EVOLUTIONARY ORIGINS AND POPULATION GENETICS OF RED RICE IN THE SOUTHERN UNITED STATES

by

ELEANOR JOHANNA KUNTZ

(Under the Direction of Rodney Mauricio)

ABSTRACT

Domestication has generally been used as a powerful exemplar of the power of selection to create novel phenotypes in short amounts of time. Reversion of cultivated species to an ancestral state is a less well-understood phenomenon and has been suggested as being the process responsible for the evolution of red rice in the southern United States. Field-collected red rice populations are highly morphologically and genetically diverse, forming a distinct taxonomic group. Contrary to findings suggesting red rice seed morphology is indicative of evolutionary history, we find no association with seen phenotype and genetic assignment, as well as evidence for mixed ancestry, with four distinct genetic groups co-occurring across the region. Our results provide strong support for multiple hybrid origins of red rice in Asia with subsequent dispersal to the United States and no support for a "de-domestication" event.

Understanding the introduction history, including source population localities and the dynamics of invasion is important in order to generate hypotheses regarding the environmental and evolutionary factors responsible for the successful establishment and persistence of invading organisms. A complex

evolutionary history places the origin red rice in Asia, arising as the result of crop to weed hybridization between cultivated rice (*Oryza sativa*) and wild rice (*Oryza rufipogon*). A long and well documented history of weed and invasive plant movement between the United States and China, in addition to the reported existence of hybrid derived red rice populations in China lead us to investigate populations of Chinese red rice in an attempt to locate the source population(s) of US red rice. Our results provide no evidence for a Chinese origin in the population of red rice sampled. Direct comparisons between populations of red rice in the United States and China surprisingly show no genetic overlap. These comparisons also indicate that US populations are more diverse then their Chinese counterparts.

Gene flow is a general term used to describe mechanisms which move genetic information between individuals in the same population and among populations. Occurring through the movement of seeds and pollen, gene flow has the potential to act as a genetic bridge between individuals, populations and species. The balance between selfing and outcrossing, largely characterized by mating system, has a direct effect on the ability individuals to incorporate genetic variability into their populations. Red rice persists in sympatry with large numbers of conspecific cultivated congeners. Genetic admixture gained as the result of seed introduction and long flowering periods in red rice provide ample opportunity genetic exchange. Inter-specific gene flow in this system additionally opens the prospect of introgression of cultivated genetic material in the red rice, potentially creating novel, highly fit genotypes. We find direct evidence for crop to weed gene flow, suggesting that the introgression of cultivated alleles into weedy rice populations may be playing a role in the weeds persistence and survival. Out crossing rates comparable to other wild *Oryza* taxa imply that selection for the maintenance of a mixed-mating system is acting to increase diversity in red rice populations. Collectively, these results indicate that red rice has a complex evolutionary history that has lead to the generation of a diverse and dynamic weed complex.

INDEX WORDS: red rice, weed evolution, gene flow, invasive species, de-domestication

EVOLUTIONARY ORIGINS AND POPULATION GENETICS OF RED RICE IN THE SOUTHERN UNITED STATRES

by

ELEANOR JOHANNA KUNTZ

Bachelor of Arts, Smith Collage, 2001

A Dissertation to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

EVOLUTIONARY ORIGINS AND POPULATION GENETICS OF RED RICE IN THE SOUTHERN UNITED STATRES

by

ELEANOR JOHANNA KUNTZ

Major Professor: Rodney Mauricio

Committee: Jim Hamrick
Jeff Bennetzen
John Wares
John Burke
Jim Leebens-Mack

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2011

ACKNOWLEDGEMENTS

Foremost I would like to acknowledge and thank my advisor, Rodney Mauricio, for his unfailing support, enthusiasm, friendship, and sense of humor. His mentorship and teachings have been invaluable, and will continue to guide me as I continue my scientific endeavors. I am also grateful to the other members of my committee, Jim Hamrick, Jeff, Bennetzen, John Wares, John Burke, and Jim Leebens-Mack all of whom gave me much of their time and capabilities in an effort not only to direct and improve this dissertation, but also the scientist that I am today. I thank past and present members of the Mauricio and Wessler labs, in particular, Vanessa Koelling, Gina Baucom, Dawn Holligan, and Joan West, for their technical help, and friendships over the years. I also must thank Jason and Andrew Colvard for their invaluable assistance in the field and greenhouse; without them this dissertation would not have been possible. I also must thank the National Science Foundation, National Institutes of Health and the University of Georgia, Department of Genetics for the funding that made this research possible. And lastly, but certainly not least, I am grateful for the love and support of my family and friends.

TABLE OF CONTENTS

LIST OF TABLES

Page

Page

LIST OF FIGURES

Page

Page

CHAPTER I:

INTRODUCTION AND LITERATURE REVIEW

Domestication

Humans drastically alter their environments, changing natural processes and landscapes to suit their needs and desires. One of the most dramatic examples of humans' ability to affect change on their environment occurred during the transition from nomadic hunting and gathering driven existence to a sedentary, community based, agriculturally driven lifestyle (Harlan, 1992). The ability to move from nomadic to sedentary communities was dependent on the creation of an agricultural system. Most major crop species were domesticated prior to 4,000 years ago (Doebley et al., 2006). The process of domestication transformed these wild and often weedy progenitors into domesticated varieties through selection by humans on traits that were desirable to them. This collective suite of characteristics is often referred to as the 'domestication syndrome' (Hammer, 1984). When compared to their wild counterparts, crops generally have larger, fleshier fruits, more compact floral structures and seeds/fruits that remain on the plant facilitating collection (Doebley et al., 2006; Koinange et al., 1996). Humans have selected for greater control on the production of their food supply, creating domesticates which are annual, flower more synchronously, have often lost their photosensitivity and no longer have seed dormancy (Koinange et al., 1996; Tian et al., 2009). Further selection of plants with erect, compact and often dwarf stature have caused genetic modifications allowing for easier collection and growth in the agricultural setting (Doebley, 2004; Jaenicke-Despres et al., 2003). The most notable loss when moving from a wild to domesticated form is the loss of seed dispersal mechanisms, such as shattering and dispersal organs such as awns (Burger et al., 2008; Doebley, 2004; Doebley et al., 2006; Ross-Ibarra et al., 2007; Vaughan et al., 2007). Certain domesticated species are so altered from their ancestral form (such as maize and

broccoli) that they are unable to propagate themselves without humans, rendering them useless in the wild.

Oryza, O. sativa cultivation and importance

Rice is one of the most important food crops worldwide, acting as the major source of caloric intake for over one third of the world's population (Vaughan et al., 2003; Vaughan et al., 2008). In addition to providing a stable food supply, rice production is an important component of the economic stability of many Asian economies (Khush, 1997; Xiao et al., 1998). The genus *Oryza* is comprised of approximately 21 species, which are grouped based on chromosome number and genomic composition (Khush, 1997). Two species of rice, *Oryza sativa* and *Oryza glaberrima*, are cultivated, with *O. sativa* (subspecies *indica* and *japonica*) farmed extensively throughout Asia, Europe and the Americas. The remaining species comprising *Oryza* are widely distributed, occurring on four continents, with most of their geographic range located in tropical areas (Doi et al., 2000; Langevin et al., 1990; Vaughan et al., 2005). Members of *Oryza* are self compatible, with a reported outcrossing rate varying from 0-50% depending on the species (Chen et al., 2004; Li et al., 2009; Song et al., 2003). *Oryza sativa*, cultivated rice, is a member of the AA genome sub-group. Cultivated rice is primarily self-pollinating, with an average out crossing rate of 0.5% (ranging from 0-3.4%) for different cultivars (Adair et al., 1962). Outcrossing rates vary across the six wild species (*O. rufipogon*, *O. barthii*, *O. glaberrima*, *O. latifolia*, *O. longistamminata*, and *O. punctata*), that share the same diploid AA genome with the cultivated types; and thus have the potential to hybridize readily (Khush, 1997; Li et al., 2001). Many of these wild species exist in close proximity to the cultivated types and often persist as agricultural pests in actively cultivated and fallow fields (Gealy et al., 2002; Langevin et al., 1990; Londo and Schaal, 2007; Vaughan et al., 2008). The co-occurrence of these wild species and the cultivated rice is greatest in central Asia, where the ranges of many of the wild *Oryza* species overlap with areas of rice cultivation (Chen et al., 2004; Gealy et al., 2003).

Rice Domestication

Cultivated *Oryza sativa* is generally thought to have been domesticated from populations of the wild species *Oryza rufipogon* in central Asia (Chen et al., 2004; Rakshit et al., 2007; Yamanaka et al., 2003). *O. rufipogon* is found as far north as southern China and occurs throughout Southeast Asia with populations in India, Thailand, Indonesia and as far south as northern Australia (Lu et al 2003). *O. rufipogon* is commonly found along the banks of lakes and rivers, in disturbed habitats such as ditches along roadsides, and can often be found in marginal habitats in and around cultivated rice fields (Song et al., 2003; Vaughan et al., 2003). *O. rufipogon* is considered a major pest of rice agriculture in many locations in Southeast Asia, particularly in areas where traditional forms of rice cultivation are utilized (Chen et al., 2004). Both annual and perennial forms of *O. rufipogon* exist, often inhabiting the same populations. The annual form is often referred to as *O. nivara*, although species level differences between *O. rufipogon* (perennial forms) and *O. nivara* (annual forms) are not widely recognized (Vaughan et al., 2001b). *O. rufipogon* is known to experience episodic outcrossing with cultivated varieties of *O. sativa* in areas where they co-occur, with estimates of outcrossing rates varying from 1.2-5% across space and time (Song et al., 2004; Song et al., 2003).

Many distinct studies support the occurrence of multiple domestication events leading to the formation of two sub-species of cultivated rice; *O. sativa indica* and *O. sativa japonica* (Cheng et al., 2003; Li et al., 2006; Li et al., 2009; Londo et al., 2006; Prathepha, 2009). Phylogeography of the wild progenitor of rice, *O. rufipogon*, suggests two independent domestication events from geographically distinct populations. One domestication event is thought to have occurred in the southeastern portion of *O. rufipogon's* native range, most likely in eastern India, while the other is thought to have occurred west of the Himalayas, most likely in southern China (Londo et al., 2006). Multiple domestication events have had a substantial impact on the diversity seen between *O. sativa indica* and *japonica* cultivars and isolating barriers exist between the two sub-species such that it is difficult to successfully cross them (Li et al., 2001; Luo et al., 2001). Population genetic structure across historic populations of *O. rufipogon* is

evident when comparing allele sharing between *indica* and *japonica* cultivars, with *indica* cultivars being more similar to western *O. rufipogon* populations and *japonica* cultivars being more similar to eastern populations. Estimates of divergence between the two sub-species are much older (0.2-3 mya) than estimates of the length of time that rice has been domesticated (10,000 years), suggesting that historic differentiation and population structure across the progenitor *O. rufipogon's* native range has exaggerated estimates of divergence between the cultivated types (Vitte et al., 2004).

Red Rice, brief introduction

Red rice is the most economically and ecologically problematic weed in rice agriculture world-wide (Gealy et al., 2000; Gealy et al., 2003; Lawton-Rauh and Burgos, 2010). Although it is primarily a pest of rice cultivation, it is also known to invade other upland crops, such as: jute, maize and soybean (Baki et al., 2000). Traditional classification groups red rice with *O. sativa L*. due to their sexual compatibility (Vaughan et al., 2001b). However, red rice shows differences in mating system with higher levels of outcrossing and longer flowering compared to the almost exclusively selfing cultivated varieties in both greenhouse and field experiments (Chen et al., 2004; Song et al., 2004). In addition, red rice shows morphological diversity within and between populations for flowering date, seed dormancy, seed morphology (color, pubescence, grain size and awn presence), cold tolerance, herbicide resistance, and tillering capacity (Gealy et al., 2002). This has prompted some to consider red rice a distinct taxonomic unit (Vaughan et al., 2001b).

What makes a weed?

As with domestication syndromes, there are a suite of morphological characteristics indicative of weed morphology, including shattering, dispersal organs, seed dormancy and high seed fecundity. The most consistent characteristics possessed by weed populations is the ability to shatter, such that seeds can be released from the panicle and distributed into the environment; and seed dormancy, such that weed seeds can persist in the field multiple seasons forming a seed bank and only germinating when conditions

are favorable (Bossuyt and Hermy, 2001; Hopfensperger, 2007). Seed dormancy is common in wild plants, seen as a sort of evolutionary "bet-hedging' allowing plants to avoid germinating all of their progeny in the following season (Pake and Venable, 1996; Venable, 2007). This allows weeds to persist in fields through crop rotations, fallow periods and environmentally stressful years (such as drought, or extreme weather conditions).

Defining an organism as an invasive or a weedy pest is often a matter of semantics. For our purposes we consider invasive species those that persist outside their native range and actively invade natural habitats such that native flora are displaced. Weeds broadly defined are any plant found 'out of place' or in unwanted areas. Weeds are found exclusively within the agricultural environment, making them a specific sub-category of invasive plants which do not necessarily displace native flora, but invade a very specific environment, the agricultural field. The persistence of these plants in agricultural fields leads to yield loss through competition for space, nutrition and water.

Harlan (1975) conjectured that the agricultural environment, with its disturbed habitat of densely planted monoculture, has promoted the formation of weedy and invasive plants. He hypothesized three major mechanisms driving weed evolution. The first recognized general wild colonizing plants which were thought to specialize in disturbed environments much like successional species. The second recognized the formation of weeds via hybridization of crops with closely related wild species. This second mechanism is also referred to as exo-ferality. Through this process, crop X wild hybrids are able to enter the cultivated habitat, persist in the environment often times outcompeting cultivated types. The third mechanism of weed formation is endo-ferality, often referred to as de-domestication, a process by which feral types are generated from abandoned domesticates, or cultivars that generate 'off-types' Ellstrans 2010) (De Wet and Harlan, 1975; Ellstrand et al., 2010; Harlan and de Wet, 1965).

De-domestication

De-domestication or endo-ferality is a local phenomenon. With this mechanism of weed evolution local cultivars are the source for weed generation and as such the weedy populations that are generated can only be as diverse as the cultivated starting material. Low levels of diversity will be exacerbated by initial small population size. A classic example of de-domestication generating weed populations is seen in populations of feral rye in the western United States (Burger et al., 2007; Burger et al., 2006). Weedy rye populations have been problematic pests for over 100 years. Originally they were thought to have a hybrid origin resulting from a cross between domesticated rye (*Secale cereale*) and the wild mountain rye (*S. strictum*), but a closer investigation into the diversity seen in feral populations confirmed a de-domesticated origin resulting from the feralization of multiple cultivars (Burger et al., 2007). Feral rye became such a problematic pest in the 1960s that most rye cultivation was abandoned in the western United States.

