculated by dividing PH by HD to get an average rate in cm/day. SS was measured in grams of force required to remove the seed from the panicle; measurements were taken from ten mature seeds collected thirty days after heading from a single panicle, where possible, and averaged per individual, as described in (Thurber *et al.*, 2010). HC was scored as straw (0) and black (1) on seeds collected thirty days after heading. A small number (<20) of individuals showed brown or gold hull colors and were not considered for analysis. AW was recorded as presence (1) vs. absence (0) at the same time HC was scored. Phenotypic data for all individuals can be found in Tables 5.7 and 5.8.

Broad-sense heritability (H^2) for each trait in each population was calculated as in (Xu *et al.*, 2009). Briefly, the average of the parental variances grown in the greenhouse environment was used as the environmental variance (V_e). V_e was subtracted from the total phenotypic variance of the F_2 population (V_p) to obtain the genetic variance (V_g). H^2 was then calculated as V_g / V_p for each trait and is reported in Table 5.5.

5.2.3 Marker Analysis

DNA was extracted from frozen tissue collected from greenhouse grown F_2 plants using a CTAB method (Reagon *et al.*, 2010). Over 188 microsatellite (SSR) markers from previously published studies (e.g. Chen *et al.*, 1997; McCouch *et al.*, 2002) were genotyped in the three parental lines. SSR markers are identified as numbers that correspond to the "RM" markers from previous studies. Additionally, two and six

insertion-deletion (indel) markers were adapted from (Shen *et al.*, 2004) for the S and B populations, respectively. These were given names R#M#. Lastly, due to inadequate coverage of chromosome 1 in the S population, an additional three indel markers were developed from whole genome sequence data (i4, i22, and i23) (Hyma & Caicedo, unpublished).

Indel and SSR markers were PCR amplified similar to (Panaud *et al.*, 1996) except that the reaction volume was reduced to 15 ul and PCR cycling conditions were as follows: 94°C for 5 min, followed by 35 cycles of 94°C for 30s, 55°C for 30s, and 68°C for 1 min; finished by 5 min at 72°C. Indel marker genotypes were directly scored from 2% agarose gels. Amplified SSR products were run on an ABI 3130XL genetic analyzer at the Genomics Resource Laboratory at the University of Massachusetts Amherst. FSA files were analyzed using the PeakScanner software to determine the sizes of bands. All marker genotypes were scored as 0, 1, or 2 depending on whether the individual was homozygous for the cultivated allele (0), heterozygous (1), or homozygous for the weedy parent allele (2). Marker segregation analysis was carried out using chi-square tests to detect significant distortion from Mendelian inheritance. Linkage maps were created using the Kosambi map function under default conditions in R/qtl, resulting in maps of ~1587 centimorgans (cM) for both populations. The average interval size is ~31.9 cM with a minimum of 2.9 cM and a maximum of 143.6 cM. Marker positions were found to be in similar locations as previously published maps for cultivated rice (e.g. Lee *et al.*, 2005; Thomson *et al.*, 2003). Marker genotypes for all individuals can be found in Supplementary File 1.

5.2.4 QTL Mapping

The normality of phenotypic data was checked using Normal Quantile Plots (Tan *et al.,* 2004). If the plot showed the data to be non-normal it was log transformed. If transformation was unsuccessful, non-parametric analysis was performed for that trait (Tilquin *et al.,* 2001). For normalized traits, QTLs were identified using Single Marker Analysis (SMA) and Composite Interval Mapping (CIM) in WinQTL Cartographer (Wang 2011). SMA was run under default conditions while CIM was run using forward-backward regression and a walk speed of 5 cM due to low marker density. Non-parametric analysis was performed on SS using a Kruskal-Wallis rank sum test (K-W test) in R/qtl. This method ranks the individuals by phenotypic value and then sorts them by genotypic value, one locus at a time. For both mapping methods, LOD scores over 2 were considered significant due to the low marker density (see Results)(Van Ooijen, 1999). The locations of the QTL identified in this study were compared to QTL previously published using the "QTL" search feature on Gramene (http://www.gramene.org/).

5.3 Results

5.3.1 Phenotypes

We compared phenotypes of the parents grown under greenhouse conditions with phenotypes previously obtained under growth chamber conditions (Table 5.1). The greenhouse environment differed from the growth chamber in having more variable temperatures and seasonally variable day-length. Phenotypic differences among environments were seen for some traits and some parental lines. Most strikingly, the SH-

RR09 weed parent had an increase of over 47 days in HD under the greenhouse environment. This is consistent with a photoperiod sensitive response, given the differences in day-length between the growth chamber (12 hours) and greenhouse (seasonally variable, but day-length consistently exceed 12 hours), and with the short day flowering behavior of many rice varieties (Yano *et al.*, 2000). In contrast, the heading date of the BHA-RR20 weed and the crop parent were consistent across environments, suggesting limited photoperiod sensitivity. For SS, the cultivated parent showed nearly half as much shattering resistance in the greenhouse compared to the growth chamber, while there was no sizable change in the weeds' shattering abilities. PH changed dramatically for the BHA-RR20 parent, which nearly doubled in plant height in the greenhouse, while the other two parents remained close to growth chamber values. On average, the SH-RR09 weed parent grew nearly twice as fast in the greenhouse than in the growth chamber while the BHA-RR20 parent grew nearly half as fast under greenhouse conditions, but both consistently exceeded the crop parent. Despite environmental influence, phenotypic differences between weed and crop parents were still evident for most traits under greenhouse conditions. The SH-RR09 weed and DGWG crop parent differed appreciably in PH, GR and SS. The BHA-RR20 weed and DGWG crop parent differed substantially in PH, HD, GR, SS, HC and AW.

We examined the phenotypic distributions of all traits in the F_2 populations (Figure 5.2). For the qualitative traits, segregation of HC fits the 3:1 ratio expected for a trait controlled by a single gene, while for AW there is an excess of weedy parent phenotypes suggesting the involvement of more than one locus ($\chi^2 = 9.358$, $P < 0.01$) (Table 5.2; Figure 5.2B). For quantitative traits, continuous, nearly normal distributions

were observed for HD, PH, and GR. The results of Normal Quantile Plots (NQP) could not reject that the $\log_{10}$ values for HD, PH, and GR from either population came from a normal distribution ($\alpha = 0.01$). In contrast, normality was rejected for SS in both populations and attempts to normalize this trait failed. Transgressive segregation was seen in all traits in both populations yet it was most noticeable for HD (Figure 5.2A). Trait means were fairly similar between the two populations for all four quantitative traits. The distributions of two traits (PH and GR) were nearly identical between populations suggesting similar genetic architectures. For both HD and SS in the S population, multiple peaks were observed suggesting a role for few major effect genes (Figure 5.2A). Interestingly, the B population distributions for HD and SS were more normal, suggesting multiple weaker effect loci and a different genetic architecture from the S population.

Correlations between traits were tested using Pearson's correlations. A weak but significant positive correlation was found between HD and PH in both populations (Table 5.6). Stronger positive correlations between plant height and heading date ($r = 0.467$ to $r = 0.76$) have been seen in other studies of both greenhouse and field grown cultivated rice plants (see Bres-Patry *et al.*, 2001; Lee *et al.*, 2005).

Broad sense heritability was calculated for all quantitative traits in each population separately (Table 5.5). Despite the evidence for environmental effects on some of these traits, heritabilities in our greenhouse environment were fairly high, ranging from ~67% for SS to nearly 100% for HD. Heritabilities were also remarkably similar across mapping populations.

5.3.2 Marker Linkage Maps

Of the 188 SSR tested in the polymorphism screen, about 31% were polymorphic between DGWG/SH-RR09 and about 45% were polymorphic between DGWG/BHA-RR20. These low levels of polymorphism are not unusual, as US weedy rice is likely descendant from cultivated rice from the *indica/aus* groups (Reagon *et al.,* 2010) (Figure 5.1). We expected to find less polymorphism between the SH weed and *indica* parent than the BHA weed and the same parent, as *indica* is putatively ancestral to SH weeds (Reagon *et al.,* 2010). Due to PCR failures, the total SSR markers used were 52 and 59 for the S and B populations respectively. With the addition of polymorphic indels, we mapped using 59 markers for the S population and 65 markers for the B population.

Segregation distortion at the $P < 0.01$ level was seen in twelve and ten markers in the S and B populations, respectively. The distorted markers consisted of ~15 to 20% of the total markers per population and about half of the markers had excess weedy alleles while the rest had excess cultivated alleles. Only two markers were distorted in both populations. Despite the low levels of distorted markers, we compared mapping both with and without these. The presence of QTL linked to distorted markers can affect their ability to be detected; yet distorted markers do not cause false positive associations and are not a problem if randomly distributed across a genetic map (Zhang *et al.,* 2010). No differences in the number or locations of significant QTL were found for any trait when segregation distorted markers were excluded.

5.3.3 Mapping Quantitative Traits

Seed Shattering (SS)

Two shattering QTL were identified in the S population using Kruskal-Wallis test (Figure 5.3; Table 5.3). One is located on chromosome 2 near position 20.3 cM (qSS2s) while the other is located on chromosome 11 near position 0.2 cM (qSS11s). Weedy alleles at both QTL increase seed shattering ability as expected, given the weedy parent's propensity for shattering. Currently, Gramene reports no shattering QTL on chromosome 2. However, our QTL on chromosome 11 may be close or overlapping with one found in a cross between wild *O. rufipogon* and an *indica* cultivar, where the wild allele increased seed shattering (Cai & Morishima, 2000). In the B population, one QTL was identified on chromosome 1 near position 189 cM (qSS1b) (Figure 5.4; Table 5.3). Weedy alleles at this QTL also work to increase seed shattering. This QTL is linked to a shared marker with a shattering QTL from a cross between *O. rufipogon* and a *tropical japonica* cultivar, where the wild allele increased seed shattering (Thomson *et al.*, 2003).

Heading Date (HD)

For the S population, a single QTL on chromosome 8 was identified by both SMA and CIM (qHD8s) (Figure 5.3). Weedy alleles at this QTL work to increase the days to heading, consistent with the later flowering seen in the weedy parent in the greenhouse (Table 5.3). In the B population, CIM identified two QTL also on chromosome 8 (qHD8.1b and qHD8.2b), while SMA only identified a single QTL (qHD8.1b) (Figure 5.4). Weedy alleles at these QTL work to increase the heading date, again consistent with the weedy parent's phenotype. It is possible that qHD8s and qHD8.1b share similar or linked causal genes as their 1.5 LOD intervals overlap and include marker 310.

Additionally, these QTL may be in a similar location as QTL identified by (Yano *et al.* 1997) and (Xiao *et al.,* 1998), in mapping populations involving *japonica* by *indica* crosses (with the *japonica* allele reducing flowering time), and *O. rufipogon* by cultivated crosses (with wild alleles increasing flowering time), respectively. Recently, potential candidate genes in this region have been cloned (Cai *et al.*, 2011; Wei *et al.*, 2010; Yan *et al.,* 2011).

Plant Height (PH)

Three QTL were detected for PH in the S population using CIM (qPH4s, qPH8s, qPH10s), while only one (qPH10s) was detected using SMA (Figure 5.3; Table 5.3). Of the three QTL, qPH10s (chromosome 10, 0.1 cM) explains most of the variation followed by qPH8s (chromosome 8, 127 cM) and qPH4s (chromosome 4, 40.9 cM). For qPH4s and qPH8s, the weedy allele increases plant height, as expected, while for qPH10s the cultivated allele increases plant height. Although we do not share any neighboring or linked markers with (Li *et al.,* 2006a), who mapped using F_2 from an *indica* crossed to a wild *O. nivara*, the markers associated with our chromosome 4 QTL are in similar physical locations. We do share a neighboring marker (284) with a QTL on chromosome 8 from (Thomson *et al.*, 2003) and a neighboring marker (239) with a QTL on chromosome 10 with (Septiningsih *et al.,* 2003), from mapping populations involving crosses between *japonica* or *indica* cultivars with a wild *O. rufipogon*; in all three studies, the wild allele increases plant height.

A single QTL was identified using both SMA and CIM in the B population located near position 147.4 cM on chromosome 1 (qPH1b) (Figure 5.4; Table 5.3). The weedy allele at this QTL increases plant height, as expected. Interestingly, the peak

marker for this QTL (5407) is physically located at ~39.5 megabase pairs (mbp), which is very close to *SD1*, a cloned gene of major effect (38.7 mbp). A large deletion in this gene, which encodes a critical enzyme involved in the final steps of gibberellin (GA) biosynthesis, has been shown to cause a semi-dwarf phenotype in cultivated rice, and our cultivated parent is known to have this deletion (Monna *et al.*, 2002; Reagon *et al.*, 2011).

Growth Rate (GR)

Unfortunately, no significant QTL were detected with both SMA and CIM in either population. GR is a complex trait likely involving many genes and epistatic interactions, so it is not surprising that we were unable to detect significant QTL. Interestingly, despite clear importance of this trait for plant fitness and competitiveness, we have found only one study mapping growth rate in *Oryza* (Li *et al.*, 2006c), which may be indicative of its complex genetic basis. Although Li *et al*. found several QTL underlying growth rate, their measurements were based on dry weight accumulated over time and are likely not comparable to ours.

5.3.4 Mapping Qualitative Traits

The results of the K-W test in R/qtl showed two significant loci for HC, one on chromosome 1 near position 75.2 cM and another on chromosome 4 near position 127.2 cM (Table 5.2). The locus on chromosome 1 may be novel. A locus on chromosome 4 controlling black to straw hull color change, identified in a cross between an *indica* cultivar and *O. rufipogon* was cloned recently (Os04g0460000; ~22.78mbp). Known as *Bh4*, this locus is physically close to our significant marker on chromosome 4 (~ 29.7mbp) (Zhu *et al.*, 2011), and may harbor mutations underlying our QTL.