Crop x wild relative hybridization

Hybridization between crops and related wild species offers a mechanism to generate populations of weedy plants that are more diverse than their cultivated counterparts. The formation of weedy plant populations as the result of crop to wild hybridization is known in many systems. When these hybridization events recur, or independently in geographically disparate regions and are brought together via seed movement the resulting hybrid weed populations can be quite diverse. The weedy pest, wild radish is the result of hybridization between cultivated radish (*Raphenus sativus*) and its close relative wild jointed charlock (*R. raphanistrum*) (Hegde et al 2006). In all areas where cultivated radish and wild jointed charlock co-occur hybridization is known to occur, often forming local hybrid swarms (Campbell et al., 2006; Snow and Campbell, 2005; Snow et al., 2001). Populations of hybrid plants have been seen in agricultural fields for the past 100 years along the west coast of the United States causing major yield loss. In the last 50 years the weedy wild radish has extended its range, moving from agricultural fields

into local habitats. Wild radish is known to invade coastal plains and disturbed inland habitats from southern Oregon to Baja California in Mexico. This transition, from weed to invasive illustrates the great damaging potential unmanaged weeds have, to not only destroy agriculture, but also to evolve invasive habits displacing local flora (Ellstrand et al., 2010; Schierenbeck and Ellstrand, 2009; Stewart et al., 2009; Warwick et al., 2009).

Crop mimics

Strong selection against weeds in the agricultural environment often will result in the formation of crop mimics. A classic example of a crop mimic is barnyard grass *(Echinochloa crus-galli* var. *oryzicola*), which evolved to mimic domesticated rice. Seedlings of barnyard grass, its progenitor (*E. crus-galli*) and cultivated rice (*O. sativa*) were grown in a common garden to evaluate the extent to which barnyard grass mimicked rice morphology (Barrett, 1983). Morphological characteristics throughout the life cycle reveled that barnyard grass did not significantly differ from *O. sativa* in any morphological measurements taken during the vegetative potion of their life cycles, although both barnyard grass and *O. sativa* differed statistically from *E. crus-galli* (the progenitor of barnyard grass). Crop mimicry is the result of strong selective pressure resulting from humans actively weeding obvious non-cultivar plants. Weeds that are indistinguishable from the cultivated types are able to escape removal and are allowed to persist (Bartsch et al., 1999; Ellstrand et al., 2010; Harlan, 1992; Warwick et al., 2009).

Red rice

Red rice is very morphologically diverse; a particular biotype is described by its collective combination of morphological seed characteristics; hull color (straw, brown, black), the presence or absence of an awn, pubescence, and seed color (Gealy et al., 2002). A wide range of phenotypes exist in the field, with morphologies ranging from crop mimics with straw colored hulls, no awns, tightly packed panicles and short stature to plants that look similar to the wild progenitor of rice, *O. rufipogon*, with black colored hulls, long awns, diffuse panicle architecture and tall stature. Seeds of red rice, as the name

implies, are most often red in color, but weedy plants with white pericarp tissue have also been found (personal observation).

Red rice is such a problematic pest of rice cultivation because of its close evolutionary relationship with the cultivated varieties. When present in the field red rice takes resources away from cultivated plants, including space, nutrients and light, causing lower yields. Red rice will generally invade marginal areas of the field, such as irrigation ditches and depressed areas of the field were water levels are too high for cultivated types to thrive, but have the potential to infest entire fields causing up to 80% yield loss (Gealy et al., 2000; Vaughan et al., 2001a; Vaughan et al., 2001b). Red rice seeds will generally shatter, or fall into the field prior to harvesting but if red rice plants maintain their seeds and are harvested the incorporation of red rice seeds significantly lowers the quality of the crop, often making it difficult to recoup financial investments. Management of red rice is extremely difficult due to the close evolutionary relationships between cultivated rice and red rice, making it virtually impossible to kill red rice without also killing cultivated rice.

Origin of red rice

Despite scientific interest and its economic importance, red rice's relationship to cultivated *Oryza sativa* remains a mystery. Two hypotheses regarding its evolutionary history exist in the literature. The first suggests red rice is a naturalized or weedy de-domesticate of cultivated rice, *O. sativa,* while the second hypothesis points toward a hybrid origin resulting from inter-specific crosses between the wild *O. rufipogon* and *O. sativa*.

Data from global populations provides mixed results regarding the evolutionary history of red rice. Allozyme studies of Asian populations suggested the presence of three classes of red rice in the region: *indica*-like crop mimics that originated from local *indica* cultivars; and two hybrid types resulting from hybridization between *O. rufipogon* and regional cultivars (Tang and Morishima, 1996). Other large scale screens of global red rice populations from Bangladesh, Brazil, Bhutan, China, India, Japan and

Korea using microsatellite (SSR) markers have found variable results depending on the region investigated. Populations of red rice in Bangladesh had many wild *O. rufipogon alleles*, suggesting a hybrid (*O. sativa* X *O. rufipogon*) origin of this population (Cho et al., 1995; Suh et al., 1997). Populations from Brazil, Bhutan, India, southern China and Japan were shown to be *indica*-like, and most likely the result of de-domestication from tropical and temperate *O. sativa indica* cultivars (Suh et al., 1997). Populations from Bhutan northern China and Korea showed genetic similarity to *O. sativa japonica* cultivars, suggesting a de-domesticated origin (Cao et al., 2006),while populations from other areas of Korea and China were thought to have a hybrid origin resulting from *O. sativa japonica* X *O. rufipogon* genetic exchange (Zhange et al in prep). Other studies investigating red rice in Bhutan suggest that hybridization between *O. sativa japonica* and *O. sativa indica* cultivars gave rise to the weedy populations seen (Nishikawa et al., 2005). Many studies find distinct genetic clusters of red rice, yet few of these studies agree in their regional/type assignments (Gealy et al., 2003; Suh et al., 1997).

Red rice in the United States

In the United States rice is grown in two major areas; the southern Mississippi Valley, extending from southeast Missouri across Arkansas, Mississippi, Louisiana to the Gulf coastal plain in Texas, and in north-central California. Red rice in the southern United States is very diverse, in the field all morphological biotypes from crop mimics to wild phenotypes can be found, yet the amount of phenotypic diversity located in any single field, varies greatly. Some fields contain plants with every imaginable combination of seed characteristics, while other fields look like clonal stands of a single red rice type (personal observation).

A previous study on red rice found in the United States by Vaughan and colleagues (2001) screened 18 individuals from Texas representing distinct biotypes using microsatellites (SSRs). They were able to group the straw hull and brown hull biotypes together, classifying them as more *indica* like; while black hull biotypes were characterized as more genetically similar to *O. rufipogon* accessions.

Contrary to these results Gealy and colleagues (2002) screened accessions from Arkansas finding that the assayed individuals formed distinct genetic clusters based on morphological characters, the presence or absence of awns. All accessions in this study, regardless of awn character, were more genetically similar to tropical *japonica* cultivars.

More recent studies have a historic panel of 27 historic red rice accessions and both microsatellites and SNP markers to describe low diversity levels in these populations. These findings lead researchers to postulate about evolutionary history of red rice in the United States, suggesting that red rice in the US is the result of de-domestication (Gross et al., 2009; Gross et al., 2010; Reagon et al., 2010). A closer look into the origin of red rice using current population level sampling across the southern US reveals a more complex evolutionary origin (Kuntz Chapter 2). High levels of genetic diversity and evidence for hybrid evolutionary history in multiple populations suggests that at least some portion of weedy red rice found in the US are introduced weeds originally from Asia. The lack of concordance between the varieties involved gives little clarity to the problem. This situation is not surprising given the potential for continuing gene flow in the field via both the commercial dispersal of seeds and ongoing *in situ* pollen flow between regional red rice populations and their cultivated congeners.

Why do we care about evolutionary and geographic origins of red rice?

Understanding the level of diversity and structure in red rice populations will contribute to our knowledge of the weed's evolutionary history, and its ability to persist in and adapt to local field conditions. Clarity regarding how red rice partitions its variation within and among populations will give important information regarding the evolutionary dynamics of the weed and insights into its introduction history. The potential for continuing evolution in these populations cannot be understood without knowledge of the scale of gene flow between populations. Inter-specific gene flow and introgression between red rice and the cultivated rice may contribute to the weeds tenacity and persistence in cultivated areas.

Many agricultural species are grown in areas outside their progenitor's native range. This is the case with our species of interest, cultivated rice *Oryza sativa*. Knowledge of historic routes of invasion is important in order to understand how environmental and evolutionary factors can affect the successful establishment and persistence of introduced weeds (Estoup and Guillemaud, 2010; Lachmuth et al., 2010; Pairon et al., 2010; Sakai et al., 2001). Locating the geographic origin of source populations and the routes of invasion of red rice has taken may allow researchers to design strategies to prevent continual introductions. A deeper understanding of the underlying factors contributing to weed generation, dissemination, and persistence help us to address important questions in evolutionary biology; including: the mechanisms involved in adaption and the environmental and evolutionary factors that influence the success of biological invasions and the successful generation of weeds (Ellstrand et al., 2010; Parker et al., 2003; Stewart et al., 2009; Wilen et al., 1995). Many factors contribute to the ability of these organisms to persist and eventually thrive in the agricultural environment, including but not limited to, the number of introductions, genetic diversity in founding populations, hybridization with native flora, and generalized life history traits such as, mating system requirements (Bacigalupe, 2009; Dlugosch and Hays, 2008; Ellstrand et al., 1999b).

Why Gene Flow is Important?

Assessing levels of crop to weed gene flow is fundamental to our understanding of red rice diversity, persistence, and general field dynamics. Within this agronomic environment we are particularly interested in red rice outcrossing events which involve their cultivated congeners. Gene flow between cultivated and red rice may lead to the introgression of alleles responsible for traits such as herbicide resistance, dwarf stature and various seed morphology traits, all of which have the potential to contribute to the ability of these populations' survival in the face of strong selection against them. Knowing the proportion of outcrossing events that are inter-specific is critical, if we are to estimate the potential for gene flow, particularly of transgenes or other selectively beneficial traits into red rice populations from cultivated rice. Such introgression events may lead to drastic changes in the genetic diversity of other

local wild rice species, or may prompt the creation of red rice 'super-weeds' (Ellstrand et al., 1999a; Ellstrand, 2003; Ellstrand and Schierenbeck, 2000; Lu and Snow, 2005; Snow et al., 2003). In either case understanding gene flow dynamics in the field is critical for understanding the potential trajectories of the system.

The chapters that follow are a series of manuscripts describing our findings on the evolutionary origins and population genetics of red rice in the southern United States. In Chapter 2 we measure genetic diversity and population structure in red rice populations, finding evidence for a hybrid origin of red rice, with the possibility of multiple introductions into the southern United States. Chapter 3 investigates potential source populations of red rice in China to better understand the introduction history of US populations. In the final chapter we directly measure rates of gene flow with paternity analysis to gain knowledge of intra- and inter-specific gene flow dynamics in these populations.

REFERENCES

Adair C.R., Miller M., Beachell H. (1962) Rice improvement and culture in the United States. Advances in Agronomy 14:61-108.

Bacigalupe L. (2009) Biological invasions and phenotypic evolution: a quantitative genetic perspective. Biological Invasions:2243-2250.

Baki B., Chin D., Mortimer M. (2000) Wild and weedy rice in rice ecosystems in Asia-a review.

Barrett S.H. (1983) Crop mimicry in weeds. Economic Botany 37:255-282.

Bartsch D., Lehnen M., Clegg J., Pohl-Orf M., Schuphan I., Ellstrand N.C. (1999) Impact of gene flow from cultivated beet on genetic diversity of wild sea beet populations. Molecular Ecology 8:1733-1741.

Bossuyt B., Hermy M. (2001) Influence of land use history on seed banks in European temperate forest ecosystems: a review. Ecography 24:225-238.

Burger J., Holt J., Ellstrand N. (2007) Rapid phenotypic divergence of feral rye from domesticated cereal rye. Weed Science:204-211.

Burger J.C., Lee S., Ellstrand N.C. (2006) Origin and genetic structure of feral rye in the western United States. Molecular Ecology 15:2527-2539.

Burger J.C., Chapman M.A., Burke J.M. (2008) Molecular insights into the evolution of crop plants. American Journal of Botany 95:113-122.

Campbell L.G., Snow A.A., Ridley C.E. (2006) Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. Ecology Letters 9:1198-1209.

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

Chen L.J., Lee D.S., Song Z.P., Suh H.S., LU B.A.O.R. (2004) Gene flow from cultivated rice (Oryza sativa) to its weedy and wild relatives. Annals of Botany 93:67.

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67.

Cho Y.C., Chung T.Y., Suh H.S. (1995) Genetic characteristics of Korean weedy rice (Oryza sativa L.) by RFLP analysis. Euphytica 86:103-110.

De Wet J.M.J., Harlan J.R. (1975) Weeds and domesticates: evolution in the man-made habitat. Economic Botany 29:99-108.

Dlugosch K., Hays C. (2008) Genotypes on the move: some things old and some things new shape the genetics of colonization during species invasions. Molecular Ecology:4583-4585.

Doebley J. (2004) The genetics of maize evolution. Annual Review of Genetics 38:37-59.

Doebley J.F., Gaut B.S., Smith B.D. (2006) The molecular genetics of crop domestication. Cell 127:1309-1321.

Doi K., Nonomura M.N., Yoshimura A., Iwata N., Vaughan D.A. (2000) RFLP relationships of A-genome species in the genus Oryza. Journal of the Faculty of Agriculture Kyushu University 45:83-98.

Ellstrand N., Prentice H., Hancock J. (1999a) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics:539-563.

Ellstrand N., Heredia S., Leak-Garcia J., Heraty J., Burger J., Yao L., Nohzadeh-Malakshah S., Ridley C. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evolutionary Applications:494-504.

Ellstrand N.C. (2003) Current knowledge of gene flow in plants: implications for transgene flow. pp. 1163-1170.

Ellstrand N.C., Schierenbeck K.A. (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? pp. 7043-7050.

Ellstrand N.C., Prentice H.C., Hancock J.F. (1999b) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Estoup A., Guillemaud T. (2010) Reconstructing routes of invasion using genetic data: why, how and so what? Molecular Ecology.

Gealy D.R., Saldain N.E., Talbert R.E. (2000) Emergence of red rice (*Oryza sativa*) ecotypes under dry-seeded rice (*Oryza sativa*) culture. Weed Technology 14:406-412.

Gealy D.R., Tai T.H., Sneller C.H. (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Science 50:333-339.

Gealy D.R., Mitten D.H., Rutger J.N. (2003) Gene flow between red rice (*Oryza sativa*) and herbicide-resistant rice (*O. sativa*): Implications for weed management. Weed Technology 17:627-645.

Gross B.L., Skare K.J., Olsen K.M. (2009) Novel Phr1 mutations and the evolution of phenol reaction variation in US weedy rice (Oryza sativa), New Phytol, England. pp. 842-50.

Gross B.L., Reagon M., Hsu S.C., Caicedo A.L., Jia Y., Olsen K.M. (2010) Seeing red: the origin of grain pigmentation in US weedy rice, Mol Ecol, England. pp. 3380-93.

Hammer K. (1984) The domestication syndrome. Kulturpflanze 32:11-34.

Harlan J.R. (1992) Crops and Man, American Society of Agronomy, Madison, WI.

Harlan J.R., de Wet J.M.J. (1965) Some thoughts about weeds. Economic Botany 19:16-24.

Hopfensperger K.N. (2007) A review of similarity between seed bank and standing vegetation across ecosystems. Oikos 116:1438-1448.

Jaenicke-Despres V., Buckler E.S., Smith B.D., Gilbert M.T.P., Cooper A., Doebley J., Paabo S. (2003) Early allelic selection in maize as revealed by ancient DNA. Science 302:1206-1208.