The results of the K-W test in R/qtl showed a significant locus on chromosome 11 near position 37.8 cM for AW (Table 5.2). A QTL for awn length identified in a cross between an *indica* cultivar and wild *O. rufipogon* by (Cai & Morishima, 2002) on chromosome 11 is close to our locus, yet we do not share any co-localized markers.

5.4 Discussion

The repeatability of evolution can be seen as parallel changes at the phenotypic and/or genetic level between organisms evolving under similar environmental conditions. Questions remain about the extent to which shared genes are likely to underlie trait convergence among distant and closely related organisms (Arendt & Reznick, 2007; Hodin, 2000; Schluter *et al.*, 2004). Note that, following Arendt and Reznick (2007), we do not distinguish between the terms 'parallel" and 'convergent'. Various species of weedy plants repeatedly invade agricultural fields and are often subjected to similar selective pressures such as soil disruption, hand- and machine-weeding, herbicide treatment and competition with crop plants. These pressures are believed to lead to convergent phenotypic evolution of traits such as rapid growth, high seed production, increased seed dispersal and deep roots in weeds (Ellstrand *et al.*, 2010; Harlan & DeWet, 1965), which may be caused by genetic changes in similar genes or genetic pathways. Because weedy red rice in the US consists of two independently evolved groups, descendant from closely related cultivated ancestors (Reagon *et al.*, 2010), we sought to determine if parallel genetic changes were involved in the evolution of weediness in these groups, and if weedy traits could be attributed to variation present in wild and/or cultivated rice.

5.4.1 Lack of Parallel Genetic Evolution for Convergent Seed Shattering in US Weedy Rice

Of all the traits that differentiate weedy rice from its cultivated progenitors, seed shattering is likely the one that most characterizes the weedy phenotype. Selection against shattering to facilitate harvesting is a hallmark of cereal domestication (Purugganan & Fuller, 2009). In contrast, efficient seed dispersal is likely crucial to weed fitness, as it allows weeds to increase their presence in the seed bank and spread to new areas (Harlan & DeWet, 1965). We have previously shown that, despite separate origins, both US weedy rice populations are highly shattering compared to their putative cultivated progenitors (Thurber *et al.,* 2010). Thus, seed shattering is a trait for which true phenotypic convergence has occurred.

The genetic basis of loss of shattering in cultivated rice has been much explored, and *sh4*, a gene coding for a transcription factor involved in the formation and degradation of the abscission layer, has been identified as the most significant shattering gene to have been selected on during domestication (Li *et al.,* 2006b; Lin *et al.,* 2007; Zhang *et al.,* 2009). Studies have shown that all cultivated rice sampled to date share a single nucleotide substitution in *sh4,* which leads to loss of shattering (Thurber *et al.,* 2010; Zhang *et al.,* 2009). Recently, we found that both US weedy rice groups possess the same "non-shattering" substitution as cultivated rice, with weeds and cultivars carrying similar or identical *sh4* alleles (Thurber *et al.,* 2010). This implies that weedy groups must have re-acquired the shattering trait through involvement of other loci, rather than acquisition of ancestral or novel *sh4* alleles. We have additionally shown that both weed groups have convergence of the shattering trait at the morphological level − formation and degradation of the abscission layer is similar among weedy rice groups,

but distinct from shattering wild rice (Thurber *et al.*, 2011). This phenotypic evidence would suggest that parallel genetic changes underlie convergence of shattering in weedy rice groups.

Surprisingly, both the observed phenotypic distributions in our mapping populations and identified QTL do not support parallel genetic changes in seed shattering between weedy groups. In the S population a highly non-normal distribution with multiple peaks suggests that a few major effect genes contribute to this trait. In contrast, the more normal shattering distribution in the B population suggests involvement of multiple weaker effect loci. Differing genetic architecture for seed shattering does not necessarily exclude the possibility of shared loci. We identified two QTL for seed shattering in the S population and one in the B population; however, these are located on three different chromosomes (1, 2 and 11) and are not shared. Only one of the three shattering QTL we mapped (qSS1b) is located on the same chromosome as a previously cloned shattering gene (*qsh1*), yet we do not share neighboring markers and previous work has suggested that this gene does not play a role in US weedy rice seed shattering (Thurber *et al.*, 2010). Thus, despite the potential of shared genetic biases due to shared ancestry of the weed progenitor groups, and despite convergence of the trait at various phenotypic levels, shattering in US weedy rice does not seem to be due to parallel genetic changes.

5.4.2 The Potential for Parallel Genetic Evolution in Divergent Weedy Traits

Despite predictions that convergence of weed-adaptive traits should occur among weeds evolving in agricultural settings, we have found divergence for several traits among our closely related weedy rice groups (Reagon *et al.*, 2011). In particular, SH

weeds flower significantly earlier and BHA weeds significantly later than their cultivated progenitors in growth chamber conditions, and SH weeds tend to be shorter and BHA weeds taller than their ancestors (Reagon *et al.,* 2011) (Table 5.4). These differences translate into divergence among weed groups that have also been reported in the field (Shivrain *et al.,* 2010a). These phenotypic patterns suggest that both flowering time and height have been under selection during weed evolution, but that, despite identical environmental conditions, multiple "adaptive solutions" exist for weedy phenotypes.

Mutations in shared genes could still underlie divergent traits if shared signaling and/or metabolic pathways shape multiple alternative trait features (Hodin, 2000). Under this scenario, mutations would not be shared among divergent groups, but alternate mutations of the same gene could underlie the divergent phenotypes. We thus looked for any evidence of shared QTL between the S and B populations for plant height and heading date. Because our *indica* parent was a semi-dwarf, we were limited to exploring QTL that increase plant height relative to semi-dwarfness.

The similar distribution patterns in PH between populations suggested similar genetic architectures for this trait. However, QTL locations were not shared between the S and B populations, and effect directions were not always predictable from parental phenotypes. We identified a single QTL in the B population (chromosome 1) where the weedy allele increases height, yet we identified three QTL (chromosomes 4, 8, and 10) in the S population, with the weedy allele increasing height at only two loci. A major locus controlling plant height in cultivated rice has been identified as *SD1*, within which a large deletion has been shown to cause a semi-dwarf phenotype that was employed in breeding during the green revolution (Monna *et al.,* 2002). We expected to detect this QTL in both

populations, as our mapping parents differ in *SD1* alleles, and our cultivated parent contains this deletion (Reagon *et al.,* 2011). The single QTL identified in the B population is physically close to *SD1*, indicating that we did not detect QTL specific to evolution of plant height in the BHA lineage. In contrast, the three QTL in the S population may have contributed to evolution of plant height in the SH lineage. These QTL may come from standing variation in the crop or wild rice, as previously reported height QTL appear to be near these. This implies there may be parallel evolution between some cultivars and weeds.

Among growth related traits, phenotypic divergence between weedy groups and between weeds and their cultivated ancestors is most apparent for flowering time (Reagon *et al.,* 2010). Our two mapping populations do not share similar phenotypic distributions for this trait, with the involvement of a few major effect loci suggested for the S population, and multiple weaker effect loci suggested for the B population. Surprisingly, our results indicate that heading date is the trait with the most potential for similar genes underlying evolutionary changes in both weed groups. All three QTL identified, one in the S population and two in the B population, are located on chromosome 8, and the QTL in the S population shares a neighboring marker with one found in the B population. Consistent with the switch to later flowering exhibited by the SH parent in the greenhouse, in all three cases the weedy allele increases days to heading.

Flowering in rice is known to be controlled by several genes that interact to create a wide range in heading dates across different environments (Takahashi *et al.,* 2009). In particular, variations in *Hd1*, which encodes a zinc finger domain protein responsible for the transition from vegetative to reproductive phase (Yano *et al.,* 2000), have been

implicated as major regulators of flowering time (Takahashi *et al.*, 2009). A cursory look at *Hd1* coding region alleles in weedy rice suggests the involvement of *Hd1* in weed flowering. Our BHA parent has a haplotype containing a two bp deletion that is known to result in photoperiod insensitivity (Takahashi *et al.*, 2009; Thurber & Caicedo, unpublished data); likewise our *indica* parent contains a four bp deletion that creates a frame shift, leading to a nonfunctional haplotype and photoperiod insensitivity (Takahashi *et al.*, 2009; Thurber & Caicedo, unpublished data). In contrast, our SH parent shares an intact haplotype common in cultivated rice (Thurber & Caicedo, unpublished data), which is known to cause photoperiod sensitivity and short day flowering behavior in cultivated rice (Takahashi *et al.*, 2009; Yano *et al.*, 2000). Our mapping results suggest that *Hd1* does not mediate differences in flowering time between weed groups and between weeds and *indica* cultivars under the variable, primarily long day conditions in our greenhouse. Given that our planting time reflects the timing of planting in the Southern US rice fields, our results also suggest that a novel locus or set of loci on chromosome 8 underlie the flowering time differences between weed groups in the field and are likely responsible for the divergence of both weed groups from their cultivated ancestors. The HD QTL we discovered here may also be contributing to variation in flowering time in cultivated and wild rice, as some candidate genes have been recently identified on chromosome 8 (Cai *et al.*, 2011; Wei *et al.*, 2010; Yan *et al.*, 2011).

5.4.3 The Potential for Shared Genes Involved in Reversals to Wild Phenotypes

Three of our traits (SS, HC and AW) show a clear reversal of a cultivated phenotype (non-shattering, straw colored hulls and no awns) to a wild phenotype (shattering, black hulls and long awns). Due to the diversity of the cultivated ancestral

gene pool, as seen by the wide range of hull and awn morphologies in our collection of *aus* and *indica* cultivars, it is possible that genes involved in some weedy traits could have arisen from standing ancestral variation. Alternatively, although lack of a role for wild rice *sh4* alleles in the shattering phenotype of weedy rice has been demonstrated, genes underlying hull color and awn presence in the wild ancestor of rice could be involved in weedy phenotypes either through introgression or compensatory mutations that reverse the phenotype in the weeds. Thus parallel genetic evolution can be envisioned at another level for weedy rice: weeds may also share genes underlying weedy traits with wild or cultivated rice.

We checked for evidence of shared genetic changes by examining published QTL from studies involving crosses of wild and cultivated *Oryza* groups. Seed shattering QTL have been mapped to nearly every rice chromosome, yet our QTL potentially overlap with only two previously published QTL, both identified in wild by cultivated rice crosses. Interestingly, the QTL we report on chromosome 2 in the S population is the first shattering QTL to be identified on that chromosome. The sharing of some QTL with wild rice suggests that the transition from non-shattering to shattering during weed de-domestication may involve some similar genes as the transition from shattering to non-shattering during domestication.

Although our QTL for awns did not overlap with any other published QTL, our hull color QTL on chromosome 4 is likely to be the recently cloned *Bh4* locus (Zhu *et al.*, 2011). Hull color in *Oryza* can vary from light (nearly white and straw) to medium (gold furrowed or brown) to dark (black); this trait is slightly ambiguous in its function, yet may be important for seed dormancy, camouflage (both on the plant and on the ground)

and seed dispersal (Zhu *et al.*, 2011). *Bh4* is a gene encoding for an amino acid transporter, and multiple deletions and SNPs that cause frame shifts and premature stop codons seem to be involved in the transition from black hulls in wild rice to straw hulls in cultivated rice (Zhu *et al.*, 2011). Our results suggest that for some weedy traits, causal alleles may be shared with wild rice.

5.4.4 Parallel Evolution Among Global Populations of Weedy Rice

A few other studies have involved mapping weed adaptive traits in crosses between non-US weedy rice and cultivated rice, though with no knowledge of the relationship between the two parents or the evolutionary origin of the weed. One such study mapped several traits (e.g. seed shattering, heading date, plant height, and yield components) in a weedy rice from France crossed to a *japonica* cultivar (Bres-Patry *et al.*, 2001), while another set of studies examined seed dormancy, shattering, awns and hull color in a weedy rice from Thailand crossed to an *indica* (Gu *et al.*, 2005a). We do not share any QTL for overlapping traits with either study, suggesting that parallel genetic evolution may not be the norm among worldwide weedy rice populations.

5.4.5 Conclusions and Future Directions

The QTL we detected gives us a starting point for identifying genes involved in weed adaptive traits. In coming years, with new resources, we plan to narrow down the genomic regions underlying evolution of US weeds. In particular, our study was hampered by the close relationship between our cross parents. Population structure analyses based on multiple loci cannot differentiate SH weeds from *indica* cultivars, and BHA weeds share alleles with both *aus* and *indica* (Reagon *et al.*, 2010), making finding

segregating markers among the parents difficult. Currently, the genomes of the three parents are being sequenced (Hyma & Caicedo, unpublished), which will improve marker density. Additionally, we are generating RIL populations derived from these crosses (Jia, Caicedo & Olsen, unpublished), which will be useful for narrowing down QTL regions and testing for QTL in multiple environments.

One caveat of our study is the lack of a cross between a BHA weed and its putative *aus* progenitor. Several attempts were made to create this cross in our lab, yet none were successful. Due to the relationship among the cultivated and weed groups, the QTL detected from the BHA-*indica* cross could include genomic regions that differ between BHA and *aus* as well as those that differ between *indica* and *aus* (Figure 5.1). Fortunately, this does not hurt our ability to detect QTL relevant to weed evolution. We are continuing attempts to create a BHA-*aus* cross to determine which QTL are specific to BHA weeds.

This study represents a first step towards dissecting the extent of parallel evolution in weed adaptive traits of a potent agricultural weed. Our finding of lack of parallel evolution at the genetic level for shattering, one of the most characteristic traits of weedy rice, joins others in showing that close evolutionary relationships do not imply use of the same genes in adaptation (Arendt & Reznick 2007). Conversely, shared genetic pathways can be implicated in the evolution of divergent phenotypes, as is likely for flowering time in weeds. Further fine-mapping of genes underlying adaptive traits in weedy rice groups, and search for weed alleles in wild and cultivated ancestors, will contribute to our eventual understanding of the circumstances under which convergent genetic evolution occurs.