Khush G.S. (1997) Origin, dispersal, cultivation and variation of rice. Plant Molecular Biology 35:25-34.

Koinange E.M.K., Singh S.P., Gepts P. (1996) Genetic control of the domestication syndrome in common bean. Crop Science 36:1037-1044.

Lachmuth S., Durka W., Schurr F.M. (2010) The making of a rapid plant invader: genetic diversity and differentiation in the native and invaded range of *Senecio inaequidens*. Mol Ecol 19:3952-67.

Langevin S.A., Clay K., Grace J.B. (1990) The incidence and effects of hybridization between cultivated rice and its related weed red rice (Oryza sativa L.). Evolution 44:1000-1008.

Lawton-Rauh A., Burgos N. (2010) Cultivated and weedy rice interactions and the domestication process, Mol Ecol, England. pp. 3243-5.

Li C.B., Zhou A.L., Sang T. (2006) Genetic analysis of rice domestication syndrome with the wild annual species, Oryza nivara. New Phytologist 170:185-193.

Li X.R., Tan L.B., Zhu Z.F., Huang H.Y., Liu Y., Hu S.N., Sun C.Q. (2009) Patterns of nucleotide diversity in wild and cultivated rice. Plant Systematics and Evolution 281:97-106.

Li Z.K., Luo L., Mei H., Wang D., Shu Q., Tabien R., Zhong D., Ying C., Stansel J., Khush G. (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158:1737.

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

Londo J.P., Chiang Y.C., Hung K.H., Chiang T.Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proceedings of the National Academy of Sciences of the United States of America 103:9578-9583.

Lu B.R., Snow A.A. (2005) Gene flow from genetically modified rice and its environmental consequences. Bioscience 55:669-678.

Luo L., Li Z.K., Mei H., Shu Q., Tabien R., Zhong D., Ying C., Stansel J., Khush G., Paterson A. (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 158:1755.

Nishikawa T., Vaughan D.A., Kadowaki K. (2005) Phylogenetic analysis of *Oryza species*, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. Theoretical and Applied Genetics 110:696-705.

Pairon M., Petitpierre B., Campbell M., Guisan A., Broennimann O., Baret P., Jacquemart A., Besnard G. (2010) Multiple introductions boosted genetic diversity in the invasive range of black cherry (Prunus serotina; Rosaceae). Annals of Botany:881-890.

Pake C.E., Venable D.L. (1996) Seed banks in desert annuals: implications for persistence and coexistence in variable environments. Ecology 77:1427-1435.

Parker I., Rodriguez J., Loik M. (2003) An evolutionary approach to understanding the biology of invasions: Local adaptation and general-purpose genotypes in the weed *Verbascum thapsus*. Conservation Biology:59-72.

Prathepha P. (2009) Molecular analysis of utility of a retrotransposon, p-SINE1-r2 in the Asian wild rice and weedy rice populations. Pak J Biol Sci 12:270-5.

Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N.T., Terauchi R. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. TAG Theoretical and Applied Genetics 114:731-743.

Reagon M., Thurber C.S., Gross B.L., Olsen K.M., Jia Y., Caicedo A.L. (2010) Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice, BMC Evol Biol, England. pp. 180.

Ross-Ibarra J., Morrell P.L., Gaut B.S. (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences 104:8641.

Sakai A.K., Allendorf F.W., Holt J.S., Lodge D.M., Molofsky J., With K.A., Baughman S., Cabin R.J., Cohen J.E., Ellstrand N.C., McCauley D.E., O'Neil P., Parker I.M., Thompson J.N., Weller S.G. (2001) The population biology of invasive species. Annual Review of Ecology and Systematics 32:305-332.

Schierenbeck K., Ellstrand N. (2009) Hybridization and the evolution of invasiveness in plants and other organisms. Biological Invasions:1093-1105.

Snow A.A., Campbell L.G. (2005) Can feral radishes become weeds. Crop ferality and volunteerism:193–208.

Snow A.A., Uthus K.L., Culley T.M. (2001) Fitness of hybrids between weedy and cultivated radish: Implications for weed evolution. Ecological Applications 11:934-943.

Snow A.A., Pilson D., Rieseberg L.H., Paulsen M.J., Pleskac N., Reagon M.R., Wolf D.E., Selbo S.M. (2003) A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. Ecological Applications 13:279-286.

Song Z., Lu B.R., Chen J. (2004) Pollen flow of cultivated rice measured under experimental conditions. Biodiversity and Conservation 13:579-590.

Song Z.P., Lu B.R., Zhu Y.G., Chen J.K. (2003) Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657-665.

Stewart C.N., Tranel P.J., Horvath D.P., Anderson J.V., Rieseberg L.H., Westwood J.H., Mallory-Smith C.A., Zapiola M.L., Dlugosch K.M. (2009) Evolution of Weediness and Invasiveness: Charting the Course for Weed Genomics. Weed Science 57:451-462.

Suh H., Sato Y., Morishima H. (1997) Genetic characterization of weedy rice (*Oryza sativa* L.) based on morpho-physiology, isozymes and RAPD markers. TAG Theoretical and Applied Genetics 94:316-321.

Tang L., Morishima H. (1996) Genetic characteristics and origin of weedy rice. Paper on Origin and Dissemination of Cultivated Rice in China:211-218.

Tian F., Stevens N.M., Buckler E.S. (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proceedings of the National Academy of Sciences of the United States of America 106:9979-9986.

Vaughan D., Balazs E., Heslop-Harrison J. (2007) From crop domestication to super-domestication. Annals of Botany 100:893.

Vaughan D.A., Morishima H., Kadowaki K. (2003) Diversity in the *Oryza* genus. Current Opinion in Plant Biology 6:139-146.

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

Vaughan D.A., Kadowaki K., Kaga A., Tomooka N. (2005) On the phylogeny and biogeography of the genus *Oryza*. Breeding Science 55:113-122.

Vaughan K.L., Ottis B.V., Prazak-Havey A.M., Bormans C.A., Sneller C., Chandler J.M., Park W.D. (2001a) Is all red rice found in commercial rice really Oryza sativa? Weed Science 49:468-476.

Vaughan L.K., Ottis B.V., Prazak-Havey A.M., Bormans C.A., Sneller C., Chandler J.M., Park W.D. (2001b) Is all red rice found in commercial rice really *Oryza sativa*? Weed Science 49:468-476.

Venable D.L. (2007) Bet hedging in a guild of desert annuals. Ecology 88:1086-1090.

Vitte C., Ishii T., Lamy F., Brar D., Panaud O. (2004) Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). Molecular Genetics and Genomics 272:504-511.

Warwick S.I., Beckie H.J., Hall L.M. (2009) Gene flow, invasiveness, and ecological impact of genetically modified crops, Ann N Y Acad Sci, United States. pp. 72-99.

Wilen C., Holt J., Ellstrand N., Shaw R. (1995) Genotypic diversity of Kikuyugrass (*Pennisetum clandestinum*)populations in California. Weed Science:209-214.

Xiao J., Li J., Grandillo S., Ahn S.N., Yuan L., Tanksley S.D., McCouch S.R. (1998) Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. Genetics 150:899.

Yamanaka S., Nakamura I., Nakai H., Sato Y.I. (2003) Dual origin of the cultivated rice based on molecular markers of newly collected annual and perennial strains of wild rice species, *Oryza nivara* and *O. rufipogon*. Genetic Resources and Crop Evolution 50:529-538.

CHAPTER II:

MULTIPLE HYBRID ORIGINS OF RED RICE IN SOUTHERN UNITED STATES: EVIDENCE FROM NATURAL POPULATIONS[1]

[1] Kuntz, E. J., Mauricio, R. To be submitted to Molecular Ecology.

ABSTRACT

Domestication has generally been used as a powerful exemplar of the power of selection to create novel phenotypes over time periods. Reversion of cultivated species to an ancestral state is a less well-understood phenomenon and has been suggested as being the process responsible for the evolution of red rice in the southern United States. Field-collected red rice populations are highly morphologically and genetically diverse, forming a distinct taxonomic group. Contrary to findings suggesting red rice seed morphology is indicative of evolutionary history, we find no association with visible phenotypic and genetic assignment. A Bayesian clustering analysis as well as a private alleles-based analysis reveals evidence for mixed ancestry, with four distinct genetic groups co-occurring across the region. Our results provide strong support for multiple hybrid origins of red rice in Asia with subsequent dispersal to the United States and no support for a "de-domestication" event.

INTRODUCTION

Understanding plant domestication into modern crops has been a major focus of evolutionary genetics for the past century (Beadle 1980). Processes by which wild, and often weed, plants are transformed via artificial selection into cultivated species can have profound effects on the phenotype (Buckler *et al.* 2006; Doebley *et al.* 2006). Most crops possess similar combinations of traits that make them amenable to cultivation. These traits are collectively termed the domestication syndrome (Harlan 1992). Often the wild progenitors of cultivated varieties were populations of asynchronously-flowering plants that possessed a mixed mating system and a variety of seed dispersal mechanisms (Eubanks 1997; Li *et al.* 2006). Cultivated plants generally possess a more erect form, an annual habit, synchronous flowering and lack seed dispersal mechanisms. Understanding the interplay between selection for the maintenance of cultivated phenotypes and the production of weedy types can help us understand the underlying genetic architecture of domestication traits, and the strength of selection needed to maintain crop phenotypes (Doebley *et al.* 2006; Ellstrand *et al.* 1999; Snow *et al.* 2001).

Often the most noxious weeds of cultivated crops are closely related congeners. For cultivated rice, *Oryza sativa*, red rice is its most economically and ecologically problematic weed world-wide, and is found in virtually all areas of rice production (Gealy *et al.* 2003). Red rice aggressively outcompetes crops in the field by crowding, shading, and lodging cultivated individuals (Gealy& Bryant 2009; Oard *et al.* 2000). Red rice does not retain its seeds like the cultivated species and instead releases seed into the soil where they can persist in the seed bank for decades. Infestations of red rice occur in greater than 30% of cultivated rice fields in the southern US, leading to annual losses of more than 50 million US dollars (Gealy *et al.* 2003; Vaughan *et al.* 2001). In the US rice is primarily grown in the Mississippi flood plain including Arkansas, Louisiana, Mississippi, southern Missouri, and eastern Texas, where tropical *O.sativa japonica* cultivars are used. There is a small secondary area of cultivation in the Sacramento Valley in California where temperate *O. sativa japonica* and a small number of *O. sativa indica* cultivars are grown.

Two phenotypes of weedy red rice have been described; the crop mimic and the wild form (Gealy *et al.* 2002; Schwanke *et al.* 2008b). The crop mimic persists in agricultural fields undetected, due to its similarity to the cultivar; the weed possesses short stature, light straw colored hull pigmentation, and lacks awns. Red rice of this type causes economic damage utilizing field resources (fertilizer, space, etc) only to shatter upon seed maturity, or by contaminating seed stock, lowering seed grade when harvested. The wild form is easy to identify in the field, but is difficult to eradicate. Red rice aggressively out competes cultivated rice, especially in marginal areas like irrigation ditches, overflows, and depressed (or unusually deep) areas of the field.

Two major hypotheses have been proposed to explain the origin of both weed phenotypes of red rice in the southern US. The first hypothesis is that red rice is the result of 'de-domestication' by which local US cultivars occasionally produce offspring with wild characteristics, giving rise to weedy red rice populations *in situ* (Olsen *et al.* 2007; Vaughan *et al.* 2001). In an introduced range devoid of wild/weedy relatives (Ellstrand 2009; Ellstrand *et al.* 1999; Lu& Snow 2005), these feral crop volunteers establish in fields, persist in the seed bank and give rise to persistent weedy populations. The second hypothesis proposes an *ex situ* hybrid origin of red rice in which hybridization between the cultivated *O. sativa* and the wild *O. rufipogon* in Asia gave rise to red rice, with subsequent transfer of the weedy hybrids to the US as red rice seed contaminants in seed stock.

These hypotheses lead to different expectations about the population genetic structure of red rice and its genetic similarity to cultivated and ancestral rice lineages. If weedy red rice in the southern US is the result of *in situ* de-domestication, then populations of the feral de-domesticated red rice should be similar to and contain a subset of, the genetic variation present in the local parental cultivar, *O. sativa japonica* (but not variation restricted to the ancestral range). Conversely, a crop x wild hybrid origin in Asia, would lead to very different expectations. Levels of genetic diversity in hybrid red rice should be higher, and associations with parental cultivars may be obscured due to the shared evolutionary history between the cultivated and *O. rufipogon* parents.

This study is designed to understand the evolutionary history of red rice populations in the Southern United States. Population level sampling of in situ red rice, local cultivars, diverse *O. sativa* cultivars and *O. rufipogon* range-wide accessions were used to characterize genetic diversity and genetic structure in order to elucidate the origin of the weeds in the US.

MATERIALS AND METHODS

Study System

Red rice has been classified as *O. sativa* because the plants are sexually compatible (Vaughan *et al.* 2001). However, red rice exhibits higher levels of out-crossing (0.1-52%) and an extended flowering time, compared to the almost exclusively selfing cultivated varieties (Estorninos *et al.* 2005; Gealy *et al.* 2003; Langevin *et al.* 1990). In addition, red rice shows morphological variation within and among populations for date of first flower, seed dormancy, seed morphology (color, pubescence, grain size and awn presence), cold tolerance, herbicide resistance, and tillering capacity (Gealy et al. 2002). These differences between red and cultivated rice have prompted some to consider red rice a distinct taxonomic unit (Schwanke *et al.* 2008a; Vaughan *et al.* 2001). Red rice in the US has been further classified into ecotypes based on the presence or absence of an awn and the color of the hull, either straw or black (Gealy *et al.* 2000; Schwanke *et al.* 2008b). The straw-hull coloration and absence of the awn are shared with *O. sativa*, while the black-hull coloration and presence of the are shared with *O. rufipogon* (Gealy *et al.* 2002; Vaughan *et al.* 2001).

Sampling

Populations of red rice were collected from infested rice fields throughout the southern US in July 2006. A total of 294 individual red rice samples were collected from 19 field sites in three regions (Texas, southern Louisiana, and Arkansas) characterized by distinct climatic differences, river drainages,

cultivation practices and regionally specific *O. sativa* cultivars (Supplement 1). Field samples were collected a few weeks prior to rice harvest to ensure seed maturity. Individual panicles were collected at one meter intervals to prevent collection of clonal ramets. One seed per panicle was hulled and prepped for DNA extraction, ensuring that individual red rice plants were only sampled once. In addition to red rice individuals, cultivated rice growing in each field was collected and analyzed.

Twenty-seven additional red rice accessions were obtained from the Dale Bumpers Rice Research Center in Stuttgart, Arkansas. Each accession consisted of a single plant collected between 1994 and 1999 from across the southern growing region, representing Missouri, Arkansas, Mississippi, Texas and Louisiana. This collection is propagated via single seed descent and is maintained as a red rice core reference panel. For additional details regarding collection and propagation see Gealy *et al* (2002). An additional 44 *O. sativa* cultivars and 37 *O. rufipogon* accessions were included in the study (Supplement 1). The *O. rufipogon* samples were propagated by the International Rice Research Institute (IRRI) and acquired as extracted DNA from the B. Schaal lab at Washington University, St. Louis USA.