Table 5.1. Phenotypes of the parental lines crossed to create F_2 mapping populations. Plants were initially chosen for their different phenotypes in the growth chamber. Greenhouse measurements are averages of three plants. Growth chamber measurements are averages of two plants.

Greenhouse

Plant ID	Plant Height (cm)	Heading Date (days)	Growth Rate (cm/day)	Seed Shattering[a] (grams)	Hull Color	Awn
SH-RR09	76.8	107.3	0.716	0	0	0
DGWG	42.7	101	0.423	28.4	0	0
BHA-RR20	104.7	116	0.903	5.5	1	1

Growth Chamber

Plant ID	Plant Height (cm)	Heading Date (days)	Growth Rate (cm/day)	Seed Shattering[a] (grams)	Hull Color	Awn
SH-RR09	83	60	1.383	0.3	0	0
DGWG	59	109	0.541	60.9	0	0
BHA-RR20	67	116	0.578	7.2	1	1

[a] Seed Shattering was measured as described in Thurber et al., 2010 using the Breaking Tensile Strength method.

Table 5.2. Segregation and mapping of qualitative trait loci. In both cases the weedy phenotype is dominant.

	Phenotypic Ratio						
Trait	D	R	χ^2	Chr.	Linked Markers	Distance (cM)	LOD
AW	136	23	9.36	11	202	37.8	2.37
HC	94	40	0.236	1	9	75.2	2.56
				4	6748	127.2	2.87

Table 5.3. QTLs for quantitative traits detected in the F_2 populations.

QTL	Chr.	Position (cM)	Nearest Marker	LOD			R^{2a}	Phenotypic Means[b]			Increased Effect[c]	Allele Effect[d]
				CIM	SMA	NP		0	1	2		
SS												
qSS1b	1	189	104	N/A	N/A	2.23	N/A	19.55	14.21	13.02	DGWG	-5.34
qSS2s	2	15.2	236	N/A	N/A	6.85	N/A	13.14	8.12	4.44	DGWG	-5.02
qSS11s	11	0.2	332	N/A	N/A	2.71	N/A	9.16	10.21	3.84	DGWG	1.05
HD												
qHD8s	8	0.5	310	15.1	48.11	N/A	28.98	96.55	125.94	139.93	SH-RR09	29.39
qHD8.1b	8	19.7	25	2.8	8.89	N/A	11.68	112.05	114.83	123.4	BHA-RR20	2.78
qHD8.2b	8	76	44	2.7	N.S	N/A	6.63	119.2	111.96	120.94	BHA-RR20	-7.24
PH												
qPH1b	1	147.4	5407	4.6	17.32	N/A	14.38	56.72	66.8	68.95	SH-RR20	10.08
qPH4s	4	40.9	417	2.3	N.S.	N/A	5.2	66.5	63.14	78.79	SH-RR09	-3.36
qPH8s	8	127	477	3.16	N.S.	N/A	9.5	68.55	61.37	75.05	SH-RR09	-7.18
qPH10s	10	0.1	239	2.36	5.51	N/A	18.5	71.74	65.83	63.47	DGWG	-5.91

[a] R^2 indicates the percentage of phenotypic variation explained by the putative QTL; only determined when CIM was significant.

[b] Phenotypic means calculated for the DGWG homozygote (0), heterozygote (1) and weedy homozygote (2).

[c] Increased effect is the source of the allele causing an increase in the phenotypic value

[d] The allele effect is the effect associated with substituting a DGWG allele with a weedy allele

Table 5.4. Means and standard deviations of phenotypes measured in cultivated and weedy populations. Standard deviations are in parenthesis.

	Plant Height[b] (cm)	Heading Date[b] (days)	Seed Shattering[a] (grams)	Hull Color	Awn	Average Growth Rate[b] (cm/day)
Cultivated rice						
indica	75(23)	120(20)	57.83(47.07)	straw	no	0.63(0.17)
aus	80(13)	107(9)	17.36(15.82)	straw	no	0.75(0.18)
japonica	76(15)	116(30)	99.92(72.41)	straw	no	0.66(0.26)
Weedy rice						
SH	69(14)	80(11)	0.7(0.73)	straw	no	0.86(0.25)
BHA	85(25)	133(16)	0.74(2.07)	black	yes	0.64(0.22)

[a] Seed Shattering data were reported in Thurber et al., 2010.

b Plant Height, Average Growth Rate and Heading Date were reported in Reagon et al., 2011, with the exception that Heading Date and Average Growth Rate measurements reported here are from date emerged rather than date sown.

Table 5.5. Broad-sense heritability values for the quantitative traits studied.

S Population	V_p	V_{cult}	V_{weed}	V_e	V_g	H^2
SS	93.91	63.94	0	31.97	61.94	0.659568
HD	1192.21	12	0.92	6.46	1185.75	0.994581
PH	307.12	49.33	26.26	37.795	269.325	0.876937
GR	0.042	0.0032	0.002	0.0026	0.0394	0.938095
B Population	V_p	V_{cult}	V_{weed}	V_e	V_g	H^2
SS	103.91	63.94	3.38	33.66	70.25	0.676066
HD	325.3	12	3	7.5	317.8	0.976944
PH	207.03	49.33	14.33	31.83	175.2	0.846254
GR	0.025	0.0032	0.0016	0.0024	0.0226	0.904

Table 5.6. Pearson's correlations between traits in the F_2 populations. Only significant correlations are shown ($P < 0.01$).

S population	*HD*	*SS*	*PH*
HD			
SS			
PH	**0.204**		
GR			

B population	*HD*	*SS*	*PH*	*GR*	*HC*
HD					
SS					
PH	**0.189**				
GR					
HC					
AW					

Table 5.7. Phenotype values for the S population individuals.

Individual	HD	SS	PH	GR	log10HD	log10PH	log10GR
S001	147	0	60	0.669643	2.167317	1.778151	-0.17416
S002	67	0	76	1.2	1.826075	1.880814	0.079181
S003	112	0	51	0.472727	2.049218	1.70757	-0.32539
S004	152	17.8	73	0.331776	2.181844	1.863323	-0.47916
S005	108	0	59	0.261364	2.033424	1.770852	-0.58275
S006	153	4	60	0.813187	2.184691	1.778151	-0.08981
S007	112	18.5	64	0.610169	2.049218	1.80618	-0.21455
S008	77	2.1	50	0.413793	1.886491	1.69897	-0.38322
S009	107	5.95	67	0.576577	2.029384	1.826075	-0.23914
S010	146	0	77	0.563107	2.164353	1.886491	-0.24941
S011	92	0	56	0.408284	1.963788	1.748188	-0.38904
S012	99	16.8	50	0.880342	1.995635	1.69897	-0.05535
S013	82	0	53	0.519608	1.913814	1.724276	-0.28432
S014	189	0	90	0.961538	2.276462	1.954243	-0.01703
S015	91	0	80	0.924528	1.959041	1.90309	-0.03408
S018	78	17	53	0.346535	1.892095	1.724276	-0.46025
S019	165	2.82	60	0.4375	2.217484	1.778151	-0.35902
S020	91	9.2	59	0.675325	1.959041	1.770852	-0.17049
S021	95	21.8	76	0.461538	1.977724	1.880814	-0.33579
S022	111	3.62	64	0.424779	2.045323	1.80618	-0.37184
S023	145	7.2	60	0.447368	2.161368	1.778151	-0.34933
S024	101	0	65	0.387879	2.004321	1.812913	-0.4113
S025	106	5.84	70	0.782609	2.025306	1.845098	-0.10646
S027	82	0	63	0.672515	1.913814	1.799341	-0.1723
S028	171	10.8	115	0.394737	2.232996	2.060698	-0.40369
S031	154	6.3	85	0.613445	2.187521	1.929419	-0.21222
S032	113	17.3	96	0.348101	2.053078	1.982271	-0.45829
S033	115	5.7	58	0.40884	2.060698	1.763428	-0.38845
S034	105	0	40	0.264151	2.021189	1.60206	-0.57815
S035	91	6.74	42	0.757282	1.959041	1.623249	-0.12074
S037	112	0	42	0.892473	2.049218	1.623249	-0.0494
S038	109	19.1	63	0.703704	2.037426	1.799341	-0.15261
S039	88	15.3	35	0.689189	1.944483	1.544068	-0.16166
S040	91	2.2	95	0.303318	1.959041	1.977724	-0.5181
S041	158	0	55	0.627907	2.198657	1.740363	-0.2021
S043	91	14.6	58	0.539823	1.959041	1.763428	-0.26775
S044	99	3.4	60	0.339535	1.995635	1.778151	-0.46912
S045	105	0	44	0.714286	2.021189	1.643453	-0.14613
S046	110	20.1	56	0.460784	2.041393	1.748188	-0.3365
S047	109	17.6	59	1.105263	2.037426	1.770852	0.043466
S049	80	19.7	50	0.300493	1.90309	1.69897	-0.52217
S050	83	28.7	45	0.363636	1.919078	1.653213	-0.43933
S051	121	0	71	0.567308	2.082785	1.851258	-0.24618
S053	88	1.26	101	0.657754	1.944483	2.004321	-0.18194
S054	142	15	48	0.52	2.152288	1.681241	-0.284
S055	85	0.42	102	0.723404	1.929419	2.0086	-0.14062
S057	187	0	123	0.29878	2.271842	2.089905	-0.52465

Table 5.7 Cont.

S058	94	0	68	0.642202	1.973128	1.832509	-0.19233
S059	102	10.8	67	0.833333	2.0086	1.826075	-0.07918
S060	160	0	47	0.26178	2.20412	1.672098	-0.58206
S061	159	0	85	0.425532	2.201397	1.929419	-0.37107
S062	212	8.35	75	0.97619	2.326336	1.875061	-0.01047
S064	97	0	92	0.509091	1.986772	1.963788	-0.2932
S066	114	8.04	64	0.324324	2.056905	1.80618	-0.48902
S067	145	0	86	0.510204	2.161368	1.934498	-0.29226
S070	71	28.3	55	0.551948	1.851258	1.740363	-0.2581
S071	108	3.44	64	0.8	2.033424	1.80618	-0.09691
S072	104	0	63	0.346535	2.017033	1.799341	-0.46025
S073	170	0	53	0.239521	2.230449	1.724276	-0.62066
S074	119	15.7	74	0.431373	2.075547	1.869232	-0.36515
S075	87	20.6	37	0.729167	1.939519	1.568202	-0.13717
S076	142	0	54	0.551471	2.152288	1.732394	-0.25848
S077	107	7.3	90	0.480263	2.029384	1.954243	-0.31852
S078	98	0	61	0.406593	1.991226	1.78533	-0.39084
S079	110	4.1	86	0.504348	2.041393	1.934498	-0.29727
S080	143	20.4	59	0.769231	2.155336	1.770852	-0.11394
S082	104	0	80	0.455357	2.017033	1.90309	-0.34165
S083	111	8.67	65	0.555556	2.045323	1.812913	-0.25527
S084	96	27.1	44	0.745455	1.982271	1.643453	-0.12758
S085	98	30.2	62	0.496552	1.991226	1.792392	-0.30404
S087	102	1.85	65	0.361446	2.0086	1.812913	-0.44196
S088	168	27.7	51	0.584746	2.225309	1.70757	-0.23303
S089	158	0	80	0.574074	2.198657	1.90309	-0.24103
S090	186	15.9	104	0.509317	2.269513	2.017033	-0.29301
S091	119	0	116	0.542169	2.075547	2.064458	-0.26587
S093	140	19.9	48	0.545918	2.146128	1.681241	-0.26287
S094	89	1.9	49	0.695238	1.94939	1.690196	-0.15787
S096	176	17.7	66	0.569767	2.245513	1.819544	-0.2443
S097	103	4.8	85	0.375	2.012837	1.929419	-0.42597
S098	84	8.63	57	0.643564	1.924279	1.755875	-0.19141
S099	164	0	74	0.911765	2.214844	1.869232	-0.04012
S100	176	4.4	90	0.679487	2.245513	1.954243	-0.16782
S102	102	2.7	53	0.522124	2.0086	1.724276	-0.28223
S105	103	0	65	0.777778	2.012837	1.812913	-0.10914
S106	101	0	56	0.428571	2.004321	1.748188	-0.36798
S107	118	0	69	0.381356	2.071882	1.838849	-0.41867
S108	147	0	65	0.552	2.167317	1.812913	-0.25806
S109	167	6.17	40	0.408163	2.222716	1.60206	-0.38917
S110	114	4.93	46	1.294872	2.056905	1.662758	0.112227
S112	196	21.2	107	0.363636	2.292256	2.029384	-0.43933
S113	179	9.98	86	0.573034	2.252853	1.934498	-0.24182
S116	86	8.2	54	0.918367	1.934498	1.732394	-0.03698
S117	109	2.93	73	0.666667	2.037426	1.863323	-0.17609
S118	84	11.03	82	0.586777	1.924279	1.913814	-0.23153
S119	132	0	66	0.771186	2.120574	1.819544	-0.11284
S120	144	6.67	63	0.82716	2.158362	1.799341	-0.08241

Table 5.7 Cont.

S121	93	19.9	83	0.637363	1.968483	1.919078	-0.19561
S122	86	0	49	0.648352	1.934498	1.690196	-0.18819
S124	159	3.32	42	0.445783	2.201397	1.623249	-0.35088
S125	75	37.4	72	0.541284	1.875061	1.857332	-0.26657
S126	126	0	82	0.449275	2.100371	1.913814	-0.34749
S128	91	4.7	74	0.33	1.959041	1.869232	-0.48149
S129	74	9.94	51	0.544872	1.869232	1.70757	-0.26371
S130	128	1.6	45	0.669725	2.10721	1.653213	-0.1741
S132	78	0	101	0.656863	1.892095	2.004321	-0.18253
S133	108	8.64	53	0.849057	2.033424	1.724276	-0.07106
S134	112	0	75	0.263158	2.049218	1.875061	-0.57978
S135	196	0	101	0.525	2.292256	2.004321	-0.27984
S136	91	0	70	0.420513	1.959041	1.845098	-0.37622
S138	112	0	80	0.387755	2.049218	1.90309	-0.41144
S140	96	5.85	70	0.574074	1.982271	1.845098	-0.24103
S141	85	2.1	64	0.660377	1.929419	1.80618	-0.18021
S143	127	7	39	0.442177	2.103804	1.591065	-0.3544
S145	77	9.84	85	1.147727	1.886491	1.929419	0.059839
S146	118	8.6	75	0.280899	2.071882	1.875061	-0.55145
S147	81	25.44	67	0.577982	1.908485	1.826075	-0.23809
S149	125	0	69	0.352041	2.09691	1.838849	-0.45341
S151	102	9.2	93	0.606061	2.0086	1.968483	-0.21748
S152	165	24.3	78	0.288043	2.217484	1.892095	-0.54054
S153	110	3.3	82	0.431579	2.041393	1.913814	-0.36494
S154	103	20.8	78	0.768293	2.012837	1.892095	-0.11447
S158	184	0	53	0.376812	2.264818	1.724276	-0.42388
S162	104	3.4	100	0.625	2.017033	2	-0.20412
S164	214	23.2	71	0.403509	2.330414	1.851258	-0.39415
S166	192	14.7	77	0.351562	2.283301	1.886491	-0.454
S168	106	12.3	90	1.043956	2.025306	1.954243	0.018682
S170	159	17.6	100	0.355705	2.201397	2	-0.44891
S171	118	56.4	91	0.413613	2.071882	1.959041	-0.38341
S178	86	4.4	65	0.475728	1.934498	1.812913	-0.32264
S179	87	1.83	84	0.546296	1.939519	1.924279	-0.26257
S180	160	6.13	84	0.487013	2.20412	1.924279	-0.31246
S181	113	0	59	0.68	2.053078	1.770852	-0.16749
S182	131	0	72	0.582609	2.117271	1.857332	-0.23462
S183	75	27.62	51	0.386364	1.875061	1.70757	-0.413
S184	154	30.8	75	0.401042	2.187521	1.875061	-0.39681
S185	103	1.19	58	0.385321	2.012837	1.763428	-0.41418
S186	141	20.8	60	0.752941	2.149219	1.778151	-0.12324
S187	193	15.1	90	0.516484	2.285557	1.954243	-0.28694
S188	76	17.9	84	0.338028	1.880814	1.924279	-0.47105
S190	176	0	46	0.527397	2.245513	1.662758	-0.27786
S191	115	1.44	67	0.490741	2.060698	1.826075	-0.30915
S192	105	11.1	73	0.401274	2.021189	1.863323	-0.39656
S193	119	4.55	73	0.380952	2.075547	1.863323	-0.41913
S195	112	14.21	48	0.307087	2.049218	1.681241	-0.51274
S196	140	12.3	42	0.554455	2.146128	1.623249	-0.25613

Table 5.7	Cont.						
S197	161	15.7	82	0.755814	2.206826	1.913814	-0.12159
S198	185	0	60	0.608696	2.267172	1.778151	-0.2156
S200	118	15.7	72	0.605769	2.071882	1.857332	-0.21769
S202	110	26.8	44	0.646341	2.041393	1.643453	-0.18954
S204	113	0	48	0.47619	2.053078	1.681241	-0.32222
S206	108	0	53	0.433498	2.033424	1.724276	-0.36301
S207	118	0	45	0.96	2.071882	1.653213	-0.01773
S208	162	13.7	90	0.3	2.209515	1.954243	-0.52288
S209	104	7.7	59	1.134328	2.017033	1.770852	0.054739
S210	98	0	50	0.965517	1.991226	1.69897	-0.01524
S211	82	17.3	47	0.522523	1.913814	1.672098	-0.28189
S212	202	13.32	70	0.480447	2.305351	1.845098	-0.31835
S214	169	1.65	69	0.515306	2.227887	1.838849	-0.28793
S215	104	18.3	63	0.573171	2.017033	1.799341	-0.24172
S217	153	8.9	66	1.103896	2.184691	1.819544	0.042928
S218	91	3.7	37	0.490741	1.959041	1.568202	-0.30915
S220	98	4.8	90	0.419048	1.991226	1.954243	-0.37774
S221	149	10.77	53	0.5	2.173186	1.724276	-0.30103
S222	117	1.29	91	0.849558	2.068186	1.959041	-0.07081
S223	145	0	72	0.628931	2.161368	1.857332	-0.2014
S224	178	0	50	0.519608	2.25042	1.69897	-0.28432
S225	166	13.8	60	0.505051	2.220108	1.778151	-0.29667
S226	113	4.2	61	0.466321	2.053078	1.78533	-0.33131
S228	102	14.3	53	0.549618	2.0086	1.724276	-0.25994
S229	77	16.7	52	0.358108	1.886491	1.716003	-0.44599
S230	138	0	62	0.879121	2.139879	1.792392	-0.05595
S231	102	23.4	85	0.397727	2.0086	1.929419	-0.40041
S232	164	4.3	49	0.631068	2.214844	1.690196	-0.19992
S233	69	7.5	54	0.380952	1.838849	1.732394	-0.41913
S234	87	0	59	0.386076	1.939519	1.770852	-0.41333
S235	133	0	35	0.511364	2.123852	1.544068	-0.29127
S236	90	5.3	60	0.678161	1.954243	1.778151	-0.16867
S239	158	8.72	61	0.650794	2.198657	1.78533	-0.18656
S240	202	1.77	70	0.356021	2.305351	1.845098	-0.44852
S243	157	10.6	63	0.723577	2.1959	1.799341	-0.14052
S245	147	0	57	0.386364	2.167317	1.755875	-0.413
S248	108	19	62	0.4	2.033424	1.792392	-0.39794
S249	95	13.6	41	0.635593	1.977724	1.612784	-0.19682
S250	103	43.2	49	0.314465	2.012837	1.690196	-0.50243

Table 5.8. Phenotype values for the B population individuals.

Individual	HD	SS	PH	GR	HC	AW	log10HD	log10PH	log10GR
B001	120	7.1	64	0.454545	1	1	2.079181	1.80618	-0.34242
B003	83	19.9	57	0.59542	1	1	1.919078	1.755875	-0.22518
B004	137	8.96	53	0.386861	1	1	2.136721	1.724276	-0.41244
B005	145	11.5	88	0.583333	0	1	2.161368	1.944483	-0.23408
B007	127	23.1	53	0.512605	x	1	2.103804	1.724276	-0.29022
B008	135	21.3	78	0.650943	1	1	2.130334	1.892095	-0.18646
B009	112	11	72	0.463918	1	1	2.049218	1.857332	-0.33356
B010	104	0	57	0.505618	1	0	2.017033	1.755875	-0.29618
B011	131	21.9	78	0.562044	x	1	2.117271	1.892095	-0.25023
B012	125	11.5	66	0.510345	0	1	2.09691	1.819544	-0.29214
B013	104	0	44	0.427536	0	0	2.017033	1.643453	-0.36903
B014	96	0	60	0.428571	1	1	1.982271	1.778151	-0.36798
B015	131	1.24	96	0.491667	1	1	2.117271	1.982271	-0.30833
B018	102	2	52	0.602837	1	1	2.0086	1.716003	-0.2198
B019	106	36.2	57	0.475806	0	0	2.025306	1.755875	-0.32257
B020	134	17.5	63	0.451977	1	1	2.127105	1.799341	-0.34488
B023	116	0	57	0.347107	1	0	2.064458	1.755875	-0.45954
B025	99	3.45	56	0.512397	x	1	1.995635	1.748188	-0.29039
B026	117	17.7	62	0.622222	0	1	2.068186	1.792392	-0.20605
B027	97	13.5	45	0.470149	1	1	1.986772	1.653213	-0.32776
B028	113	0	63	0.368056	x	1	2.053078	1.799341	-0.43409
B029	105	17	64	0.645669	1	1	2.021189	1.80618	-0.18999
B030	98	21.4	54	0.406504	1	1	1.991226	1.732394	-0.39094
B032	93	9.8	50	0.5	1	1	1.968483	1.69897	-0.30103
B034	93	14.9	74	0.597938	1	1	1.968483	1.869232	-0.22334
B035	106	21.5	69	0.458647	1	1	2.025306	1.838849	-0.33852
B037	92	14.2	47	1.152941	0	1	1.963788	1.672098	0.061807
B038	112	16	53	0.261146	1	1	2.049218	1.724276	-0.58312
B039	112	28.3	37	0.5	1	0	2.049218	1.568202	-0.30103
B040	110	16	84	0.763636	0	1	2.041393	1.924279	-0.11711

Table 5.8 Cont.

B041	111	15.4	78	0.625	0	1	2.045323	1.892095	-0.20412
B042	131	14.7	83	0.6875	1	1	2.117271	1.919078	-0.16273
B043	102	11.5	45	0.294118	1	1	2.0086	1.653213	-0.53148
B046	127	12	82	0.650485	1	1	2.103804	1.913814	-0.18676
B047	110	10.8	68	0.566265	0	1	2.041393	1.832509	-0.24698
B049	105	2.1	73	0.435897	1	1	2.021189	1.863323	-0.36062
B051	102	12.8	99	0.531532	1	1	2.0086	1.995635	-0.27447
B052	119	2.21	87	0.597938	0	1	2.075547	1.939519	-0.22334
B053	126	18	81	0.261146	1	1	2.100371	1.908485	-0.58312
B057	110	4.4	63	0.695238	0	0	2.041393	1.799341	-0.15787
B060	136	26.4	73	0.483333	1	1	2.133539	1.863323	-0.31575
B061	104	22.1	86	0.697917	1	1	2.017033	1.934498	-0.1562
B062	128	22.6	108	0.568421	x	1	2.10721	2.033424	-0.24533
B064	103	9.04	58	0.328467	1	0	2.012837	1.763428	-0.48351
B065	113	5.68	76	0.647619	1	1	2.053078	1.880814	-0.18868
B066	129	15.8	53	0.815534	1	1	2.11059	1.724276	-0.08856
B067	113	14.8	72	0.335938	1	0	2.053078	1.857332	-0.47374
B068	93	3.5	69	0.583333	1	1	1.968483	1.838849	-0.23408
B069	95	23.5	74	0.836735	0	1	1.977724	1.869232	-0.07741
B071	127	16.1	66	0.386861	0	1	2.103804	1.819544	-0.41244
B072	132	29.5	52	0.333333	0	1	2.120574	1.716003	-0.47712
B075	109	15.2	59	0.527778	0	1	2.037426	1.770852	-0.27755
B076	130	2.91	50	0.477941	0	1	2.113943	1.69897	-0.32063
B078	126	9.94	81	0.444444	1	1	2.100371	1.908485	-0.35218
B079	112	18.3	81	0.540541	1	1	2.049218	1.908485	-0.26717
B081	107	26.9	81	0.686747	1	1	2.029384	1.908485	-0.1632
B082	85	1.12	49	0.478992	x	1	1.929419	1.690196	-0.31967
B083	123	4.55	75	0.4	1	1	2.089905	1.875061	-0.39794
B085	94	2.67	52	0.970588	1	1	1.973128	1.716003	-0.01296
B086	120	23.8	66	0.578947	0	0	2.079181	1.819544	-0.23736
B087	121	9.24	74	0.573034	x	1	2.082785	1.869232	-0.24182
B091	112	15.9	56	0.320611	1	1	2.049218	1.748188	-0.49402

Table 5.8 Cont.

B094	105	30.4	81	0.330357	0	0	2.021189	1.908485	-0.48102
B097	79	16.1	63	0.606897	0	1	1.897627	1.799341	-0.21689
B098	112	37.3	49	0.55102	1	1	2.049218	1.690196	-0.25883
B099	114	15.1	68	0.398496	x	1	2.056905	1.832509	-0.39958
B100	123	12.9	50	0.5	1	1	2.089905	1.69897	-0.30103
B101	120	0	60	0.45045	0	0	2.079181	1.778151	-0.34635
B103	106	20.3	71	0.537634	1	1	2.025306	1.851258	-0.26951
B105	81	23.4	44	0.959596	1	1	1.908485	1.643453	-0.01791
B106	99	22.2	61	0.4	1	1	1.995635	1.78533	-0.39794
B107	125	11.8	71	0.507692	0	1	2.09691	1.851258	-0.2944
B108	97	25.1	58	0.335938	x	1	1.986772	1.763428	-0.47374
B109	138	17.5	59	0.537313	1	1	2.139879	1.770852	-0.26977
B112	93	16.64	62	0.561905	1	0	1.968483	1.792392	-0.25034
B113	129	20.3	72	0.55	1	0	2.11059	1.857332	-0.25964
B114	121	12.5	45	0.571429	1	0	2.082785	1.653213	-0.24304
B116	124	15.7	63	0.815534	1	1	2.093422	1.799341	-0.08856
B117	111	5.22	60	0.371901	0	1	2.045323	1.778151	-0.42957
B118	83	2.64	47	0.333333	0	1	1.919078	1.672098	-0.47712
B119	117	19	51	0.771429	1	1	2.068186	1.70757	-0.1127
B120	110	13.6	55	0.697917	x	1	2.041393	1.740363	-0.1562
B121	110	17.8	98	0.616667	1	0	2.041393	1.991226	-0.20995
B122	121	9.1	42	0.509804	1	1	2.082785	1.623249	-0.2926
B127	91	16.6	62	0.623853	1	0	1.959041	1.792392	-0.20492
B129	105	27.2	59	1	1	1	2.021189	1.770852	0
B130	144	7	84	0.366412	1	1	2.158362	1.924279	-0.43603
B132	120	7.96	74	0.568421	1	1	2.079181	1.869232	-0.24533
B133	133	7.8	57	0.552381	1	1	2.123852	1.755875	-0.25776
B134	93	20.3	41	0.40708	1	1	1.968483	1.612784	-0.39032
B135	125	24.5	50	0.681319	1	1	2.09691	1.69897	-0.16665
B136	120	18	77	0.431818	1	1	2.079181	1.886491	-0.3647
B137	145	19.4	48	0.381679	1	1	2.161368	1.681241	-0.4183
B138	123	14.9	77	0.423077	x	1	2.089905	1.886491	-0.37358

Table 5.8 Cont.