DNA extractions

Twenty to thirty red rice seeds and cultivar seeds from each sampled population were scored for hull color, awn presence and seed color. All seeds were germinated on Watman filter paper in Petri dishes and grown in a growth chamber at 30°C. Seedlings were collected after 7-14 days, were snap frozen in liquid nitrogen and total genomic DNA was extracted using Qiagen's DNAeasy mini kit (Valencia, CA). Extracted samples were quantified using a spectrophotometer (NanoDrop/ Wilmington, DE) and standardized to approximately 30 ng DNA per μl.

Microsatellite genotyping

Fifteen microsatellite loci (simple sequence repeats, or SSRs) representing unique linkage groups were chosen from previously identified microsatellite regions (Temnykh *et al.* 2001). Polymerase chain

reaction (PCR) amplifications were performed using three primers: a SSR sequence specific forward primer with an M13 tail at its 5' end (5'- CACGACGTTGTAAAACGACA-3'), a SSR sequence specific reverse primer and a fluorescent-labeled M13 primer (Schuelke 2000). Thirteen microliter reactions contained 1.5µl of DNA sample, 0.093µl 10 mM forward SSR primer, 0.466µl 10 mM reverse SSR primer, 0.446µl 10 mM M13 primer, 1.2µl 10 mM dNTPs, 1.2µl 10 mM MgCl2, 1.2µl 10XPCR buffer, and one unit Taq polymerase. Samples were amplified using the following PCR profile: initial denaturation at 94°C for 3min; followed by 35 cycles of: 94°C for 45 sec, 55°C annealing for 45sec, 72°C extension for 45sec; and a final extension at 72°C for 15 min. Post PCR, samples were held at 4°C. The resulting PCR fragments were diluted 1:10 and genotyped on a capillary electrophoresis sequencer ABI3700 (Applied Biosystems; Foster City, Ca) using MapMarker1000 ROX (BioVentures Inc; Murfreesboro, TN) as an internal size standard.

Genotypic analysis

All microsatellite loci were scored using GeneMarker, version 1.70 (SoftGenetics; State College, PA). Samples were subjected to randomization and 96 individuals were re-sampled to confirm genotype fidelity. All produced consistent genotypes at the 15 scored loci.

Descriptive population genetic statistics were calculated using GENALEX (Peakall& Smouse 2006); including: mean number of alleles per locus, percentage of polymorphic loci, gene diversity [as measured by Nei's expected heterozygosity (H_e)](Nei 1978), and polymorphism information content (PIC). To represent graphically the relationships between taxa, genetic distance values calculated in GENALEX were used to perform a Principal Component Analysis (PCA), where a dissimilarity matrix of pairwise differences between individuals was constructed and eigenvalue analysis was used to compress the total variation between samples into a limited number of dimensions.

Spatial patterns of genetic variation and geographic associations were analyzed using an analysis of molecular variance (AMOVA), as implemented by GENALEX. The AMOVA analysis was used to

calculate F_{ST} among regions, between populations within regions and within populations. Populations of red rice were subdivided into regional groupings based on natural geographic boundaries and the cultivation of distinct rice cultivars in each area. In this Southern red rice collection three regions were investigated: Texas, Arkansas and Louisiana.

Neighbor-joining dendrograms and UPGMA phenograms were then constructed using the NEIGHBOR program in PHYLIP 3.67 (Felsenstein 2006; Felsenstein 2007). We used Nei's genetic distance between populations and sample input order was randomized (Nei 1978). Support for the final tree was calculated with 1000 bootstrap replicates using the CONSENSE program.

The program InStruct (Falush *et al.* 2003; Gao *et al.* 2007; Pritchard *et al.* 2002) was used to infer population structure and population selfing rates in the complete data set, including *O. rufipogon* accessions, *O. sativa indica* and *japonica* cultivars and the southern weedy red rice samples. InStruct limits the number of false signals of population substructure generated in the STRUCTURE program given the high selfing rate across *Oryza*. InStruct uses a Markov Chain Monte Carlo algorithm in a Bayesian framework to simultaneously estimate selfing rates and examine the number of populations and the clustering of individuals in those populations based on linkage disequilibrium between the markers. This information is used to assign individuals into a hypothetical number of populations (K).

Separate InStruct analyses were preformed on the possible parental species (*O. rufipogon* and *O.sativa*), the 19 red rice populations from the southern US, and on the total data set (i.e. including *O. rufipogon*, *O. sativa* and red rice) using an initial burn in of 100,000 and with two independent MCMC chains run for a length of 200,000 iterations. We tested a range (1-40) of K values (number of population clusters) using 10 independent runs for each K in each analysis. The most significant K value was assigned using a log likelihood ratio test. All runs used the admixture model with the population of origin information setting disabled. The data were displayed using the Distruct program (Rosenberg 2004). The

most likely K for the putative parental data set was used to test the complete dataset, as red rice is expected to be derived from at least one of these species.

Private allele analysis was then used to investigate fine scale associations between taxa. To test the hypothesis of a hybrid origin for weedy red rice, alleles specific to *O. rufipogon* or to *O. sativa* were identified and their presence and distribution of these alleles in the weedy red rice populations was assessed. To address which *O. sativa* subspecies were involved in the origin of weedy red rice, alleles specific to either the *indica* or *japonica* cultivar group were identified and the presence of these alleles in red rice was assessed. Differences between local US cultivars and the world sample of *O. sativa* were also investigated, and weedy red rice individuals were screened to see if alleles specific to either local US or global rice cultivars were found more frequently in the southern US red rice populations.

RESULTS

Field collected red rice populations contained a wide range of seed phenotypes. The two biotypes described in the literature, straw-hull/awn-less and black-hull/awned, were only a subset of the actual range of phenotypic diversity observed in the field. Most notably straw-hull/awned, and black/awn-less types were also found in populations across the range (Table1). Significant variation was seen in the amount of pubescence on the hull, hull color (straw, brown and black). When present awn length ranged from 0.5 to 8.0 cm long, and although all seeds were red there was variation in the intensity of pericarp pigmentation, ranging from light pink to dark burgundy red.

Microsatellite analysis

All 15 microsatellites were polymorphic in *O. rufipogon* and weedy red rice, 14 loci were polymorphic in *O. sativa japonica* and 13 in *O. sativa indica* (Table 2). Although red rice had the largest total number of alleles and the most alleles per polymorphic locus, the effective number of alleles (A_e)

was highest in *O. rufipogon* (A_e = 8.40), followed by red rice (A_e = 3.02). As predicted, *O. rufipogon* contained the highest level of genetic diversity (H_e = 0.855). Genetic diversity in red rice (0.645) was intermediate to the wild and cultivated groups (H_e = 0.401 for var. *japonica* and 0.470 for var. *indica*). Observed heterozygosity in both *O. sativa japonica* and *indica* were virtually zero, as we would expect for highly selfing cultivars. *O. rufipogon* had the highest level of observed heterozygosity (H_o= 0.216). In all cases observed heterozygosity was much lower than expected heterozygosity suggesting high levels of selfing in all taxa.

Red rice and *O. rufipogon* have the lowest pairwise F_{ST} value (0.068) demonstrating that they have the highest genetic similarity between all rice groups. The cultivars, *O. sativa indica* and *japonica*, have the highest pairwise F_{ST} values (0.384), indicating that the two cultivar types are more genetically differentiated from one another than they are from the other rice groups (Table 3). Comparisons between *O. rufipogon* and the cultivated *O. sativa* sub-species show that *O. rufipogon* is more similar to *O. sativa indica* (0.157) than *O. sativa japonica* (0.187).

Genetic diversity among red rice populations

At the population level we see a range of genetic diversity (0.511 to 0.107) in red rice in the southern US. When the populations are pooled across any one region each region is polymorphic at all loci, but individual populations within a particular region show differences in diversity measures (Table 4). Red rice populations in Arkansas averaged 10.27 alleles per polymorphic locus, followed by Louisiana with 9.06 and Texas with 7.80. Taken as a whole, populations in Arkansas had more genetic diversity than either the populations of Louisiana or Texas (Table 4). Analysis of molecular variance (AMOVA) shows that most variation (58%) in red rice is present at the population level (Table 5). Variation among populations within their respective regions explained 36% of the variation, and 6%, is partitioned among geographic regions.

Principal coordinate analysis (PCA) based on the 15 microsatellite loci was carried out for the entire data set; including all red rice, *O. sativa indica*, *O. sativa japonica*, and *O. rufipogon* individuals (Figure 1). The PCA analysis explained 68% of the variation in the rice taxa within the first three axes. The first two axes explain 52% of the variation, while the third explained an additional 16%. Given the evolutionary history of *O. rufipogon* as the progenitor of both cultivated sub-species of rice it is not unexpected that *O. rufipogon* individuals are found in the middle of the PCA, overlapping all other rice groups. Red rice comprises one, albeit diffuse cluster, with two sub-groups of increased individual membership occurring at opposite ends of the first axis. Half of the *O. sativa indica* cultivars are found residing within the first red rice sub-group, and thus are termed "indica-like" red rice. Interestingly, *O. sativa indica* cultivars were more closely associated with *O. rufipogon* accessions than their *O. sativa japonica* cultivar counterparts, a result consistent with the pairwise F_{ST} results. The second sub-group of red rice individuals, clustering on the opposing end of the first axis also shares some overlap with *O. rufipogon* individuals but on the whole is a distinct group of individuals showing no associations with either cultivar. There is a small subset of red rice individuals that cluster at the interface of the *O. sativa japonica* and *O. rufipogon* group overlap (*japonica*-like).

Overlaying phenotypic and genetic marker information on the individual PCA shows no association with hull color, or awn presence and red rice sub-group. Red rice individuals show moderate clustering based on population, although, there are many cases in which a particular individual is more tightly clustered with members from a different population, or to individuals in a population in a different region. There is no clustering based on region. Individuals from all three regions, Texas, Louisiana and Arkansas, are found in both red rice sub-groupings and are also associated with the *O. sativa indica* group.

Both neighbor-joining and UPGMA methods produced identical relationships between populations and as such, only the neighbor-joining dendrograms are presented (Figures 3 and 4). Populations of red rice fall into two clades (with moderate bootstrap support) much like the split in sub

populations seen in the PCA. Populations from the first group were made up of five Arkansas populations (AR1, Ar9, Ar8, Ar3 and Ar6), three Louisiana populations (La11, La12, and La13) and two Texas populations (Tx10 and Tx11) (Figure 3). The second group included populations from all three regions including three populations from Louisiana (La5, La6 and La7), four populations from Texas (Tx13, Tx3, Tx8 and Tx9) and two populations from Arkansas (Ar5 and Ar2). The fact that the placement of individual populations was not correlated with their region of origin is consistent with the AMOVA results.

When the complete data set was analyzed *O. sativa indica* was positioned closest to five Arkansas populations and *O. rufipogon* was positioned sister to *O. sativa japonica*. Interestingly, *O. sativa japonica* was positioned sister to the Tx13 population with moderate bootstrap support (58%). *Oryza rufipogon's* central position and low bootstrap support is not surprising given our knowledge of its evolutionary history. As the wild progenitor of both *O. sativa* cultivars, *O. rufipogon* should be positioned sister to both groups with little support for differentiation between the groups.

InStruct analysis:

Bayesian cluster analysis on the putative parental populations, *O. rufipogon, O. sativa indica* and *O. sativa japonica* reflects the known domestication history of cultivated rice. The analysis produced an optimal K value of eight (Figure 5). The two derived groups are largely assigned to one genetic cluster each; seen as a virtually 100% cluster for the *O. sativa indica* individuals (in pink) and almost 100% cluster in the *O. sativa japonica* group (in purple). Two *O. sativa japonica* cultivars, Nipponbare and Rexoro, are distinct from the rest of the *japonica* group (Figure 5a). These two cultivars share a more admixed genetic history relative to the rest of the *japonica* cultivars, making them appear to be of mixed ancestry, as seen in the *O. rufipogon* individuals. There are two *O. rufipogon* individuals, one from Cambodia and the other from Laos, that are predominantly assigned with *O. sativa indica* (pink); suggesting that they represent populations from which the domesticated *indica* rice were derived. The

reduction in genetic diversity relative to *O. rufipogon*, and the formation of two unique genetically reduced groups is expected given what we know about the domestication history of rice.

Most *O. rufipogon* individuals demonstrated an admixed ancestry. Two distinct genetic groups exist within *O. rufipogon*, illustrating that the two groups are genetically diverged from the remainder of *O. rufipogon* individuals. The first of these genetic groups (shown as dark green in Figure 5a) includes five *O. rufipogon* individuals from Malaysia, India and Thailand. The second group (shown in red) includes five individuals from Thailand and Papua New Guinea. A small subset of individuals is also associated with the *indica* genetic cluster.

Bayesian clustering of the 19 southern red rice populations converged on an optimal K=10, suggesting that ten genetic clusters can be found within the 294 red rice individuals analyzed (Supplemental 2). In all analyses, the five independent replications produced the same groupings of individuals.

Because the putative parental taxa established eight as the most likely number of genetic clusters within the *O. rufipogon/sativa* data we analyzed the total dataset with this K value to understand how red rice in the US fits into these parental clusters (Figure 5b). *Oryza rufipogon* was again found to be highly admixed, while the two rice cultivars formed two distinct genetic clusters. Red rice was not homogeneous; individuals fell into four genetic clusters, suggesting multiple origins. Populations in Arkansas (Ar1, 3, 6, 8 and 9) are associated with the *indica* genetic cluster 1 (yellow). The largest red rice genetic group, genetic cluster 3 (purple), shared genetic similarity with two *O. rufipogon* accessions from Malaysia. The other two predominant red rice clusters (cluster 4 &5) showed no substantial association with any of the putative parental taxa sampled.

Private alleles:

Detailed analyses of allele associations among the rice groups provide insight into the interwoven genetic histories of these *Oryza* taxa. To address the question of hybrid origin for red rice (*O. rufipogon* x *O. sativa*) alleles were sorted to identify those which are private to either *Oryza rufipogon* or *O. sativa*. Red rice individuals were screened to assess the proportion of taxon specific alleles shared with the putative parental types (Figure 6). *O. rufipogon* accessions contain 136 private alleles at its 15 loci, of which 44 were shared with red rice individuals representing 14 loci. The *O. sativa* group contained 33 private alleles in 11 of the 15 loci investigated, 13 of which were shared with red rice. Of the 13 *O. sativa* private alleles shared with red rice, one allele was present in both *O. sativa indica* and *O. sativa japonica* cultivars, while 4 alleles were *indica*-specific and 8 alleles were *japonica*-specific. Red rice individuals never simultaneously contained an *O. sativa japonica* specific allele and an *O. sativa indica* specific allele. Red rice individuals did, however, contain multiple alleles from one cultivar type, and alleles private to *O. rufipogon*.

Private alleles were found among comparisons of the local US cultivars as compared to the global *O. sativa* accessions. Twenty-one unique alleles distributed across 11 loci were found in the global *O. sativa* accessions which were absent from the field sampled US cultivars, and 51 unique alleles representing all 15 loci were found exclusively in southern cultivars as compared to the global *O. sativa* accessions. Of the private alleles found in the US cultivars 6 alleles (representing 5 loci) were found exclusively in Texas cultivars, 2 were found (at 2 loci) in Arkansas cultivars and 6 (at 5 loci) were found exclusively in Louisiana cultivars.