B140	153	19.2	51	0.528	x	0	2.184691	1.70757	-0.27737
B144	105	16.3	68	0.478992	1	1	2.021189	1.832509	-0.31967
B147	126	16.3	65	0.444444	1	1	2.100371	1.812913	-0.35218
B151	109	37.7	56	0.456	0	0	2.037426	1.748188	-0.34104
B154	103	10.5	67	0.54321	x	0	2.012837	1.826075	-0.26503
B155	85	1.55	98	0.764228	1	1	1.929419	1.991226	-0.11678
B156	120	11.3	72	0.578947	0	1	2.079181	1.857332	-0.23736
B157	134	8.9	89	0.486957	0	1	2.127105	1.94939	-0.31251
B158	108	12.3	59	0.55102	x	1	2.033424	1.770852	-0.25883
B159	96	12.1	66	0.513761	1	1	1.982271	1.819544	-0.28924
B160	131	12.7	81	0.544643	x	1	2.117271	1.908485	-0.26389
B161	121	1.8	62	0.486486	0	1	2.082785	1.792392	-0.31293
B162	111	18.3	50	0.731092	1	1	2.045323	1.69897	-0.13603
B163	109	11.6	68	0.955056	0	1	2.037426	1.832509	-0.01997
B164	177	0	80	0.333333	0	1	2.247973	1.90309	-0.47712
B165	150	7.45	59	0.493827	x	1	2.176091	1.770852	-0.30643
B166	168	12	90	0.583333	0	1	2.225309	1.954243	-0.23408
B167	161	22.1	66	0.573034	1	1	2.206826	1.819544	-0.24182
B168	105	10.7	81	0.527778	1	1	2.021189	1.908485	-0.27755
B169	126	8	88	0.331034	1	1	2.100371	1.944483	-0.48013
B173	108	17.4	57	0.809524	1	1	2.033424	1.755875	-0.09177
B175	77	15.5	39	0.626016	1	1	1.886491	1.591065	-0.20341
B176	112	14.2	61	0.540541	1	1	2.049218	1.78533	-0.26717
B178	123	24.4	94	0.989474	1	1	2.089905	1.973128	-0.0046
B180	104	17.7	52	0.51087	0	1	2.017033	1.716003	-0.29169
B181	126	16.3	42	0.557522	x	1	2.100371	1.623249	-0.25374
B183	117	19.2	62	0.441176	1	1	2.068186	1.792392	-0.35539
B184	137	42.8	45	0.890909	0	1	2.136721	1.653213	-0.05017
B186	121	20.1	55	0.366667	0	1	2.082785	1.740363	-0.43573
B187	119	14.9	74	0.535714	0	1	2.075547	1.869232	-0.27107
B191	119	10	61	0.664179	1	1	2.075547	1.78533	-0.17771
B196	130	14.2	66	0.650943	1	1	2.113943	1.819544	-0.18646

Table 5.8 Cont.

B198	124	13.5	59	0.666667	1	1	2.093422	1.770852	-0.17609
B199	162	4.56	80	0.698413	1	1	2.209515	1.90309	-0.15589
B200	147	0	62	0.616162	0	1	2.167317	1.792392	-0.21031
B204	120	3.42	58	0.333333	x	1	2.079181	1.763428	-0.47712
B205	103	20.1	84	0.76087	1	0	2.012837	1.924279	-0.11869
B207	144	23.9	53	0.548077	1	1	2.158362	1.724276	-0.26116
B208	137	37.2	67	0.446281	1	1	2.136721	1.826075	-0.35039
B209	111	7.35	59	0.528	1	1	2.045323	1.770852	-0.27737
B210	134	4.01	72	0.606897	x	1	2.127105	1.857332	-0.21689
B211	135	76	84	0.512605	x	1	2.130334	1.924279	-0.29022
B212	103	11.2	55	0.970588	x	1	2.012837	1.740363	-0.01296
B213	157	16.45	41	0.506494	0	1	2.1959	1.612784	-0.29543
B214	132	4.7	77	0.864078	1	0	2.120574	1.886491	-0.06345
B215	132	3.57	57	0.409938	1	1	2.120574	1.755875	-0.38728
B216	123	12.2	58	0.477941	1	1	2.089905	1.763428	-0.32063
B217	131	29.4	65	0.618321	0	1	2.117271	1.812913	-0.20879
B218	95	14.1	54	0.393333	x	1	1.977724	1.732394	-0.40524
B219	100	12.8	55	0.647619	1	1	2	1.740363	-0.18868
B221	131	17.7	48	0.588652	0	0	2.117271	1.681241	-0.23014
B224	131	10.9	42	0.633588	x	1	2.117271	1.623249	-0.19819
B227	128	19.4	43	0.6	x	1	2.10721	1.633468	-0.22185
B228	131	1.33	50	0.702703	x	1	2.117271	1.69897	-0.15323
B230	89	3.57	51	0.55814	1	1	1.94939	1.70757	-0.25326
B232	97	22.44	78	0.546296	1	1	1.986772	1.892095	-0.26257
B234	125	33	57	0.328467	1	1	2.09691	1.755875	-0.48351
B237	120	13.6	59	0.463918	1	1	2.079181	1.770852	-0.33356
B238	109	24.2	58	0.537736	1	1	2.037426	1.763428	-0.26943
B243	152	37.5	88	0.804124	1	1	2.181844	1.944483	-0.09468
B244	111	20.6	54	0.621849	1	0	2.045323	1.732394	-0.20632
B245	101	24.3	57	0.533333	0	1	2.004321	1.755875	-0.273
B247	112	0	64	0.471545	0	1	2.049218	1.80618	-0.32648
B248	99	11.1	95	0.583333	1	1	1.995635	1.977724	-0.23408

Table 5.8 Cont.

B249	105	20	58	0.515873	1	1	2.021189	1.763428	-0.28746

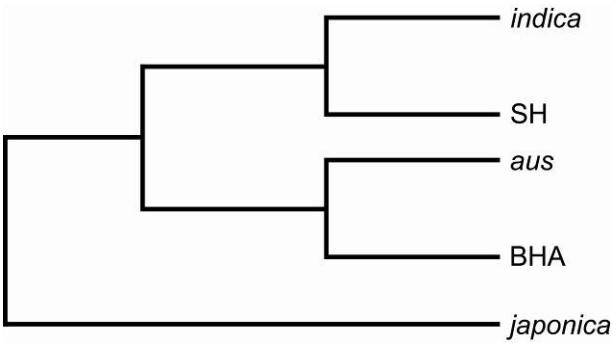


Figure 5.1. Relationships between cultivated and weedy rice groups.
In the US, SH weeds are more closely related to *indica* cultivars while BHA weeds are more closely related to *aus* cultivars. These weed groups are more distantly related to *japonica* cultivars, which are typically grown in the US fields. *Aus* and *indica* groups are believed to share a domestication origin, regardless of the number of domestications for Asian rice as a whole (Caicedo *et al.* 2007; Garris *et al.* 2005).

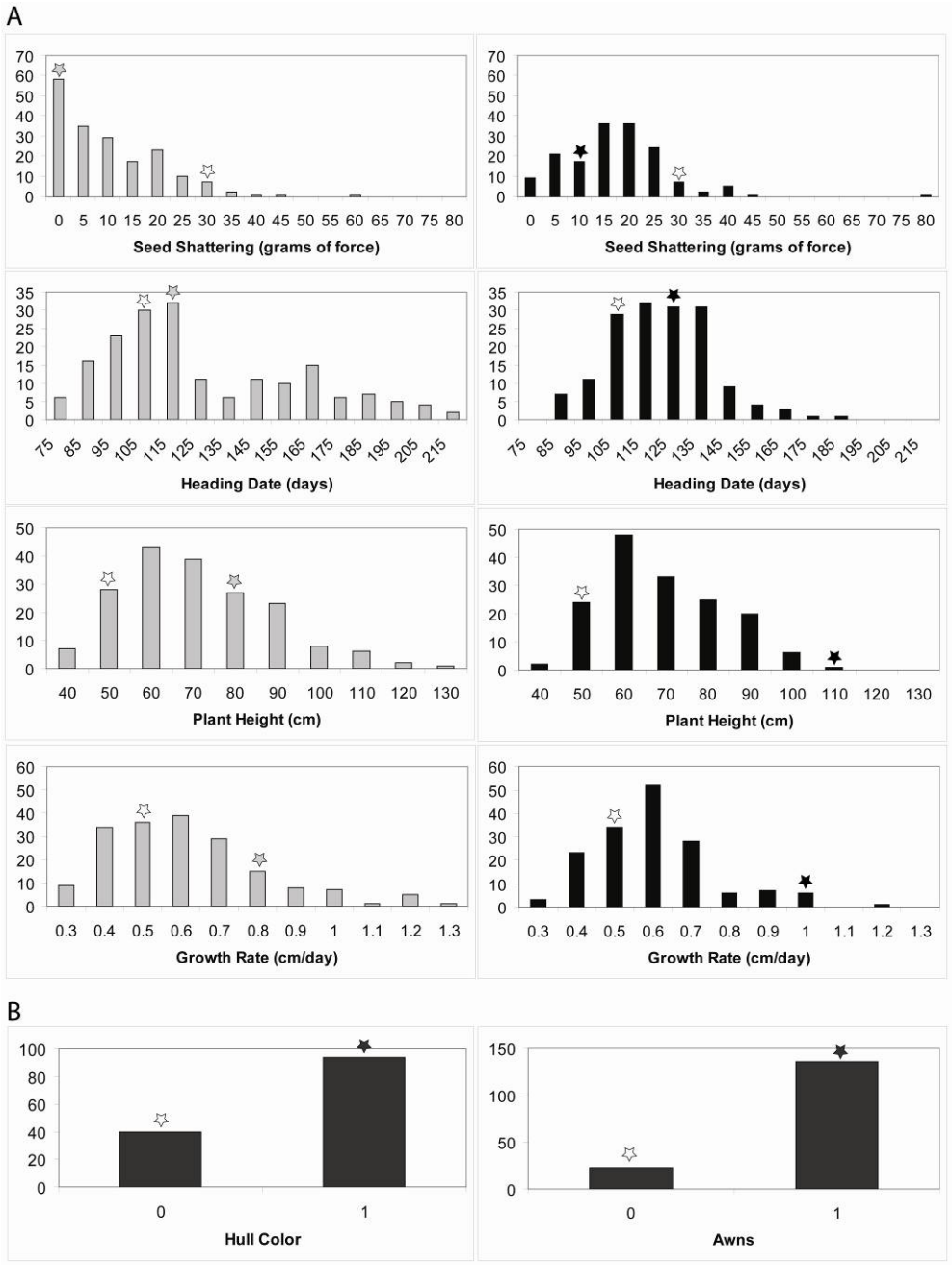


Figure 5.2. Frequency distributions of traits in the F_2 populations.
Grey bars represent traits measured in the S population, while black bars represent traits measured in the B population. The white stars correspond to trait values for the cultivated parent and the filled stars to the respective weed parent (SH_RR09 or BHA_RR20). The vertical axis of each figure represents the number of individuals.

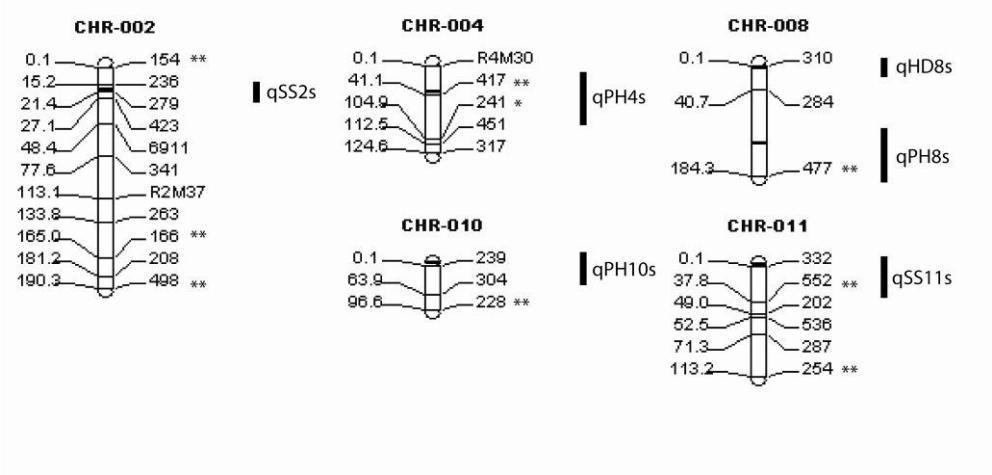


Figure 5.3. Molecular linkage map with positions of QTL for three traits in the S population.

Markers with segregation distortion are denoted with asterisks (* P<0.01, ** P<0.001). The length of the vertical line represents the 1.5 LOD confidence interval around the QTL peak. Only chromosomes with significant QTL are shown. Marker names are on the right side while marker positions drawn to scale are on the left.

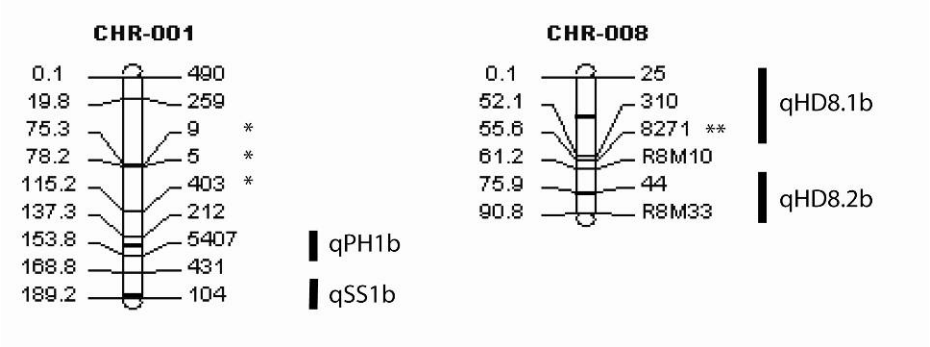


Figure 5.4. Molecular linkage map with positions of QTL for three traits in the B population.

Markers with segregation distortion are denoted with asterisks (* P<0.01, ** P<0.001). The length of the vertical line represents the 1.5 LOD confidence interval around the QTL peak. Only chromosomes with significant QTL are shown. Marker names are on the right while marker positions to scale are on the left.

CHAPTER 6

OVERALL CONCLUSIONS

6.1 Dissertation Conclusions

Weeds that colonize agricultural fields are of great interest from both a practical standpoint, as their presence often affects crop yields, and from an evolutionary standpoint, as little is known about how weeds evolve and adapt to a variety of environments. My research has increased our understanding of how weedy traits evolve and the genetic basis of convergence in weed-adaptive traits. Weedy rice is a major agronomic pest of cultivated rice and, as such, is in need of intense study. The two subpopulations of weedy rice studied here, SH and BHA, are likely de-domesticates of different varieties of cultivated rice (Reagon *et al*., 2010). It is possible that parallel evolution of weedy traits has occurred between these weedy rice subpopulations, due to their separate origins within cultivated rice followed by evolution under similar selective pressures in the US agroecosystem. My research furthers the understanding of the evolution of weedy rice by studying the relationship of different weedy rice populations to each other and investigating the genetic mechanisms by which weedy rice has acquired traits that have allowed it to spread and proliferate.

Seed shattering, or the easy release of seeds upon ripening, was a particularly interesting weedy trait. Nearly all weedy rice worldwide shatters its seeds while cultivated rice has been selected to retain its seeds through moderate levels of shatter resistance (Purugganan & Fuller, 2009). Of the many weed adaptive traits that differentiate weedy rice from its cultivated progenitors, seed shattering is one of the most important for characterizing the weedy phenotype. Efficient seed dispersal is likely

crucial to weed fitness, as it allows weeds to increase their presence in the seed bank and spread to new areas (Harlan & DeWet, 1965). My main research goals were to investigate the incidence of parallelism of seed shattering in US weedy rice while also investigating domestication-related candidate genes in order to assess where weedy alleles come from and determine if novel alleles exist in weedy rice. The seed shattering phenotype was investigated in both chapters 2 (extent of shattering ability) and 3 (shattering morphology and timing) of this thesis. With the exception of weedy rice with mixed ancestry (e.g. BRH/MX weeds; Reagon *et al.*, 2010), my results from chapter 2 show that all US weedy rice populations are highly shattering despite the range of shattering degree in the cultivated progenitors of US weeds. Additionally, all weedy rice shatters to a similar degree as wild rice, despite separate origins of major weedy rice groups, suggesting that this trait was strongly selected for during weedy rice evolution.

The flower-pedicel junction (where the base of the flower attaches to the panicle) is the site where seed release occurs after degradation of an abscission layer (Lin *et al.*, 2007; Oba *et al.*, 1995; Jin & Inouye, 1982; Jin & Inouye, 1985; Jin, 1986). In chapter 3 of this dissertation, I investigated the morphology of the abscission layer in weedy, cultivated, and wild rice and how this may affect the timing of seed shedding. Weedy rice develop abscission layers in the same location as wild rice, consistent with their shared shattering phenotype, yet the degradation of this layer is accelerated in weedy rice from both SH and BHA groups compared to wild rice. This accelerated degradation may also increase the weeds' shedding ability. Further investigation confirmed that some weedy rice individuals show an increase in seed shedding ability five days earlier than is typical in wild rice; however, some weedy rice individuals parallel the shattering ability

of wild rice, suggesting that shattering ability is not completely correlated with the rate at which the abscission layer degrades. The developmental differences between weedy and wild rice abscission layer traits further suggests that shattering in weedy rice was likely not acquired through introgression with wild rice.

In addition to phenotypic characterization, in chapter 2 of this dissertation I also investigated two major shattering loci in rice: *qsh1* and *sh4* (Konishi *et al.*, 2006; Li *et al.*, 2006b; Lin *et al.*, 2007). I found that all US weedy rice in my panel possess an ancestral *qsh1* allele that is common in all non-*temperate japonica* cultivated and wild rice groups, and is not correlated with loss of shattering outside of the *temperate japonica* clade. However, all US weedy rice accessions carry a single nucleotide substitution associated with non-shattering at *sh4*, and most weeds share an *sh4* haplotype with cultivated rice that appears to have been under strong selection, represented by high LD in genomic haplotypes surrounding this locus. *Sh4* has been identified as the most significant shattering gene to be selected on during domestication (Li *et al.* 2006b; Lin *et al.* 2007; Zhang *et al.* 2009). My identification of strongly selected upon alleles shared between weeds and cultivars supports the origin of US weeds from domesticated ancestors and suggests that this substitution, characteristic of cultivated *sh4* alleles, is not sufficient for reduction of shattering in all genetic backgrounds. Additionally, these data suggest that novel loci, potentially containing weed-specific mutations, are involved in the parallel evolution of shattering in both the SH and BHA weed groups as these weeds have evolved from closely related ancestors.

In addition to studying seed shattering, I also investigated the genetic mechanisms behind altered flowering times ("heading date") in weedy rice. The regulation of

flowering time in weeds is very important for increased competitive ability and local adaptation to various day length and temperature regimes (Greenup *et al.*, 2009, Sawers *et al.*, 2005). In our US weedy rice samples flowering time variation is the most defining growth related trait between weed groups and their cultivated ancestors (Reagon *et al.*, 2011). In chapter 4 of this dissertation, I quantified the heading date phenotype in weedy rice and found that the trait is not convergent between weedy rice groups but rather has diverged; SH weeds flower earlier and BHA weeds flower later than both local and Asian cultivated rice. That these two weed groups have evolved separate flowering phenotypes suggested to me that different mutations in major flowering time genes or potentially different genes might be playing a role in this divergence. Thus in chapter 4, I also investigated two important components of the flowering time gene network, *Hd1* and *Hd3a*. I found that at both loci weeds share haplotypes with their cultivated progenitors despite significantly different flowering times. However, only at the *Hd1* locus does haplotype significantly correlate with flowering time phenotype; at this locus BHA weeds share a common deletion resulting in photoperiod insensitivity and later flowering. As these genes only explain part of the flowering phenotype of the weeds, other genes must be involved that cause the difference in phenotype seen between weeds and progenitors. This was the motivating factor in including heading date in the QTL study done in chapter 5.

In order to further understand the incidence of parallel genetic evolution of seed shattering and flowering time in weedy rice and identify genomic regions that may contain novel candidate genes, I designed a QTL mapping study reported in chapter 5. For this study, two separate F_2 mapping populations were generated by crossing an *indica*

cultivar to a single accession from each weed population, SH and BHA. Neither the phenotypic distributions for shattering (normal in one mapping population and highly non-normal in the other) nor the locations of shattering QTL (chromosome 1 in one mapping population versus chromosomes 2 and 11 in the other) supported the hypothesis that these two weed parents share similar genetic mechanisms for the convergent shattering phenotype. More interestingly, the shattering locus on chromosome 2 does not appear to have been previously identified by any other QTL study involving crosses between cultivar types or between cultivars and wild *Oryza*. Additionally, QTL for seed shattering have also been identified in crosses between non-US weedy rice individuals and cultivated rice, yet my QTL do not overlap with either study, suggesting that parallel genetic evolution may not be the case across other weedy rice populations.

For flowering time, the phenotypic distributions were again different: normal in one mapping population and nearly bi-modal in the other. Since these weeds arose from different, but related, cultivar groups and posses different phenotypes I expected the genes involved in this trait to be different between the two groups. However, this seems a bit less likely given that QTL identified for flowering time in both populations are located in the same region of chromosome 8, coincident with a recently identified candidate gene, *Ghd8*.

My research altogether has shown that the genetics behind convergent and divergent weed traits does not always occur as predicted when weed populations are closely related. In the case of seed shattering, a trait that has evolved in parallel between the populations, different genes appear to be at work in the weed groups that also differ from the genes used during rice domestication. For flowering time, a trait that has

diverged between the groups, it is possible that the gene responsible for both phenotypes (early vs. late flowering) is the same but that the mutations within this gene, or the genes it interacts with, may be working towards opposite ends.

6.2 Future Work

In light of the results of my research, the evolution and origin of weedy traits in weedy rice is a complex matter that requires more research. In terms of seed shattering, further investigation at both the phenotypic and genetic levels is needed to fully understand this fascinating trait. At the phenotypic level, two avenues should be pursued: increased sampling of cultivated rice and finer scale developmental characterization. Although we tried to collect the broadest samples of cultivated rice possible, there is still the potential for missed phenotypic variation. Additional samples of landrace and older cultivars from *indica* and *aus* groups would greatly add to the understanding of the extent of the variation of this phenotype in the progenitors of weedy rice. These samples could then also be useful in identifying genes and alleles present in the standing variation of the weedy rice progenitor gene pool that may have led to shattering in weedy rice. Additionally, further investigation into the abscission layer formation and degradation at more floral and seed developmental stages may help identify more precisely when the abscission layer forms in weedy rice and how rapidly after formation it degrades. This rapid degradation of the abscission layer, potentially leading to earlier shattering, could be useful as a trait for future mapping and may more precisely locate the genes specific to the weedy rice shattering phenotype.

In order to learn more about parallel genetic evolution of shattering in *Oryza,* future work should also investigate other weedy rice populations worldwide. It is possible that these other weedy populations have evolved from separate cultivated rice groups (i.e. *tropical* or *temperate japonica*) or potentially through admixture of cultivars with wild rice in Asia. Some interesting questions that could be posed include: Do all weedy rice worldwide shatter to the same degree? Do they all posses the same morphological mechanisms for shattering? Are the same genes or even alleles/mutations involved in seed shattering despite the potential for a lack of shared ancestry with US weeds? Some research has already begun on these questions as several groups are working on weedy rice from different areas of the world. Two groups in particular have been mapping weed adaptive traits, including seed shattering, in non-US weeds (Bres-Patry *et al*., 2001; Gu *et al*., 2005a; Gu *et al*., 2005b; Gu *et al*., 2005c). Interestingly, my QTL for seed shattering do not overlap with any QTL from either study, suggesting that parallel genetic evolution may not be contributing to convergent phenotypes occurring worldwide.

Most recently, a group of Japanese researchers published a study using methods similar to those in chapter 3 of this thesis with weedy and cultivated rice samples found in Japan. Their shattering time course data shows an increase in shedding ability in their cultivars around 24 days after heading (DAH) while the weeds show their largest increase closer to 21 DAH (Akasaka *et al.*, 2011). Although their weeds also show an earlier increase in shedding ability compared to the cultivars they sampled, our cultivars and weeds showed a much earlier increase in shedding ability, by between 7 and 14 days, which suggests that, even within weedy and cultivated rice, variation is present.

Although they did not compare the weedy rice abscission layers to those of wild rice, these layers had formed and even started to become disorganized in preparation for separation as early as 3 DAH (Akasaka *et al.*, 2011). Further investigation into the abscission layers of individuals from other populations of weedy rice from Asia would help to figure out if parallel evolution is occurring among more distantly related weed groups.

The most important next step is to fine map both the shattering and heading date QTL reported here. Two of the most cutting edge tools and materials available to us at present are 1) a set of RILs derived from a subset of my F_2 populations and 2) whole genome sequence data for the three parents used to create my mapping populations (Jia, Caicedo & Olsen, unpublished; Hyma & Caicedo, unpublished). These genomes will generate many more markers that can be tested on the RILs in multiple environments and will not only give us a chance to identify more loci, possibly of weaker effect, but also to narrow down the regions of the genome containing genes of interest to a hopefully manageable number. It is even possible that some of the smaller effect shattering QTL could overlap between populations.

Lastly, to further understand the divergence of heading date phenotypes at the molecular level it would be prudent to investigate weedy rice alleles at the candidate gene *Ghd8*, the potential underlying cause of my heading date QTL. It is possible that this gene harbors alleles specific to each weed population and may show evidence of selection in weedy rice.

6.3 Broader Impacts

My results greatly add to the understanding of weedy rice as a dynamic and diverse set of populations. Some of these populations may well possess traits that can be beneficial to cultivated rice. Genes involved in heading date differences in weedy rice may be beneficial to further adapting cultivated rice to growing in the Southern US. More specifically, SH alleles conferring earlier flowering may be beneficial for shortening the growing season of rice in the US; this could allow for multiple rounds of planting in a single season, thus increasing the yield of a single field, or expanding the rice growing regions in the US by allowing rice to be planted in more northern latitudes with shorter summer seasons.

Although high shattering is not desired in the crop, each of the genes controlling shattering, even those found in weedy rice, could be used to further adjust the low level of shattering to suit human needs for threshing. It has been suggested that a switch from hand threshing, where seeds were expected to be pulled off easily, to machine threshing, using combines and other equipment, perpetuated a further shift to more severe non-shattering in modern cultivars of rice. As machine threshing may not be possible in some areas of the world, farmers may still need access to high yielding and disease resistant varieties being developed in non-shattering rice but with an increase in the ease of seed removal. Utilization of weedy alleles for both heading date and shattering could be accomplished by breeding or genetic engineering of elite lines with alleles from weedy rice. Weedy and cultivated rice crosses are certainly possible; I produced some crosses for this thesis, and there are repeated reports of hybridization occurring in rice fields. By

using marker-assisted selection, linkage drag of undesirable weedy traits can be mitigated.