DISCUSSION

Red rice in the Southern US is a genetically distinct but highly diverse taxon

Our data are consistent with an Asian hybrid origin of red rice. There is no evidence suggesting a feral origin for the weed in the Southern US. Red rice is a phenotypically and genetically diverse group that is more closely related to itself than either cultivated sub-species, or to wild *O. rufipogon*. The wide range of phenotypes expressed in red rice includes the entire range of phenotypes seen in both *O. rufipogon* and *O. sativa*. The data confirm genetic associations between red rice, *O. rufipogon* and both *O. sativa* cultivar types, but provides no evidence for the paired cultivar/red rice groupings we would expect to see under a de-domestication scenario. Principal coordinate, InStruct and neighbor-joining tree analyses all support the presence of unique genetic groups within red rice, distinct from one another, but more closely related to each other than to the putative parental classes.

Red rice as a taxonomic unit is most genetically similar to *O. rufipogon* relative to either putative cultivar parent (Table2), yet both PCA (Figure 2) and the presence of red rice private alleles suggest that much of the diversity seen in southern red rice is outside the genetic range of *O. rufipogon*. Bayesian clustering confirms the cohesiveness of the red rice populations. If red rice was more similar to *O. rufipogon* or either cultivar then we would expect the InStruct analysis to partition red rice individuals into the parental groups. Instead analyses of the entire data set at a K of eight (Figure 5) suggests that there are four distinct genetic clusters of red rice in the southern US. Membership in each of these clusters is not defined exclusively by region, population membership or individual seed morphology. Bayesian clustering confirmed the PCA and private allele results, recognizing *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* groups as separate from all red rice individuals.

Genetic diversity within red rice

Genetic cluster 3 (purple) is the most widespread within red rice, including populations in all three regions (Texas, Louisiana and Arkansas) and thus contains the majority of individuals. Most *O. rufipogon* individuals have a small portion of their genetic assignment in this cluster, but two of the *O. rufipogon* individuals are predominantly assigned to this group; these include an individual from Malaysia, and an annual *O. rufipogon* from Indonesia. Because these countries are adjacent, this might reflect the geographical source of this genetic group of red rice.

Louisiana 7 is an interesting population, with virtually 100% of its individuals assigned into, genetic cluster 5 (Figure 8). This population was phenotypically uniform; all red rice individuals in the population were straw-hulled and awnless (closely mimicking cultivated rice). Although there is only one population with genetic cluster 5 as its majority assignment, there are individuals from other populations with a majority assignment in this group. This could indicate that gene flow from the homogeneous LA7 region (or a similarly homogeneous population) via the movement of seed between populations is occurring. Despite the genetic similarity between these individuals and the LA7 population, both LA 5 and LA 12 individuals possessed awns unlike LA7.

Five of the seven Arkansas populations are members of the (yellow) *indica*-like genetic cluster, suggesting that red rice in these populations, and in Arkansas as a whole, are the result of hybridization between *O. rufipogon* and *O. sativa indica* cultivars. The Bayesian clustering assignments corroborate the *indica*-like genetic association suggested by the PCA and tree assignments for these populations. Individuals in these *indica*-like populations are not exclusively cultivar-like in their morphology as may be expected from the literature (Gealy *et al.* 2002; Londo and Schaal 2007; Schwanke *et al.* 2008b); instead, both the cultivar-like straw-hull/awnless and black-hull/awned individuals make up these populations.

One of the most admixed populations is LA 13, with all members being partially assigned to at least two genetic clusters and with most being partially assigned to all four groups. The high level of admixture in this population is suggestive of high rates of outcrossing between red rice individuals in this population. More often admixture is seen at the population level, with individuals within populations being assigned primarily to one genetic cluster, but the population contains individuals from multiple genetic groups. This pattern of admixture is more suggestive of seed movement between populations.

Overall the data suggest limited involvement of *O. sativa japonica* in the formation of southern red rice. A single individual found in AK5 had a majority assignment to *japonica*-like genetic cluster 2. Other minimal associations were found in two Texas populations (TX8 and TX13) and in the LA 12 population, possibly the signature of local gene flow. Private alleles were present in red rice, but were generally limited to populations where weak *O. sativa japonica* associations were found, possibly indicating a secondary effect of local gene flow.

Lack of genetic and phenotypic associations

Other studies have not found the same high level of genetic diversity in red rice (Londo and Schaal 2007). This is most likely due to differences in sampling. Previous studies used red rice weed-biotype accession as their red rice group while here we collected field populations.

In the literature the crop mimic and the *O. rufipogon* like groups are described as the primary rice biotypes occurring in the field. Other studies have found a tight association between awn and color phenotypes and ancestry (Gealy *et al.* 2002; Londo and Schaal 2007; Schwanke *et al.* 2008b). They found that both the straw-hull and black-hull types formed distinct groups with different genetic histories. Straw-hulled red rice in these studies were the most genetically similar to *O. sativa indica* cultivars, while the black-hulled individuals were genetically closer to *O. rufipogon*. These studies suggest that the crop-mimic, straw hulled, awn-less variety found in the Southern US has been associated with *O. sativa indica*

cultivars (Vaughan *et al.* 2001), while the black hulled awned type has been associated with *O. rufipogon* accessions.

When we performed analysis on 27 of the 28 red rice accessions used by Londo and Schaal (2007) we found the same patterns of diversity and low levels of variation that they reported, using our (different) panel of SSR markers. In contrast to these results, populations of red rice in the field show all combinations of phenotypic characters, and individuals with phenotypes other than the characterized straw-hull-awnless and black-hull awned types were found. The least frequent morphological type in the field was the black hull awnless type. The straw-hull-awned type was prolific, distributed across the southern range of red rice and grouped with both red rice sub-groups, possibly suggesting strong selection for phenotypes that mimic cultivated hull color. These data do not support a strong association between morphological similarity and genetic marker based similarity or ancestry. There is no association with awn presence or hull color and assignment to the major genetic red rice groups, and as such seed morphology is not an accurate indicator of genetic associations or ancestry.

Evidence for a Hybrid Origin of red rice

In addition to the genetic cohesiveness of red rice and lack of close genetic associations with specific *O. sativa* cultivars, the high levels of diversity found within and among red rice populations do not support a de-domestication hypothesis. The origin of weedy red rice by de-domestication should result in a bottleneck, with possibly only one cultivar giving rise to the weedy rice population. Our results show levels of diversity in red rice that are on par with range wide samples of *O. rufipogon*, further confirmation of a hybrid, not feral origin. PCA on all individuals nests *O. rufipogon* within red rice. The position of *O. rufipogon* in conjunction with the overlap found between *O. sativa* cultivars and red rice are suggestive of hybridization. Placement of *O. rufipogon* and the cultivars on internal nodes of the neighbor-joining tree is also consistent with these findings. These observations are further strengthened

by the co-occurrence of private alleles of *O. rufipogon* and *O. sativa* in red rice individuals. Taken together, the data strongly support a hybrid origin of red rice in Asia.

The patterns of genetic diversity seen in red rice suggest that multiple introductions of genetically distinct red rice have occurred. Populations of red rice in the southern US are more diverse than Chinese populations of hybrid red rice, despite the presence of the wild congeners in the Chinese agricultural environment (Cao *et al.* 2006). The presence of private alleles specific to both *O. sativa indica* and *O. sativa japonica* in red rice suggests both cultivars have played a role in the formation of this weed. The absence of individuals possessing private alleles from both cultivars, in conjunction with individuals carrying *O. rufipogon*-specific alleles argues for at least two independent hybrid origins in Asia followed by separate introductions into the United States.

Further evidence for multiple origins of red rice can be seen from the partitioning of genetic diversity as revealed by the InStruct analysis of the entire data. Red rice is clearly not homogeneous, as might be expected if this taxon originated only once. The partitioning of red rice into four major genetic clusters gives insight into the possible introduction dynamics of the weed.

The data support the hypothesis that US populations of red rice resulted from the admixture of multiple genotypically diverse populations of hybrid (*O. rufipogon* x *O. sativa*) red rice in Asia. Multiple introductions into the US from genetically independent red rice populations in Asia has most likely given rise to the highly genetically diverse red rice populations in the predominant rice growing regions in the United States. A difference in diversity across global red rice populations implies that the dynamics of weedy rice population formation and maintenance may differ by geographic region.

ACKNOWLEDGEMENTS

The authors thank Jason Colvard, Vivian Spangiel and Andrew Colvard for assistance in the field; and Jim Hamrick, Mark Chapman, Vanessa Koelling, and Joan West for helpful comments and suggestions regarding data analysis and the manuscript. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation. The study was funded by a grant from the National Science Foundation Plant Genome Program (to S.R.W.), Research Project Grants-in-Aid for Scientific Research 14360005 and 15380006, Pire grant, and the University of Georgia, Department of Genetics NIH training grant.

REFERENCES

Beadle GW (1980) The ancestry of corn, pp. 112-119, Scientific American.

Buckler ES, Goodman MM, Holtsford TP, Doebley JF, Sanchez J (2006) Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123-134.

Cao QJ, Lu BR, Xia H, *et al.* (2006) Genetic diversity and origin of weedy rice (*Oryza sativa f. Spontanea*) populations found in north-eastern China revealed by simple sequence repeat (SSR) markers. *Annals of Botany* **98**, 1241-1252.

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* **127**, 1309-1321.

Ellstrand NC (2009) Evolution of invasiveness in plants following hybridization **11**, 1089-1091.

Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* **30**, 539-563.

Estorninos LE, Gealy DR, Talbert RE, Gbur EE (2005) Rice and red rice interference. I. Response of red rice (*Oryza sativa*) to sowing rates of tropical japonica and indica rice cultivars. *Weed Science* **53**, 676-682.

Eubanks MW (1997) Molecular analysis of crosses between Tripsacum dactyloides and Zea diploperennis (Poaceae). *Theoretical and Applied Genetics* **94**, 707-712.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution* **23**, 691-700.

Felsenstein J (2007) PHYLIP (Phylogeny Inference Package) version 3.67. Department of Genome Sciences, University of Washington, Seattle., Distributed by the author.

Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**, 1635-1651.

Gealy DR, Bryant RJ (2009) Seed physiochemical characteristics of field-grown US weedy red rice (*Oryza sativa*) biotypes: Contrasts with commercial cultivars. *Journal of Cereal Science* **49**, 239-245.

Gealy DR, Mitten DH, Rutger JN (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, Saldain NE, Talbert RE (2000) Emergence of red rice (*Oryza sativa*) ecotypes under dry-seeded rice (*Oryza sativa*) culture. *Weed Technology* **14**, 406-412.

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

Harlan JR (1992) Crops and Man. American Society of Agronomy, Madison, WI.

Langevin SA, Clay K, Grace JB (1990) The incidence and effects of hybridization between cultivated rice and its related weed red rice (*Oryza sativa*). *Evolution* **44**, 1000-1008.

Li CB, Zhou AL, Sang T (2006) Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*. *New Phytologist* **170**, 185-193.

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

Lu BR, Snow AA (2005) Gene flow from genetically modified rice and its environmental consequences. *Bioscience* **55**, 669-678.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.

Oard J, Cohn MA, Linscombe S, Gealy D, 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.

Olsen KM, Caicedo AL, Jia YL (2007) Evolutionary genomics of weedy rice in the USA. *Journal of Integrative Plant Biology* **49**, 811-816.

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

Pritchard J, Falush D, Stephens M (2002) Inference of population structure in recently admixed populations **71**, 67.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233-234.

Schwanke AML, Andres A, Noldin JA, Concenco G, Procopio SO (2008a) Seed germination and dormancy of red rice ecotypes. *Planta Daninha* **26**, 497-505.

Schwanke AML, Noldin JA, Andres A, Procopio SO, Concenco G (2008b) Morphological characterization of red rice (*Oryza sativa*) ecotypes derived from irrigated rice areas. *Planta Daninha* **26**, 249-260.

Snow AA, Uthus KL, Culley TM (2001) Fitness of hybrids between weedy and cultivated radish: Implications for weed evolution. *Ecological Applications* **11**, 934-943.

Temnykh S, DeClerck G, Lukashova A, *et al.* (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* **11**, 1441-1452.

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

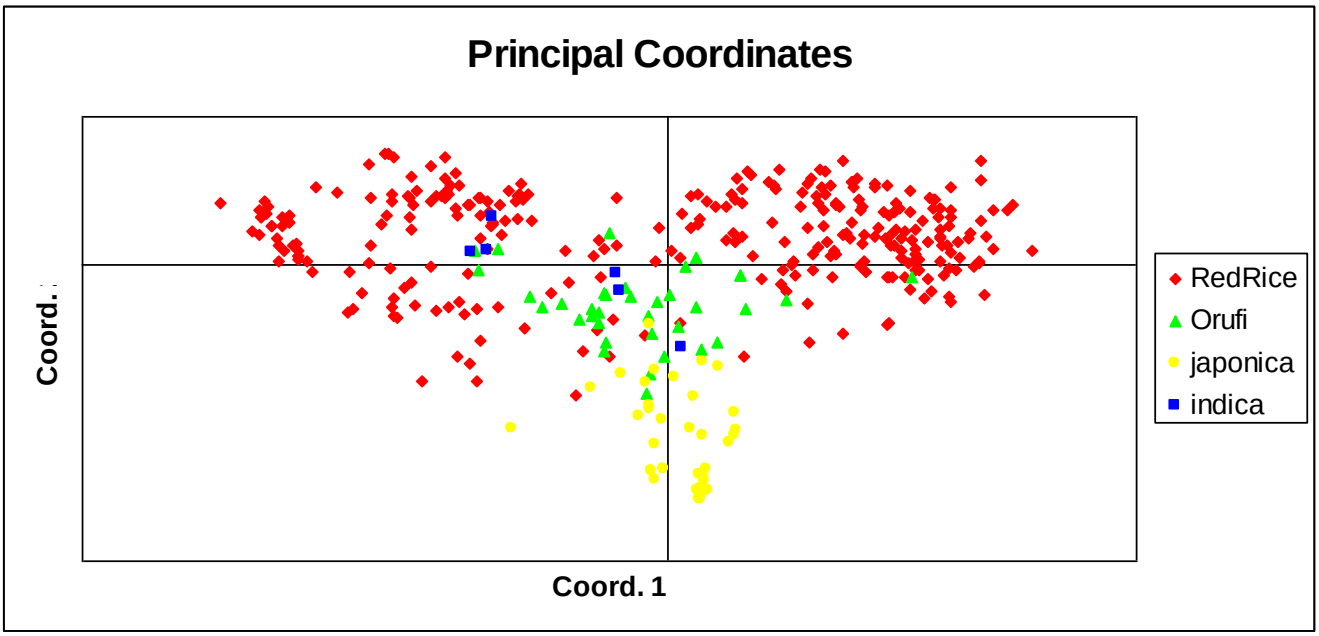


Figure 2.1: Principal Coordinate Analysis of Red Rice, *O. rufipogon, O. sativa japonica* and *O. sativa indica* individuals.

Principal Coordinate Analysis of Red Rice, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals graphically displays the two-dimensional distribution of genetic variation. The PCA explained 50% of the variation present in the data (the third axis accounts for 68%). Symbols correspond to red rice, *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* individuals.