Additionally, my research adds to our knowledge of weed evolution and the incidences of parallel evolution in closely related lineages. It is probable that weeds overall do not always use the same genes to evolve weedy traits; yet it is possible that weed species, no matter how divergent, may still use similar genes, types of genes, or pathways to arrive at the same adaptive phenotype. In the case of seed shattering and increased seed dispersal, much is known about the ecology of weed seed dispersal but little is known about the genetics behind it. Several weeds of cultivated plants display increased seed shattering (see examples in Ellstrand *et al.*, 2010). In particular, a weedy form of cereal rye (*Secale cereale* L.) is the result of de-domestication of volunteer rye and possess a seed shattering phenotype similar to that of the wild rye species (Burger *et al.*, 2007). It would be interesting to see if the genes involved in weedy rye shattering are similar to those involved in shattering of weedy rice.

Parallel evolution utilizing similar genes and mutations across plant lineages has been demonstrated in some cases (e.g. flower color in independently evolving Ipomoea lineages; Streisfield & Rausher, 2009; Des Marais & Rausher, 2010; Smith & Rausher, 2011), yet it does not appear to be universal that a shared trait will have the same genetic basis. My work shows that, for the case of seed shattering in weedy rice from the US, the same genes are likely not contributing to the parallel evolution of this trait. Given that both weedy rice populations likely originated from two closely related but highly diverse subpopulations of cultivated rice and share a near identical shattering phenotype, it was expected that similar genes would play a major role in seed shattering. However, my

QTL study results do not support this idea, suggesting that multiple mutations in different genes can lead to similar levels of seed shattering. Only once the shattering loci have been identified and investigated in a survey of weedy, wild, and cultivated rice will we know the true extent of parallel genetic evolution in this system.

BIBLIOGRAPHY

Abebe T, Wise RP, and Skadsen RW (2009) Comparative transcriptional profiling established the awn as the major photosynthetic organ of the barley spike while the lemma and the palea primarily protect the seed. *The Plant Genome* 2, 247-259.

Akasaka M, Konishi S, Izawa T, and Ushiki J (2011) Histological and genetic characteristics associated with seed-shattering habit of weedy rice (*Oryza sativa* L.) from Okayama, Japan. *Breeding Science* 61, 168-173.

Ammiraju JS, Luo M, Goicoechea JL, *et al.* (2006) The Oryza bacterial artificial chromosome library resource: construction and analysis of 12 deep coverage large insert BAC libraries that represent the 10 genome types of the genus Oryza. *Genome Research* 16, 140-147.

Arendt J and Reznick D (2007) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *TRENDS ECOL EVOL* 23, 26- 32.

Asano K, Takashi T, Miura K, *et al.* (2007) Genetic and Molecular Analysis of Utility of *sd1* Alleles in Rice Breeding. *Breeding Science* 57, 53-58.

Ayeh KO, Lee Y, Ambrose MJ, *et al*. (2009) Characterization and structural analysis of wild type and a non-abscission mutant at the development funiculus (Def) locus in *Pisum sativum* L. *BMC Plant Biology* 9:76-83

Basu C, Halfhill, MD, Mueller, TC and Stewart Jr., CN (2004) Weed genomics: new tools to understand weed biology. *TRENDS in Plant Science* 9:391-398.

Bernardo R (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from he Last 20 Years. *Crop Science* 48, 1649-1664.

Bhattacharjee P, Singhal RS, and Kulkarni PR (2002) Basmati rice: a review. *International Journal of Food Science and Technology* 37, 1-12.

Bollback JP and Huelsenbeck JP (2009) Parallel Genetic Evolution Within and Between Bacteriophage Species of Varying Degrees of Divergence. *Genetics* 181, 225-234.

Bres-Patry C, Lorieuz, M., Clement, G., Bangratz, M. and Ghesquiere, A. (2001) Heredity and genetic mapping of domestication-related traits in a temperate japonica weedy rice. *Theor Appl Genet* 102:118-126.

Burger JC, Holt JM, and Ellstrand NC (2007) Rapid phenotypic divergence of feral rye from domesticated cereal rye. *Weed Science* 55, 204-211.

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

Burgos NR, Norsworthy JK, Scott RC, and Smith KL (2008) Red rice (*Oryza sativa*) status after 5 years of imidazolinone-resistant rice technology in Arkansas. *Weed Technology* 22:200-208.

Cai H-Y, Diao S, He Y-G, *et al.* (2011) Genetic and physical mapping of qHY-8, a pleiotropic QTL for heading date and yield-related traits in rice. *Euphytica* DOI10.1007/s10681-10011-10581-10680.

Cai HW and Morishima, H (2000) Genomic regions affecting seed shattering and seed dormancy in rice. *Theor Appl Genet* 100:840-846.

Cai HW and Morishima H (2002) QTL clusters reflect character associations in wild cultivated rice. *Theor Appl Genet* 104, 1217-1228.

Caicedo AL and Purugganan MD (2005) Comparative Plant Genomics. Frontiers and Prospects. *Plant Physiology* 138, 545-547.

Caicedo AL, Williamson SH, Hernandez RD, *et al.* (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genetics* 3:e163.

Cao Q, Lu, B.-R., Xia, H. *et al.* (2006) Genetic Divsersity and Origin of Weedy Rice (*Orzya sativa f. spontanea*) Populations Found in North-eastern China Revealed by Simple Sequence Repeat (SSR) Markers. *Annals of Botany* 98:1241-1252.

Chen X, Temnykh S, Xu Y, Cho YG, and McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (Oryza sativa L.). *Theor Appl Genet* 95, 553-567.

Cho SK, Larue CT, Chevalier D, *et al*. (2008) Regulation of floral organ abscission in *Arabidopsis thaliana*. *PNAS* 105:15629-15634.

Counce PA, Keisling TC, and Mitchell AJ (2000) A Uniform, Objective, and Adaptive System for Expressing Rice Development. *Crop Science* 40:436-443.

Cui J, Fan S, Shao T, *et al.* (2007) Characterization and fine mapping of the *ibf* mutant in rice. *Journal of Integrative Plant Biology* 49, 678-685.

Delouche JC, Burgos, N.R., Gealy, D.R., *et al.* (2007) Weedy rices- origin, biology, ecology and control (ed. FAO). FAO, Rome, Italy.

Des Marais DL and Rausher MD (2010) Parallel evolution at multiple levels in the origin of hummingbird pollinated flowers in *Ipomoea*. *Evolution* 64, 2044-2054.

Doi K, Izawa T, Fuse T, *et al.* (2004) Ehd1, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of Hd1. *Genes and Development* 18, 926-936.

Dung LV, Inukai T, and Sano Y (1998) Dissection of a major QTL for photoperiod sensitivity in rice: its association with a gene expressed in an age-dependent manner. *Theor Appl Genet* 97:714-720.

Ellstrand NC, Heredia SM, Leak-Garcia JA, *et al.* (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evolutionary Applications* 3, 494-504.

Fan C, Xing, Y., Mao, H. *et al*. (2006) GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 112:1164-1171.

FAOSTAT (2008) Food and Agricultural Organization of the United Nations. http://faostat.fao.org/.

Flint J and Mott R (2001) Finding the molecular basis of quantitative traits: successes and pitfalls. *Nature Reviews* 2:437-45.

Franks SJ, Sim S, and Weis AE (2007) Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *PNAS* 104, 1278-1282.

Fujino K, Wu J, Sekiguchi H, *et al.* (2010) Multiple introgression events surrounding the *Hd1* flowering-time gene in cultivated rice, *Oryza sativa L. Mol Genet Genomics* 284, 137-146.

Fuller DQ, Qin, L., Zheng, Y. *et al*. (2009) The Domestication Process and Domestication Rate in Rice: Spikelet Bases from the Lower Yangtze. *Science* 323:1607-1609.

Fuller D, Sato Y-I, Castillo C, *et al*. (2010) Consilience of genetics and archaeobotany in the entangled history of rice. *Archaeological and Anthropological Sciences* 2:115-131.

Garris AJ, Tai, T.H., Coburn, J., Kresovich, S. and McCouch, S. (2005) Genetic Structure and Diversity in *Oryza sativa L*. *Genetics* 169:1631-1638.

Gealy DR, Tai TH, and Sneller CH (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. *Weed Science* 50:333-339.

Gealy DR, Mitten, D.H. and Niel-Rutger, J. (2003) Gene Flow Between Red Rice (Oryza sativa) and Herbicide-Resistant Rice (O. sativa): Implications for Weed Management. *Weed Technology* 17:627-645.

Gealy DR (2005) Gene movement between rice (Oryza sativa) and weedy rice (Oryza sativa)-- a US temperate rice perspective. *PNAS* 99:11256-11259.

Gealy DR, Agrama, HA, and Eizenga, GC (2009) Exploring Genetic and Spatial Structure of US Weedy Red Rice (*Oryza sativa* L.) in Relation to Rice Relatives Worldwide. *Weed Science.* 57:627-643

Goff SA, Ricke D, Lan TH, *et al.* (2002) A draft sequence of the rice genome (Oryza sativa L. ssp japonica). *Science* 296, 92-100.

Greenup A, Peacock WJ, Dennis ES, and Trevaskis D (2009) The molecular biology of seasonal flowering-responses in Arabidopsis and the cereals. *Annals of Botany* 103, 1165-1172.

Gross BL, Reagon, M., Hsu, S.-C., Caicedo, A.L., Jia,Y., and Olsen, K.M. (2010) Seeing Red: The Origin of Grain Pigmentation in US Weedy Rice. *Molecular Ecology* 2:3380-3393.

Gu X-Y, Chen Z-X, and Foley ME (2003) Inheritance of Seed Dormancy in Weed Rice. *Crop Science* 43, 835-843.

Gu X-Y, Kianian, S.F. and Foley, M.E (2005a) Seed Dormancy Imposed by Covering Tissues Interrelates to Shattering and Seed Morphological Characters in Weedy Rice. *Crop Science* 45:948-955.

Gu X-Y, Kianian, S.F., Hareland, G.A., Hoffer, B.L. and Foley, M.E. (2005b) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theor Appl Genet* 110:1108-1118.

Gu X-Y, Kianian SF, and Foley ME (2005c) Phenotypic Selection for Dormancy Introduced a Set of Adaptive Haplotypes From Weedy Into Cultivated Rice. *Genetics* 171, 695-704.

Gu X-Y, Turnipseed, E.B. and Foley, M.E (2008) The qSD12 Locus Controls Offpring Tissue-Imposed Seed Dormancy in Rice. *Genetics* 179:2263-2273.

Haldane JBS and Waddington CH (1931) Inbreeding and linkage. *Genetics* 16, 357-374.

Harlan JR and DeWet, J.M. (1965) Some thoughts about weeds. *Economic Botany* 19:16-24.

Harlan JR (1992) *Crops & Man* American Society of Agronomy, Madison, Wisconsin.

Hillis DM, Mable BK, and Moritz C (1996) Applications of molecular systematics: the state of the field and a look to the future. In: *Molecular systematics* (eds. Hillis DM, Moritz C, Mable BK). Sinauer Associates, Inc., Sunderland, MA.

Hodin J (2000) Plasticity and Constraints in Development and Evolution. *Mol Dev Evol* 288, 1-20.

Hoekstra HE and Nachman MW (2003) Different genes underlie adaptive melanism in different populations of rock pocket mice. *Molecular Ecology* 12:1185-1194.

Hu GL, Zhang DL, Pan HQ, *et al.* (2011) Fine mapping of the awn gene on chromosome 4 in rice by association and linkage analyses. *Chinese Science Bulletin* 56, 835-839.

Ji H-S, Chu, S.-H., Jiang, W. *et al*. (2006) Characterization and Mapping of a Shattering Mutant in Rice That Corresponds to a Block of Domestication Genes. *Genetics* 173:995-1005.

Ji H, Kim S-R, Kim Y-H, *et al*. (2010) Inactivation of the CTD phosphatase-like gene *OsCPL1* enhances the development of the abscission layer and seed shattering in rice. *The Plant Journal* 61:96-106.

Jin I-D and Inouye J (1982) Relation between grain shedding and pedicel morphology near the abscission layer of japonica-indica hybrid rices bred in Korea. *Jpn J Crop Sci* 51:271-275.

Jin I-D and Inouye J (1985) On the degree of grain shedding, histological peculiarity of abscission region and esterase isozyme genotype of Bulu and Tjereh rice varieties originated in Indonesia. *Jpn J Crop Sci* 54:373-378.

Jin I-D (1986) On the formation and development of abscission layer in rice plants, *Oryza sativa L*. *Jpn J Crop Sci* 55:451-457.

Khush GS (1997) Origin, dispersal, culitvation and variation of rice. *Plant Molecular Biology* 35, 25-34.

Kinglsey EP, Manceau M, Wiley CD, and Hoekstra HE (2009) Melanism in Peromyscus is Caused by Independent Mutations in Agouti. *PLoS One* 4:e6435.

Kojima S, Takahashi Y, Kobayashi Y, *et al.* (2002) Hd3a, a Rice Ortholog of the Arabidopsis FT Gene, Promotes Transition toFlowering Downstream of Hd1 under Short-Day Conditions. *Plant Cell Physiology* 43, 1096-1105.

Konishi S, Izawa, T., Lin, S.Y. *et al*. (2006) An SNP Caused Loss of Seed Shattering During Rice Domestication. *Science* 312:1392-1396.

Lee S-J, Oh C-S, Suh J-P, McCouch SR, and Ahn S-N (2005) Identification of QTLs for domestication-related and agronomic traits in an Oryza sativa X O. rufipogon BC1F7 population. *Plant Breeding* 124, 209-219.

Li C, Zhou A, and Sang T (2006a) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytologist* 170, 185-194.

Li C, Zhou A, and Sang T (2006b) Rice Domestication by Reducing Shattering. *Science* 311, 1936-1939.

Li S-B, Zhang Z-H, Hu Y, *et al.* (2006c) Genetic dissection of developmental behavior of crop growth rate and its relationships with yield and yield related traits in rice. *Plant Science* 170, 911-917.

Lin Z, Griffith, M.E., Li, X. *et al*. (2007) Origin of seed shattering in rice (*Oryza sativa L*.). *Planta* 226:11-20.

Londo JP, Chiang YC, Hung KH, Chiang TY, and Schaal BA (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *PNAS* 103:9578-9583.

Londo J.P., and Schaal, B.A. (2007) Origins and population genetics of weedy red rice in the USA. *Molecular Ecology* 16:4523-4535.

Luan W, Chen H, Fu Y, *et al.* (2009) The effect of the crosstalk between photoperiod and temperature on the heading-date in rice. *PLoS One* 4, e5891.

Mather KA, Caicedo AL, Polato NR, *et al.* (2007) The extent of linkage disequilibrium in rice (*Oryza sativa L*.). *Genetics* 177:2223-2232.

McCouch SR, Teytelman L, Xu Y, *et al.* (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa L*.). *DNA Research* 9, 199-207.

Mollenhauer HH (1964) PLASTIC EMBEDDING MIXTURES FOR USE IN ELECTRON MICROSCOPY. *Stain Technology* 39:111-114.

Monna L, Kitazawa N, Yoshino R, *et al.* (2002) Positional cloning of rice semidwarfing gene, sd-1: rice "green revolution gene" encodes a mutant enzyme involved in gibberellin synthesis. *DNA Research* 9, 11-17.

Nachman MW, Hoekstra HE, and D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *PNAS* 100:5268-5273.

Nadeau NJ and Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *TRENDS in Genetics* 26, 484-492.

Neff MM, Turk E., and Kalishman, M. (2002) Web-based Primer Design for Single Nucleotide Polymorphism Analysis. *TRENDS in Genetics* 18:613-615.

Nei M and Li, WH. (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 76:5269-5283.

Noldin JA, Chandler, J.M. and McCauley, G.N. (1999) Red Rice (Oryza sativa) Biology. I. Characterization of Red Rice Ecotypes. *Weed Technology* 13:12-18.

Oard J, Cohn, M.A., Linscombe, S., Gealy, D., and Gravois, K. (2000) Field evaluation of seed production, shattering, and dormancy in hybrid populations of transgenic rice (*Oryza sativa*) and the weed, red rice (*Oryza sativa*). *Plant Science* 157:13-22.

Oba S, Sumi N, Fujimoto F, and Yasue T (1995) Association between Grain Shattering Habit and Formation of Abscission Layer Controlled by Grain Shattering gene *sh-2* in Rice (*Oryza sativa L.*). *Jpn J Crop Sci* 64:607-615.

O'Brien TP, Feder N, and McCully ME (1964) Polychromatic Staining of Plant Cell Walls by Toluidine Blue O. *Protoplasma* 59:367-373

Oka HI (1988) *Origin of Cultivated Rice.* Tokyo: Japan Scientific Societies Press and Elsevier Science Publishers.

Olsen KM, Caicedo AL, Polato N, *et al.* (2006) Selection under domestication: Evidence for a sweep in the rice Waxy genomic region. *Genetics* 173:975-983.

Olsen KM, Caicedo, A.L. and Jia, Y. (2007) Evolutionary Genomics of Weedy Rice in the USA. *Journal of Integrative Plant Biology* 49:811-816.

Onishi K, Takagi, K., Kontani, M., Tanaka, T. and Sano, Y. (2007) Different patterns of genealogical relationships found in the two major QTLs causing reduction of seed shattering during rice domestication. *Genome* 50:757-766.

Panaud O, Chen X, and McCouch S.R. (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (Oryza sativa L.). *Molecular & General Genetics* 252, 597-607.

Paterson AH, Freeling M, and Sasaki T (2005) Grains of knowledge: Genomics of model cereals. *Genome Research* 15, 1643-1650.

Patterson SE. (2001) Cutting Loose. Abscission and Dehiscence in Arabidopsis. *Plant Physiology* 126:494-500.

Purugganan MD and Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature* 457:843-848.

Reagon M, Thurber, C.S., Gross, B.L., Olsen, K.M., Jia, Y., and Caicedo, A.L. (2010) Genomic patterns of nucleotide diversity in parallel populations of US weedy rice. *BMC Evolutionary Biology*. 10:180.

Reagon M, Thurber CS, Olsen KM, Jia Y, and Caicedo AL (2011) The long and the short of it: SD1 polymorphism and the evolution of growth trait divergence in US weedy rice. *Molecular Ecology* 20, 3743-3756.

Roberts JA, Whitelaw CA, Gonzalez-Carranza ZH, *et al*. (2000) Cell Separation in Plants-- Models, Mechanisms and Manipulation. *Annals of Botany* 86:223-235.

Rompler H, Rohland N, Lalueza-Fox C, *et al.* (2006) Nuclear gene indicates coat-color polymorphism in Mammoths. *Science* 313, 62.

Rozas J, Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003) DnaSP,DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.

Rozen S and Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (ed. Krawetz S MS), pp. 365-386. Humana Press, Totowa, NJ.

Sabeti PC, Reich, D.E., Higgins, J.M. *et al*. (2002) Detecting recent positive selection in the human genome from haloptype structure. *Nature* 419:832-837.

Sahli HF, Conner JK, Shaw FH, Howe S, and Lale A (2008) Adaptive differentiation of quantitative traits in the globally distributed weed, wild radish (*Raphanus raphanistrum*). *Genetics* 180, 945-955.

Sang T and Ge, S. (2007a) The Puzzle of Rice Domestication. *Journal of Integrative Plant Biology* 49:760-768.

Sang T and Ge, S. (2007b) Genetics and phylogenetics of rice domestication. *Current Opinion in Genetics & Development* 17:533-538.

Sang T. (2009) Genes and Mutations Underlying Domestication Transitions in Grasses. *Plant Physiology* 149:63-70.

Sasaki A, Ashikari M, Ueguchi-Tanaka M, *et al.* (2002) A mutant gibberellin-synthesis gene in rice. *Nature* 416, 701-702.

Sawers RH, Sheehan MJ, and Brutnell TP (2005) Cereal phytochromes: targets of selection, targets for manipulation? *TRENDS in Plant Science* 10, 138-143.

Schluter D, Clifford, E.A., Nemethy, M., and McKinnon, J.S. (2004) Parallel Evolution and Inheritance of Quantitative Traits. *American Naturalist* 163:809-822.

Schwinn K, Venail J, Shang Y, *et al.* (2006) A small family of MYB-regulatory genes controls floral pigmentation intensity and patterning in the genus *Antirrhinum*. *Plant Cell* 18, 831-851.

Septiningsih EM, Prasetiyono J, Lubis E, *et al.* (2003) Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the Oryza sativa variety IR64 and the wild relative O. rufipogon. *Theor Appl Genet* 107, 1419-1432.

Shen Y-J, Jiang H, Jin J-P, *et al.* (2004) Development of Genome-Wide DNA Polymorphism Database for Map-Based Cloning of Rice Genes. *Plant Physiology* 135, 1198-1205.

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

Shivrain VK, Burgos NR, Agrama HA, *et al.* (2010a) Genetic diversity of weedy red rice (*Oryza sativa*) in Arkansas, USA. *Weed Research* 50, 289-302.

Shivrain VK, Burgos NR, Scott RC, *et al.* (2010b) Diversity of weedy red rice (*Oryza sativa L.*) in Arkansas, USA. in relation to weed management. *Crop Protection* 29:721-730.

Smith SD and Rausher MD (2011) Gene loss and parallel evolution contribute to species difference in flower color. *Mol. Biol. Evol.* 28, 2799-2810.

Spielmeyer W, Ellis MH, and Chandler PM (2002) Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *PNAS* 99, 9043-9048.

Steiner CC, Rompler H, Boettger LM, Shoneberg T, and Hoekstra HE (2008) The genetic basis of phenotypic convergence in beach mice: Similar pigment patterns but different genes. *Mol. Biol. Evol.* 26, 35-45.

Stephens M, Smith, N.J., and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 68:978-989.

Stephens M and Scheet, P. (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics* 76:449-462.

Stinchcombe JR and Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100, 158-170.

Streisfeld M and Rausher MD (2009) Genetic changes contributing to the parallel evolution of red floral pigmentation among *Ipomoea* species. *New Phytologist* 183, 751-763.

Sweeney M and McCouch S (2007) The complex history of the domestication of rice. *Annals of Botany* 100:951-957.

Tabuchi T, Ito S, and Arai N (2001) Anatomical Studies of the Abscission Process in the Tomato Pedicels at Flowering Stage. *J Japan Soc Hort Sci* 70:63-65.

Tajima F (1989) Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123:585-595.

Takahashi Y, Teshima KM, Yokoi S, Innan H, and Shimamoto K (2009) Variations in Hd1 proteins, Hd3a promoters, and Ehd1 expression levels contribute to diversity of flowering time in cultivated rice. *PNAS* 106, 4555-4560.

Tamaki S, Matsuo S, Wong HL, Yokoi S, and Shimamoto K (2007) Hd3a protein is a mobile flowering signal in rice. *Science* 316, 1033-1036.

Tamura K, Dudley J, Nei M, and Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-1599.

Tan WD, Gan FF, and Chang TC (2004) Using normal quantile plot to select an appropriate transformation to achieve normality. *Computational Statistics & Data Analysis* 45, 609-619.

Thomson MJ, Tai, T.H., McClung, *et al*. (2003) Mapping quantitative trait loci for yield, yield components and morphological traits in an advanced backcross population between *Oryza rufipogon* and the *Oryza sativa* cultivar Jefferson. *Theor Appl Genet* 107:479-493.

Thurber CS, Hepler P, and Caicedo AL (2011) Timing is Everything: Early Degradation of Abscission Layer May Lead to Increased Shattering in US Weedy Rice. *BMC Plant Biology* 11, 14.

Thurber CS, Reagon M, Gross BL, *et al*. (2010) Molecular evolution of shattering loci in US weedy rice. *Molecular Ecology* 19:3271-3284.

Tilquin P, Coppieters W, Elsen JM, *et al.* (2001) Statistical power of QTL mapping methods applied to bacteria counts. *Genetical Research* 78, 303-316.

Tsuji H, Taoka K-I, and Shimamoto K (2010) Regulation of flowering in rice: two florigen genes, a complex gene network and natural variation. *Current Opinion in Plant Biology* 14, 1-8.

Van Ooijen JW (1999) LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83, 613-624.

Vaughan LK, Ottis, B.V., Prazak-Havey, A.M., *et al.* (2001) Is all red rice found in commercial rice really *Oryza sativa*? *Weed Science* 49:468-476.

Vaughan DA, Lu, B-R., and Tomooka, N. (2008) The evolving story of rice evolution. *Plant Science* 174:394-408.

Vavilov NI (1922) The law of homologous series in variation. *J Genet.* 12:47-89.

Wang S, Basten, C.J., and Zeng, Z.-B. (2011) Windows QTL Cartographer 2.5, Raleigh, NC.

Weber E and Schmid B (1998) Latitudinal population differentiation in two species of Solidago (Asteraceae) introduced in Europe. *American Journal of Botany* 85, 1110-1121.

Wei X, Xu J, Guo H, *et al.* (2010) DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology* 153, 1747-1758.

Wing RA, Kim HR, Goicoechea JL, *et al.* (2005) The Oryza map alignment project (OMAP): a new resource for comparative genomics studies within Oryza. *Plant Molecular Biology* 59, 53-62.

Wittkopp PJ, Williams BL, Selegue JE, and Carroll SB (2003) Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. *PNAS* 100, 1808-1813.

Xiao, J., Li, J., Grandillo, S. *et al*. (1998) Identification of Trait-Improving Quantitative Trait Loci Alleles From a Wild Rice Relative, *Oryza rufipogon*. *Genetics* 150:899-909.

Xing YZ, Tang, W.J., Xue, W.Y., Xu, C.G., and Zhang, Q. (2008) Fine mapping of a major quantitative trait loci, qSSP7, controlling the number of spikelets per panicle as a single Mendelian factor in rice. *Theor Appl Genet* 116:789-796.

Xiong LZ, Liu, K.D., Dai, X.K., Xu, C.G., and Zhang, Q. (1999) Identification of genetic factors controlling domestication-related traits of rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*. *Theor Appl Genet* 98:243-251.

Xu NW, Xu S, and Ehlers J (2009) Estimating the Broad-Sense Heritability of Early Growth Cowpea. *International Journal of Plant Genomics*, doi:10.1155/2009/984521.

Yan W-H, Wang P, Chen H-X, *et al.* (2011) A major QTL, Ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant* 4, 319-330.

Yang Y, Klejnot J, Yu X, Liu X, and Lin C (2007) Florigen (II): It is a mobile protein. *Journal of Integrative Plant Biology* 49, 1665-1669.

Yano M, Harushima Y, Nagamura Y, *et al.* (1997) Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theor Appl Genet* 95, 1025-1032.

Yano M, Katayose Y, Ashikari M, *et al.* (2000) Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. *The Plant Cell* 12, 2473-2483.

Yoon H-S and Baum DA (2004) Transgenic study of parallelism in plant morphological evolution. *PNAS* 101, 6524-6529.

Yu S, Li JX, Xu CG, *et al.* (2002) Identification of quantitative trait loci and epistatic interactions for plant height and heading date in rice. *Theor Appl Genet* 104, 619-625.

Zhang G and Mergoum M (2007) Molecular mapping of kernel shattering and its association with Fusarium head blight resistance in a Sumai3 derived population. *Theor Appl Genet* 115:757-766.

Zhang L, Wang S, Li H, *et al.* (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. *Theor Appl Genet* 121, 1071-1082.

Zhang L-B, Zhu, Q., Wu, Z-Q. *et al*. (2009) Selection on grain shattering genes and rates of rice domestication. *New Phytologist*. 184:708-720.

Zheng X-M and Ge S (2010) Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Molecular Ecology* 19:2439-2454.

Zhu B-F, Si L, Wang Z, *et al.* (2011) Genetic Control of a Transition from Black to Straw-White Seed Hull in Rice Domestication. *Plant Physiology* 155, 1301-1311.

Zhu QH and Ge S (2005) Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *New Phytologist* 167:249-265.