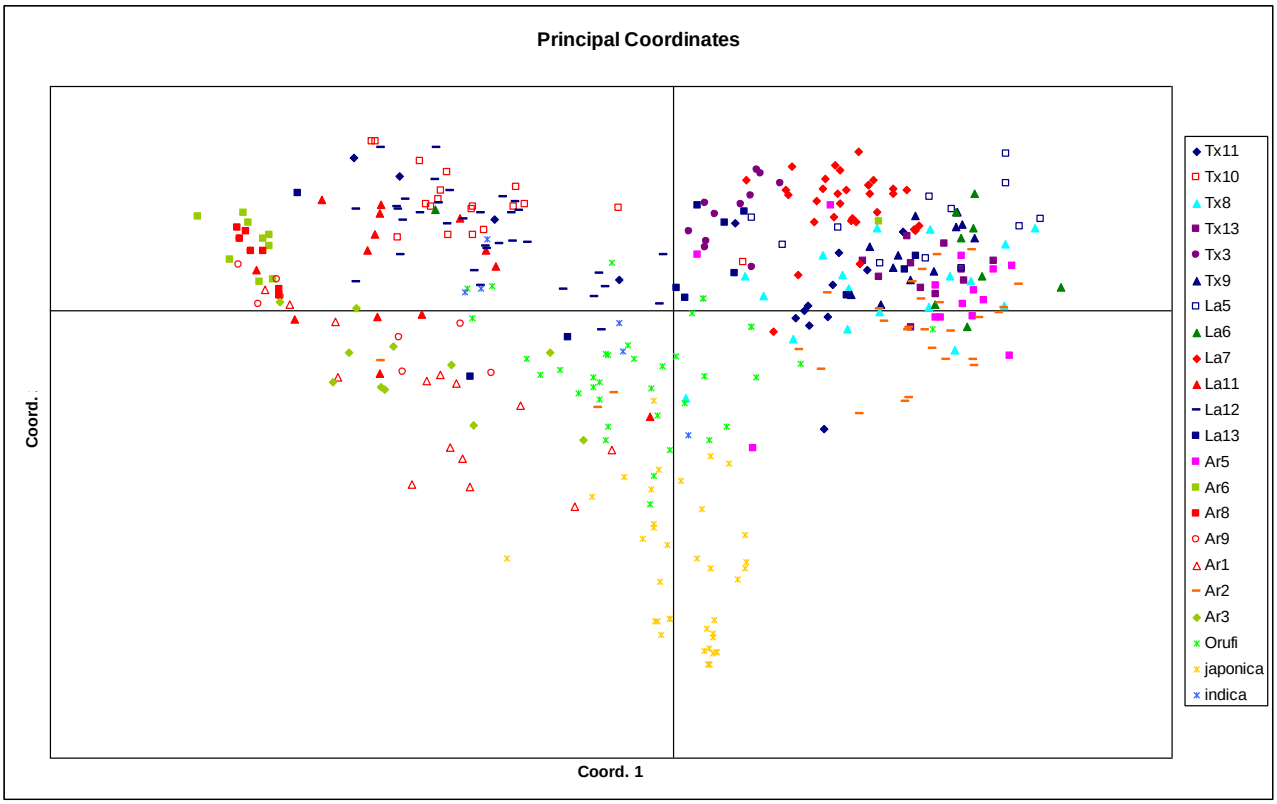


Figure 2.2: Principal Coordinate Analysis of red rice, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals by geographic location.

Two-dimensional PCA explained 52% of the variation present in the data (the third axis accounts for 68%). Symbols correspond to the population of origin for all red rice, *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* individuals; populations are color codes indicating seed morphology. Population seed morphology is color coded: red populations have straw-hull/awnless individuals, dark green populations have black-hull/awned individuals, purple populations have straw-hull/awned individuals, dark blue populations have straw-hull/awned & black-hull/awned individuals, turquoise populations have black-hull/awned & black-hull/awnless individuals, pink populations have straw-hull/awnless & straw-hull/awned individuals, olive green populations have black-hull/awned & straw-hull/awnless individuals, and the orange populations have straw-hull/awnless, straw-hull/awned & black-hull/awned individuals.

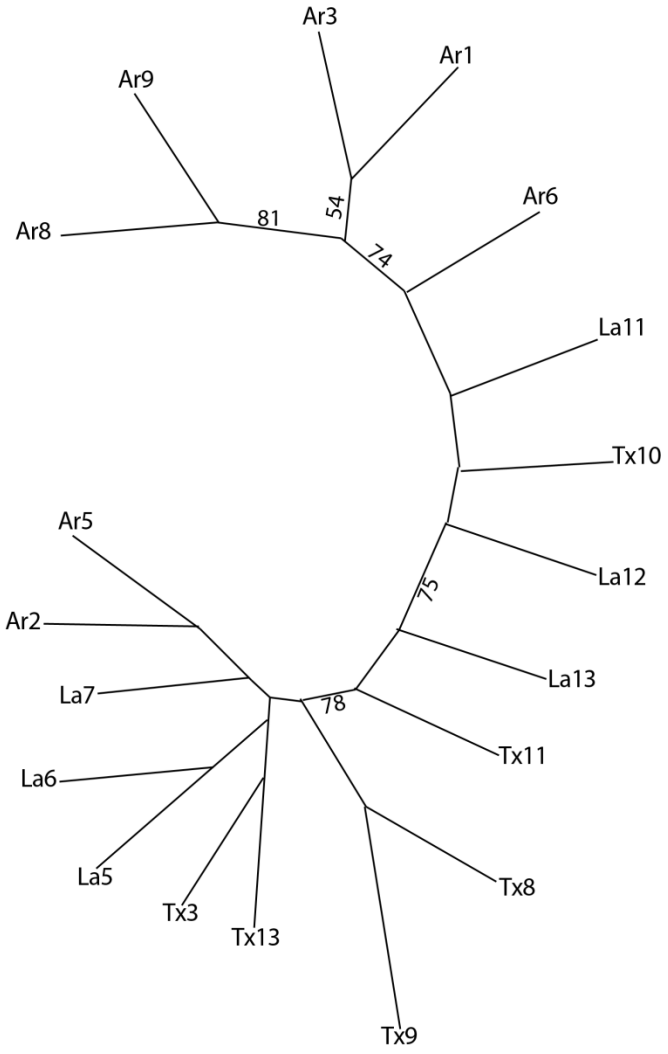


Figure 2.3: Unrooted Neighbor-joining dendogram of Southern US Red Rice populations

Neighbor-joining dendogram based on Nei's (1972) genetic distance values for all Southern US red rice populations constructed using population genotypic data from all 15 SSR loci. Numbers along the branches represent bootstrap support greater than 50% after 1000 replicates.

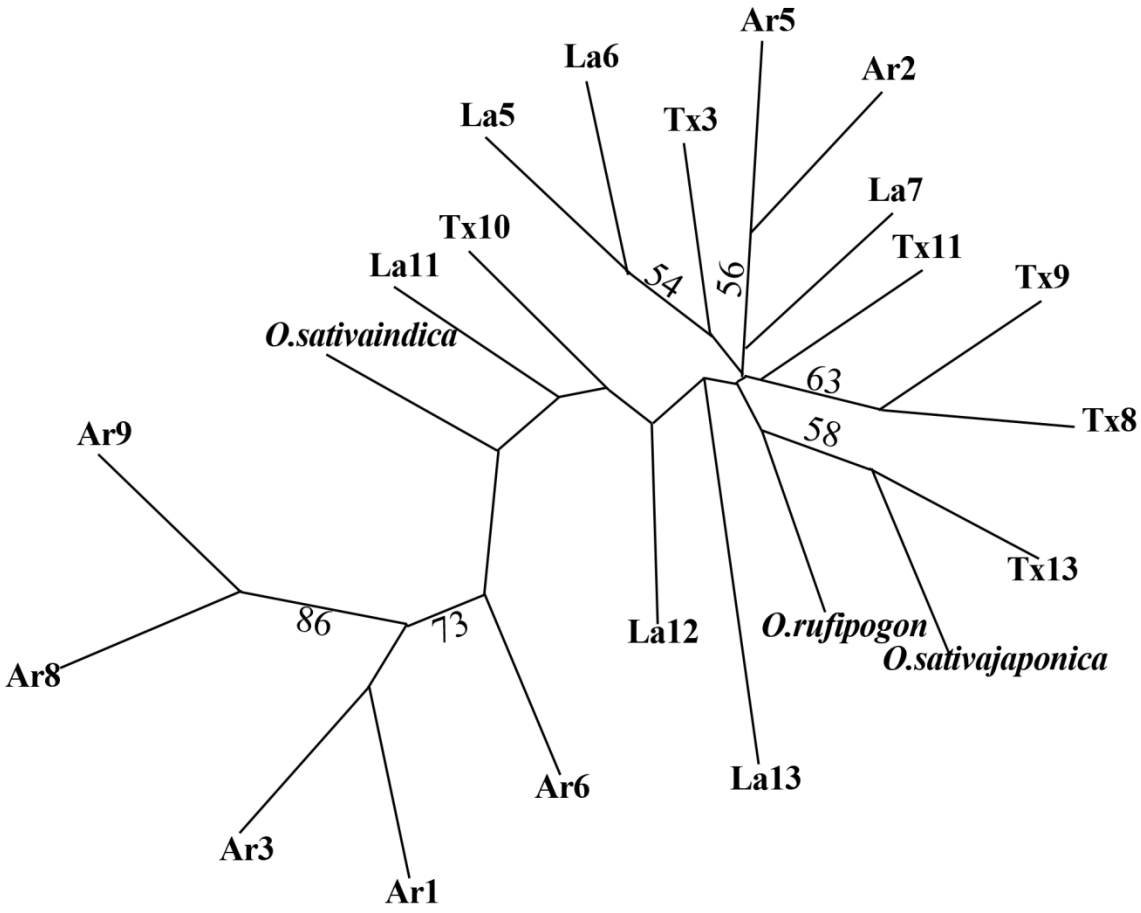


Figure 2.4: Unrooted Neighbor-joining dendogram of all rice groups.

Neighbor-joining dendogram for all rice categories constructed using population genotypic data from all 15 SSR loci. Numbers along the branches represent bootstrap support greater then 50% after 1000 replicates.

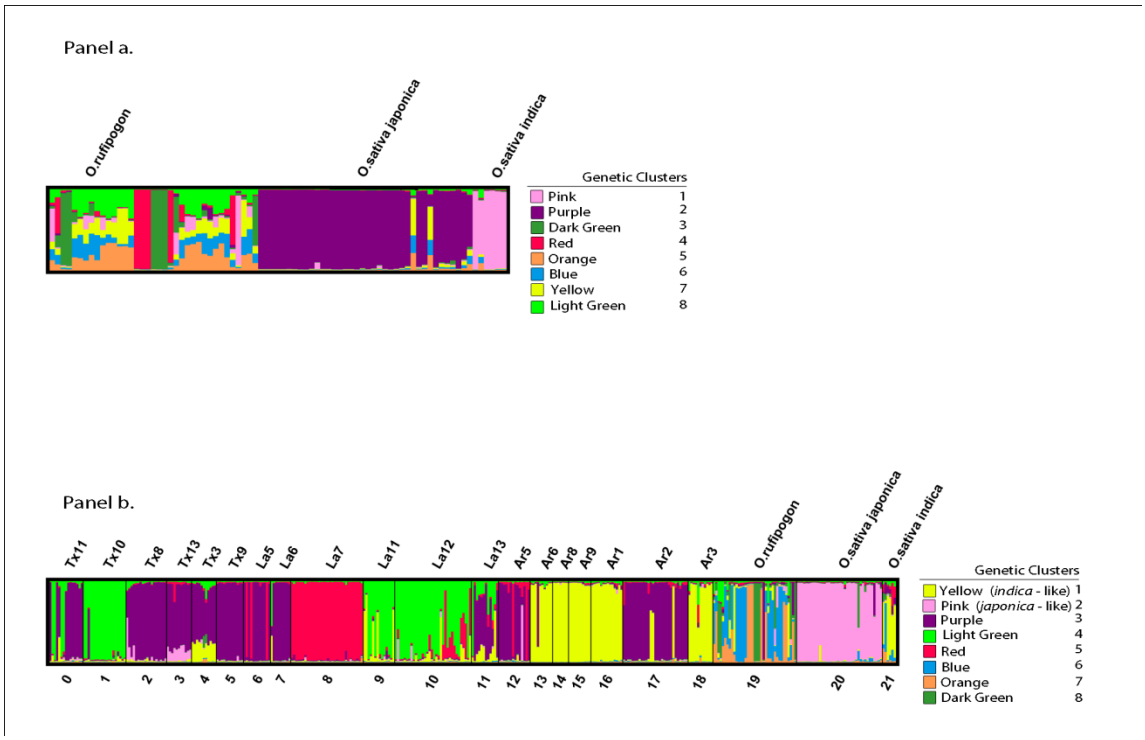


Figure 2.5: Distruct graphical representation of the InStruct results.

Panel a) Results for the parental analysis using *O. rufipogon* and *O. sativa* accessions. An optimal K=8 was found for the putative parental group. Panel b) Results for the total dataset analyzed with K=8.

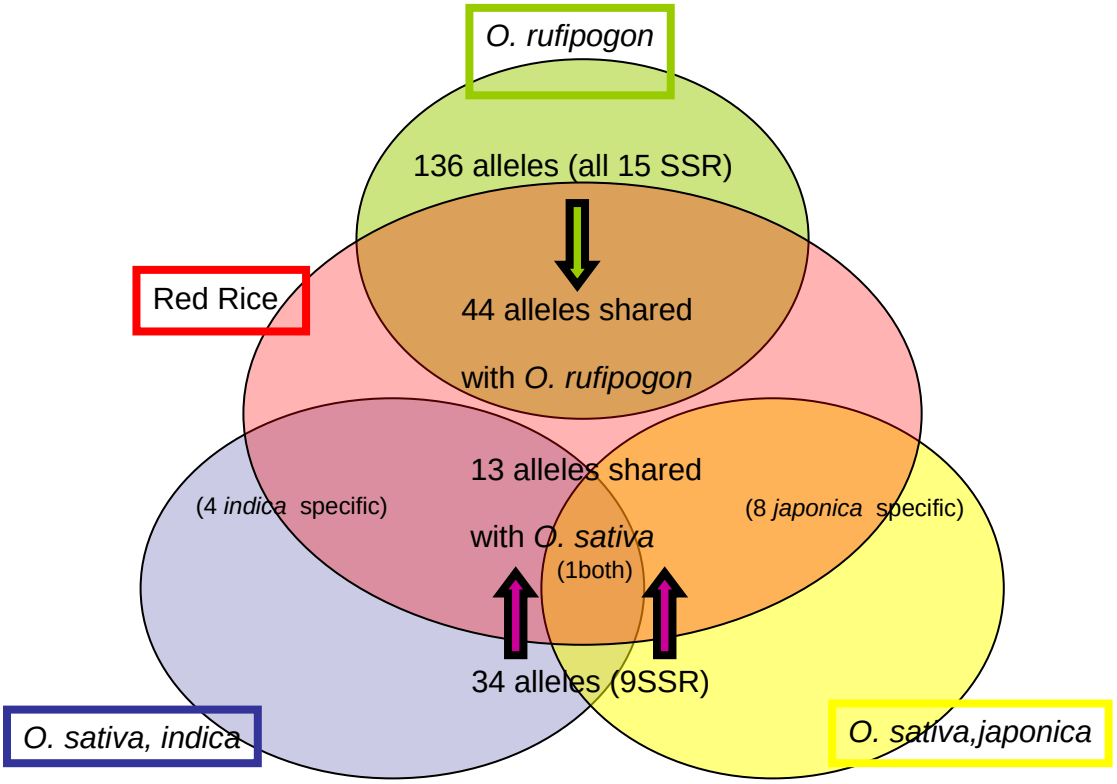


Figure 2.6: Private alleles shared with red rice.

Private allele analysis confirms the occurrence of alleles unique to *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* in red rice individuals. Alleles found in *O. rufipogon* which are absent in cultivated *O. sativa* are represented in the green domain, while alleles found exclusively in *O. sativa* are represented in the blue (*indica*) and red (*japonica*) domains. Only one *O. sativa* private allele was found in both cultivars. The yellow domain contains the private alleles from each group that are shared with red rice.

Population Identification	Number Individuals	GPS coordinates	Seed Morphology
TX11	15	N29° 33.8281 W95°.804055	strawhull/awn &blackhull/awn
TX10	19	N26°.127006 W98°23.198	strawhull/awnless
TX8	18	N29°33.750 W96°31.421	blackhull/awnless & black/awn
TX13	11	N29°12.314 W96°20.997	strawhull/awn
TX3	11	N29°11.859 W96°20.726	strawhull/awn
TX9	12	N29°11.859 W96°20.726	strawhull/awn &blackhull/awn
LA5	12	N32°38.898 W91°16.309	strawhull/awn &blackhull/awn
LA6	9	N31°49.195 W91°23.198	blackhull/awn
LA7	32	N31°33.750 W91°31.421	strawhull/awnless
LA11	14	N30°12.314 W92°20.997	strawhull/awnless
LA12	34	N28°11.859 W92°20.726	strawhull/awn &blackhull/awn
LA13	12	N26°10.716 W92°19.543	strawhull/awn &blackhull/awn
AR5	14	N34°28.571 W91°26.61	strawhull/awnless & strawhull/awn
AR6	10	N34°28.580 W91°26.296	strawhull/awnless & blackhull/awn
AR8	7	N34°28.261 W91°24.225	strawhull/awnless
AR9	10	N34°28.652 W91°24.199	strawhull/awnless
AR1	14	N34°07.626 W91°15.991	strawhull/awnless
AR2	29	N34°05.011 W91°22.349	strawhull/awnless & strawhull/awn & blackhull/awn
AR3	11	N34°14.339 W91°17.681	strawhull/awnless & blackhull/awn

Table 2.1: Southern red rice population information, including number of individuals sampled, location and seed phenotypes found.

Taxonomic group	P_p(%)	A_p	AP_p	A_e	H_e	H_o
Red Rice	100.00%	241	16.070	3.023	0.645	0.085
O. sativa japonica	93.33%	71	5.070	2.247	0.401	0.008
O .sativa indica	86.67%	39	3.000	2.200	0.470	0.000
O. rufipogon	100.00%	225	15.000	8.403	0.855	0.216
Mean	95.00%	144	9.780	3.968	0.593	0.077

Table 2.2: Species level genetic diversity comparisons. P_p = percent polymorphic loci; A_p = total number of alleles per taxon (including monomorphic loci); AP_p = mean number of alleles per polymorphic locus; A_e = mean effective number of alleles per polymorphic locus; H_e = genetic diversity (expected heterozygosity); H_o = observed heterozygosity.

Pairwise F_{ST} (via Frequency) for all Rice Groups

	Red Rice	*O. rufipogon*	*O. sativa japonica*	*O. sativa indica*
Red Rice	0.000			
O. rufipogon	0.068	0.000		
O. sativa japonica	0.224	0.187	0.000	
O. sativa indica	0.175	0.157	0.384	0.000

Table 2.3: Pairwise F_{ST} values between rice groups. Low F_{ST} values indicate a high level of similarity between red rice and *O. rufipogon*. High F_{ST} values between *O. sativa indica* and *O. sativa japonica* cultivar groups suggest that these groups are most differentiated from one another.

Population	P_p (%)	A_p(%)	AP_p	A_{ep}	Mean I	H_{op}	H_{ep}
Tx11	100.00%	55	3.670	2.374	0.919	0.069	0.511
Tx10	93.33%	49	3.430	1.570	0.571	0.048	0.305
Tx8	86.67%	47	3.350	1.810	0.687	0.056	0.381
Tx13	53.33%	30	2.880	1.550	0.318	0.037	0.176
Tx3	66.67%	40	2.670	1.700	0.478	0.101	0.248
Tx9	66.67%	39	3.400	2.250	0.533	0.082	0.288
La5	80.00%	44	3.230	1.805	0.638	0.063	0.349
La6	93.33%	43	3.000	1.864	0.692	0.047	0.392
La7	93.33%	58	3.800	1.566	0.555	0.085	0.271
La11	86.67%	52	3.850	1.893	0.688	0.098	0.356
La12	100.00%	74	4.930	1.824	0.777	0.102	0.381
La13	100.00%	48	3.200	1.951	0.766	0.087	0.430
Ar5	86.67%	52	3.640	1.643	0.668	0.147	0.344
Ar6	93.33%	37	2.570	1.464	0.477	0.051	0.265
Ar8	40.00%	22	2.170	1.165	0.178	0.048	0.107
Ar9	73.33%	36	2.910	1.569	0.485	0.084	0.272
Ar1	73.33%	46	3.580	1.903	0.679	0.093	0.355
Ar2	100.00%	86	5.730	2.924	0.963	0.104	0.500
Ar3	80.00%	53	4.170	2.026	0.779	0.172	0.405
Mean	82.46%	47.9	3.483	1.834	0.624	0.083	0.333
Pooled pops							
TX	100.00%	117	7.800	2.806	1.212	0.063	0.600
La	100.00%	136	9.067	2.348	1.132	0.084	0.540
Ar	100.00%	154	10.267	3.015	1.359	0.109	0.628

Table 2.4: Genetic diversity statistics for southern red rice populations

P_p = percent polymorphic loci; A_p = total number of alleles per population (including monomorphic loci); AP_p = mean number of alleles per polymorphic locus per population; A_{ep} = mean effective number of alleles per polymorphic locus per population; I = genetic identity; H_e = genetic diversity (expected heterozygosity); H_o = observed heterozygosity.

Summary AMOVA Table

Source	df	Est. Var.	%
Among Regions	2	0.304	6%
Among Pops/Regions	16	1.965	36%
Within Pops	569	3.188	58%
Total	587	5.456	

Table 2.5: AMOVA: Southern red rice populations

Analysis of molecular variance shows that most variation in red rice is present within the populations (58%). Populations within regions contain some variation (36%), while 6% of the variation is partitioned among regions.

CHAPTER III:

NO EVIDENCE FOR A CHINESE ORIGIN OF RED RICE IN THE UNITED STATES[2]

[2] Kuntz, E. J., Mauricio, R. To be submitted to Heredity.

ABSTRACT

The movement of organisms around the globe has had a profound impact on the health and diversity seen in the habitats and communities that they are introduced, often becoming invasive and/or weedy in their new environments. Understanding the introduction history, including source population localities and the dynamics of invasion is important in order to generate hypotheses regarding the environmental and evolutionary factors responsible for the successful establishment and persistence of invading organisms. Red rice is a noxious weed of rice agriculture, causing significant yield loss in United States rice production. A complex evolutionary history places the origin of red rice in Asia, arising as the result of crop to weed hybridization between cultivated rice (*Oryza sativa*) and wild rice (*Oryza rufipogon*). A long and well documented history of weed and invasive plant movement between the United States and China, in addition to the reported existence of hybrid derived red rice populations in China lead us to investigate populations of Chinese red rice in an attempt to locate the source population(s) of US red rice. We used populations of red rice from China which represent the diversity seen in red rice across China to test for putative source populations. Our results provide no evidence for a Chinese origin in the population of red rice sampled. Direct comparisons between populations of red rice in the United States and China surprisingly show no genetic overlap. These comparisons also indicate that US populations are more diverse then their Chinese counterparts.

INTRODUCTION

The intentional and unintentional transportation of plants, animals, insects and other organisms around the globe has had a major impact on the biological world around us. A minutia of these biological globetrotters become invaders in their introduced environments, affecting the native biology, diversity and health of the communities into which they are introduced (Sax et al., 2005). The roles of humans in the movement of biodiversity across the globe are widely recognized; acting as unintentional vectors relocating organisms as a byproduct of international trade, travel and global transport, or via intentional introductions such as with horticultural specimens, and agricultural plant and animals.

Several factors influence the ability of these organisms to persist and eventually thrive in their introduced environment, including but not limited to, the number of introductions, genetic diversity in founding populations, hybridization with native flora, and generalized life history traits such as, mating system requirements (Bacigalupe, 2009; Dlugosch and Hays, 2008; Ellstrand et al., 1999). Organisms that persist in their introduced localities may become invasive and/or weedy pests in their introduced range. Although many factors contribute to an organism's ability to persist in the introduced area and ultimately become invasive or weedy, their ability to respond to selection and adapt to their new environments is heavily dependent on genetic variability (Facon et al., 2008; Leung et al., 2004). Genetic variation for ecologically important traits may increase the likelihood that an invasive and/or weedy population will be able to survive and respond to changes in their new environment (Kolbe et al., 2004; Kolbe et al., 2007; Lachmuth et al., 2010; Lavergne and Molofsky, 2007). Multiple introductions from either one or many source populations (Kolbe et al., 2004; Pairon et al., 2010), and/or hybridization with closely related native species in the introduced range are the two most common ways introduced populations acquire genetic variation (Ellstrand, 2009a; Ellstrand and Schierenbeck, 2006; Ellstrand, 2009b).

Many agricultural species are grown in areas outside their progenitor's native range. This is the case with our species of interest, cultivated rice *Oryza sativa*. Red rice, the major agricultural pest of rice cultivation worldwide, co-occurs with rice cultivation in all areas where rice is grown. The evolutionary

origins of red rice around the globe remain unclear; some populations appear to be the result of de-domestication (where a cultivated types become feral by either, losing domestication traits, or re-acquiring wild traits (Jing et al., 2007), while others red rice populations appear to be the product of hybridization between cultivated and wild rice. Populations of red rice in the southern United States have been shown to have a hybrid evolutionary history, formed by the crossing of *Oryza sativa* the cultivated rice and *Oryza rufipogon*, the wild progenitor of cultivated rice (Kuntz Chapter 2). Since the native range of *O. rufipogon* is in Asia, weedy red rice found in the US was likely introduced from Asia. The goal of this research is to clarify the Asian source of that introduction.

Understanding historic routes of invasion is important for generating and testing reasonable hypotheses regarding the environmental and evolutionary factors responsible for the successful establishment and persistence of invading organisms (Estoup and Guillemaud, 2010). Understanding the geographic origin of source populations and the routes of invasion of invasive pests and weeds will help researchers design strategies to prevent continual introductions. Once specific geographical sources are identified for US red rice populations, areas prone to generating weedy red rice can be further examined and tested to uncover the underlying mechanisms generating these weedy plants. For example, promiscuous cultivars that readily outcross with the wild *O. rufipogon* or cultivation practices that increase the chance of weed formation via de-domestication can be identified and used to design measures that will function to limit the generation of additional weedy red rice biotypes in the future. Understanding the underlying factors that contribute to weed generation, dissemination, and persistence, help us to address important questions in evolutionary biology, specifically: which mechanisms are involved in adaption and which environmental and evolutionary factors that influence the success of biological invasions and weed generation.

In this study, we examine Chinese populations of red rice in order to investigate the possible origins of red rice in the United States. Knowledge of the geographical routes invasive and/or weedy pests have taken from their source populations to their introduced areas will aid in understanding the history of

invasion, the origin of genetic variation, and the genetic composition of the invaders in the introduced range (Estoup and Guillemaud, 2010). We investigate the potential red rice sources in China by comparing genetic variation in both the US and China using SSR molecular markers. We use these data to reveal patterns of population structure and genetic diversity among populations of both regions in order to identify genetic signatures that would suggest one or more Chinese source populations for US red rice.

MATERIALS AND METHODS

Study System

Red rice is one of the most problematic agricultural pests in the world, causing up to 80% yield loss in fields with serious infestations (Gealy and Bryant, 2009). Red rice can be found in all rice growing regions and is extremely problematic throughout Asia, Europe, the United States, Central and South America. Red rice, also known as weedy rice, is classified as *Oryza sativa* due to its sexual compatibility with cultivated rice (Vaughan et al., 2001). Although red rice is able to readily cross with cultivated rice, it is morphologically distinct and highly morphologically variable in comparison to cultivated varieties. Red rice displays variation both within and among populations for height, seed morphology (color, pubescence, grain size and awn presence), date to first flower, seed dormancy, herbicide resistance, and tillering capacity (Gealy and Bryant, 2009; Gealy et al., 2002). The exact origin and evolutionary history of red rice is still debated, and the uniformity of origin from one rice growing region to another is unclear. A recent study suggests that the weed populations found in the Southern United States are hybrids (*O. sativa* X *O. rufipogon*), likely originating in Asia, the native range of the range limited parent (Kuntz Chapter 2).

A large portion of Chinese rice cultivation lies in the northern range of *Oryza rufipogon*, which extends through most of Southeast Asia with western populations in India and southern most populations occurring in Indonesia and Papua New Guinea (Figure 3.1). Throughout its native range *O. rufipogon*, the

wild progenitor, experiences population degradation due to changes in local farming practices, urbanization claiming wild habitats and other generalized human disturbances (Gao, 2004). In the northern areas of its natural range *O. rufipogon* is considered an endangered species (Gao and Zhang, 2005; Song et al., 2006). Perennial populations of *O. rufipogon* are wind-pollinated, self-compatible plants with a mixed mating system that also reproduce clonally via ramet formation (Barbier et al., 1991; Morishima and Barbier, 1990; Oka and Morishima, 1967; Song et al., 2003a; Xie et al., 2001; Zhou et al., 2003). Variation in the portion of sexual and asexual reproduction found in Chinese populations is correlated with latitude, with outcrossing increasing as latitude increases, suggesting an increase in sexual reproduction in the margins of the species range (Gao, 2004; Gao et al., 2000; Gao and Zhang, 2005). As these northern portions of *O. rufipogon's* range experience higher levels of outcrossing and co-occur with cultivated rice they have the potential to generate hybrid red rice biotypes.

In China rice cultivation is found in two regions; one small area in the north located in the Liaoning province, and a larger area encompassing most of the southern provinces. In this area populations of *O. rufipogon* are found sympatric to cultivated rice fields throughout the larger southern rice growing region, however its range limits exclude it from the smaller northern, Liaoning, rice growing area. Population structure and diversity data from Chinese *O. rufipogon* populations demonstrate that populations near cultivated fields contain rare alleles that are commonly present in rice cultivars growing in the region. These data suggest introgression from cultivated rice in to *O. rufipogon* in areas of sympatry (Song et al., 2003a; Song et al., 2006; Xu et al., 2006). In addition to *O. rufipogon* populations near cultivated fields containing cultivated alleles, morphological traits typically found only in cultivated varieties of rice, such as the loss of awn organs, are also seen in co-occurring populations, which further suggests introgression of cultivated alleles into *O. rufipogon* populations (Song et al., 2003b).

Red rice is found in both the northern and southern rice growing areas in China. The northern Liaoning rice growing area grows exclusively *O. sativa japonica* rice. The larger southern cultivation region predominantly utilized *O. sativa indica* cultivars. Data suggests that northern Liaoning populations

of red rice are *O. sativa japonica* de-domesticated types arising from local *japonica* cultivars (Cao et al., 2006; Yu et al., 2005), while the data are mixed regarding the genesis of red rice weeds in the southern region; some reports suggesting de-domestication (Zhang et al in prep.) while others suggest a hybrid origin (*O. sativa* x *O. rufipogon*) (Jing et al., 2007). The exact origins of US rice cultivars are debated, but the co-occurrence of *O. rufipogon* with cultivated rice and a history of genetic exchange between the taxa make China a logical place to look for the source populations of US red rice.

Sampling

Populations of red rice from the United States were collected from fields in July of 2006. A total of 19 populations from three distinct regions (Texas, Southern Louisiana, and Arkansas) were collected for a total of 294 red rice individuals (Table 3.1). All field samples were collected just prior to the rice harvest to ensure seed maturity. Single panicles were harvested at one meter intervals to prevent sampling clonal ramets. One seed per panicle was germinated and prepped for DNA extraction.

Additionally, twenty-seven US red rice accessions were obtained from the Dale Bumpers Rice Research Center in Stuttgart, Arkansas. Each accession consisted of a single plant collected between 1994 and 1999 from across the southern growing region, which represents Missouri, Arkansas, Mississippi, Texas and Louisiana. This collection of plants is maintained at the Dale Bumpers Rice Research Center and is used as a reference panel for studying red rice diversity in the US. The accessions are propagated via single seed descent and propagated for scientific use. For additional details regarding collection and propagation see Gealy *et al* (2002).

Chinese populations of red rice were obtained as extracted DNA from the Weed Research laboratory of Dr. Sheng Qiang at the Nanjing Agricultural University in Nanjing China. A total of 88 individuals from 4 populations representing the two major rice growing regions in China are included in this study. Preliminary data generated by the Sheng lab suggests that these 4 populations represent the genetic diversity of Chinese red rice (Zhang et al in prep). The northern and southern growing regions are

climatically distinct, with the northeastern rice growing region only cultivating *japonica* cultivars (represented by 27 individuals from Liaoning), and the larger southern rice growing region predominantly cultivating *indica* cultivars (represented by 21 individuals from Jiangsu, 20 individuals from Guangdong, and 20 individuals from Hainan) (Figure 3.2).

In addition to red rice populations from the southern US and China, 83 *O. sativa* cultivars, 45 *japonica* and 38 *indica* cultivars from both China and the US and 37 *O. rufipogon* accessions were included in the study (Table 3.1). The *O. rufipogon* samples were collected from populations across Southeast Asia, representing much of the native range. These seeds were propagated by the International Rice Research Institute (IRRI) and acquired as extracted DNA from the Barbara Schaal lab at Washington University, St. Louis USA.

DNA extractions

All seeds were germinated on Whatman filter paper in Petri dishes and grown in a growth chamber at 30°C. Seedlings were collected after 7-14 days, were snap frozen in liquid nitrogen and total genomic DNA was extracted using Qiagen's DNAeasy mini kit (Valencia, CA). Extracted samples were quantified using a spectrophotometer (NanoDrop/ Wilmington, DE) and standardized to approximately 30ng DNA per µl.

Microsatellite genotyping

Sixteen microsatellite loci (simple sequence repeats, or SSRs) representing all 12 chromosomes were chosen from previously identified microsatellite regions (Temnykh et al., 2001). Polymerase chain reaction (PCR) amplifications were performed using three primers: a SSR sequence specific forward primer with an M13 tail at its 5' end (5'- CACGACGTTGTAAAACGACA-3'), a SSR sequence specific reverse primer and a fluorescent-labeled M13 primer (Schuelke, 2000). Thirteen microliter reactions contained 1.5µl of DNA sample, 0.1µl 10 mM forward SSR primer, 0.5µl 10 mM reverse SSR primer,

0.5µl 10 mM M13 primer, 1.2µl 10 mM dNTPs, 1.2µl 10 mM MgCl2, 1.2µl 10XPCR buffer, and one unit Taq polymerase. Samples were amplified using the following PCR profile: initial denaturation at 94°C for 3min; followed by 35 cycles of: 94°C for 45 sec, 55°C annealing for 45sec, 72°C extension for 45sec; and a final extension at 72°C for 15 min. Post PCR, samples were held at 4°C. The resulting PCR fragments were diluted 1:10 and genotyped on a capillary electrophoresis sequencer ABI3700 (Applied Biosystems; Foster City, Ca) at the UGA sequencing facility using MapMarker1000 ROX (BioVentures Inc; Murfreesboro, TN) as an internal size standard.

Genotypic analysis

All microsatellites loci were scored using the software program GeneMarker, version 1.70 (SoftGenetics; State College, PA). Samples were subjected to randomization and 50 US red rice individuals and 40 Chinese red rice individuals were re-sampled to confirm genotype fidelity. All produced consistent genotypes at the 16 scored loci.

Descriptive population genetic statistics were calculated; including: average number of alleles per locus, percentage of polymorphic loci, gene diversity [as measured by Nei's expected heterozygosity (H_e)](Nei, 1978), and F_{IS} as a measure of individual inbreeding, using the software program GENALEX (Peakall and Smouse, 2006). To graphically represent the relationships between taxa, genetic distance values calculated in GENALEX were used to perform a Principal Component Analysis (PCA), where a dissimilarity matrix of pairwise differences between samples is created and the total variation between those samples is compressed into a limited number of dimensions using eigen value analysis. The compressed variation is then plotted visually in two dimensions with the first dimension (x-axis) accounting for the majority of variation seen in the samples, and the second dimension (y-axis) accounting for the second most descriptive dimension of variation in the samples.

Rarefaction was used to address differences in sampling size across the US and Chinese red rice populations. Rarefaction is a method that aims to normalize data so that inconsistencies in biological

measurements driven exclusively by differences in sampling size are corrected, in order to accurately compare groups. Using rarefaction techniques to address inconsistencies in biological sampling was originally utilized by Sanders (1968) to estimate the number of species in a given area for conservation purposes in marine benthic diversity. The technique was extended to include the ability to compare differences in population genetic measures, most specifically comparisons of genetic diversity in unequally sized samples (Mousadik and Petit, 1996). This method specifically, standardized measures of allelic richness across samples of unequal size such that diversity in allelic richness can be used to compare diversity among the samples and major differences in diversity driven by sampling can be discovered by comparing measures of genetic diversity (H_e) and allelic richness. This method is employed in order to verify that differences in sampling between our Chinese populations (N = 88) and our US populations (N = 294) do not distort our ability to make valid comparisons of genetic diversity across the two regions. Rarefaction was used to measure allelic richness as per the methodology employed by El Mousadik and Petit (1996) using the software package FSTAT.9.3.2 (Goudet, 1995).

Spatial patterns of genetic variation and geographic associations between US and Chinese populations were analyzed using an analysis of molecular variance (AMOVA), as executed by GENALEX. The AMOVA analysis was used to calculate F_{ST} among regions, between populations within regions and within populations (Peakall and Smouse, 2006). Populations of red rice in the US were subdivided into regional groupings based on natural geographic boundaries and the cultivation of distinct rice cultivars in each area. In the southern US three regions were investigated: Texas, Arkansas and Louisiana. Populations from China were also subdivided based on geographic boundaries, corresponding generally to differences in cultivar types grown, with *japonica* cultivars grown in the northeastern region and *indica* cultivars predominantly grown in the southern areas (Zhang et al. in prep).

Neighbor-joining dendrograms were built using the NEIGHBOR program in PHYLIP 3.67 (Felsenstein, 2006; Felsenstein, 2007). We used Nei's genetic distance between populations as our input and sample input order was randomized (Nei, 1978). The data were sub-sampled to generate 1000 data

sets using the BootSeq application in PHYLIP 3.67 and support for the final tree was calculated with 1000 bootstrap replicates using the CONSENSE application.

The program InStruct (Falush et al., 2003; Gao et al., 2007; Pritchard et al., 2002) was used to infer population structure in the entire data set, including: *O. rufipogon* accessions, *O. sativa indica* and *japonica* cultivars, the southern US and Chinese red rice samples. InStruct limits the number of false signals of population substructure generated in the STRUCTURE program in predominantly selfing taxa. Given the high selfing rate across the *Oryza* genus InStruct program is more appropriate for our analysis. InStruct uses a Markov Chain Monte Carlo algorithm in a Bayesian framework to estimate the number of populations and the clustering of individuals in those populations based on linkage disequilibrium between the genetic markers. This information is used to assign individuals or portions of individuals into a hypothetical number of populations (K).

Separate InStruct analyses were performed for the Chinese red rice populations with the cultivar and the wild *O. rufipogon* samples, for the US red rice populations with the cultivar and the wild *O. rufipogon* samples, and for red rice as a group; including all 19 US and the 4 Chinese red rice populations. All runs used an initial burn in of 100,000 and with two independent MCMC chains run for a length of 200,000 iterations. We tested a range (1-30) of K values (number of population clusters) using 10 independent runs for each K in each analysis. The most significant K value was assigned using a log likelihood ratio test (Evanno et al., 2005). Sample population of origin information was not used and all runs were performed using both the admixture and no admixture models. The data are displayed using the Distruct program (Rosenberg, 2004) .

Private alleles analysis was used to address associations between the putative parental taxa of red rice in both the US and China. Alleles specific to *O. rufipogon* and the cultivated *O. sativa* (both *indica* and *japonica* cultivars grouped together) were first identified and the presence and distribution of these alleles in all red rice individuals across both regions was assayed. The presence of cultivar and *O.*

rufipogon private alleles was documented and the sharing of specific private alleles between US and Chinese red rice populations was recorded. In order to look more closely at the contribution and influence of either *indica* or *japonica* cultivars to the red rice populations in both regions we assayed for private alleles held between the two cultivar groups and assayed the red rice samples for their presence or absence, again recording which of the alleles are held in common between the US and Chinese red rice groups. The cultivar private alleles were further limited by the removal of alleles private to one cultivar type but also found in *O. rufipogon*. This more specific set of exclusive cultivar alleles was used to verify cultivar contributions to red rice.

RESULTS

Microsatellite analysis

All rice groups; red rice, *Oryza sativa indica*, *O. sativa japonica* and *O. rufipogon*, are polymorphic at all 16 microsatellites scored (Table 3.2). Comparing red rice populations by region shows that while the US red rice populations are polymorphic at all loci, the Chinese populations are not, with 81.25% polymorphism. US red rice populations contain the largest total number of alleles (A_p = 285), followed by *O. rufipogon* with 239 alleles. Chinese red rice (A_p = 96) and both cultivars (*O. sativa japonica* A_p = 96 and *O. sativa indica* Ap = 106) had fewer than half the total alleles present in either US red rice or *O. rufipogon*. The mean number of alleles per polymorphic locus is also highest in US red rice (AP_p = 17.80) although, not significantly different from the number observed in *O. rufipogon* (AP_p = 14.94). Chinese red rice populations (AP_p = 7.39) and both *O. sativa japonica* and *indica* cultivars (AP_p = 6 and 6.63 respectively) had fewer alleles at polymorphic loci than US red rice and *O. rufipogon*. Although US red rice had more total alleles and the highest mean number of alleles per polymorphic locus than any other group, *O. rufipogon* has the highest number of effective alleles per locus (A_e = 8.498). Both rice cultivars, Chinese red rice and US red rice did not significantly differ from one another in the number of effective alleles per locus despite US red rice having over twice as many alleles in total.

As expected *O. rufipogon* had both the highest genetic diversity (H_e = 0.858) and highest observed heterozygosity (H_o = 0.221). Genetic diversity in US red rice (H_e = 0.658) and *O. sativa* indica (H_e = 0.640) is higher than that observed in Chinese red rice (H_e = 0.477) and *O. sativa japonica* cultivars (H_e = 0.453). Diversity, as measured by heterozygosity, is significantly lower ($p < 0.0001$) in the Chinese populations as compared to US red rice (Table 3.2). Northern and southern red rice populations in China were not significantly different from one another, but the Northern Chinese region is significantly different from all three US red rice regions (TX, La and Ar), while southern Chinese region is significantly different from the La and Ar regions (Table 3.5). Observed heterozygosity is lower than expected heterozygosity in all groups, as we would predict for these predominantly selfing taxa.

Genetic similarity is highest between US red rice and *O. rufipogon* (F_{ST} = 0.065) and lowest between Chinese red rice and *O. sativa japonica* (F_{ST} = 0.297). These results are contrary to our expectations, as we predict that the highest genetic divergence would be between the two cultivated groups, *O. sativa japonica* and *O. sativa indica* due to their separate evolutionary histories, which result from two isolated domestication events (Cheng et al., 2003; Liu et al., 2007; Londo et al., 2006). Instead we find that the Chinese red rice populations are most genetically differentiated from *O. sativa japonica* cultivars, and least differentiated from *O. rufipogon* (F_{ST} = 0.183) relative to the other rice group. Interestingly, despite Chinese red rice being most genetically similar to *O. rufipogon*, *O. rufipogon* shares the least genetic similarity with the Chinese red rice populations as compared to the other rice groups. Although US red rice populations are most similar to *O. rufipogon*, suggesting an Asian origin, they are found to be least genetically similar to Chinese red rice populations, concurrently suggesting that Chinese populations are not the source populations for the red rice we find in the United States.

Rarefaction analysis

Rarefaction analysis confirmed that unequal sampling between the US and China did not affect our ability to effectively compare diversity between regions (Figure 3.3). Although measures of H_e was significantly different between US and Chinese populations of red rice, comparisons of allelic richness

not significantly different from one and other (Table 3.4). Allelic richness (R') was, as expected, highest in *O. rufipogon* (R' = 11.565). Northern Chinese red rice was the least diverse in both measures of diversity (R' = 2.463, H_e = 0.207). In China the southern populations (R' = 3.227, H_e = 0.364) are more diverse then the northern red rice populations across all diversity measures, but based on allelic diversity, southern red rice populations are not significantly different from one another. For the US sub-regions, diversity as measured by H_e ranks Arkansas as the most diverse (0.655), followed by Texas (0.614) and Louisiana (0.555). However, when using allelic diversity as our measure, the order shifts so that Arkansas is still the most diverse (5.795), followed by Louisiana (4.937) and Texas (4.908).

Inbreeding as measured by F_{IS} is positive in all populations (Table 3.4). *O. sativa japonica* cultivars have the highest F_{IS} (0.936), followed by *O. sativa indica* cultivars (0.897), which we expect given that the cultivars are assumed to be 100% selfing. Northern Chinese red rice has the lowest inbreeding coefficient (0.347) relative to the other rice groups, which is interesting because this group is the least diverse by all metrics. This finding may suggest that outcrossing rates are higher in northern Chinese red rice populations, but that generally lower diversity in the region is driven by limited inter-specific gene flow with local cultivars and/or that the limited genetic diversity in the exclusively *O. sativa japonica* cultivars grown in the region have influenced diversity in these red rice populations. All US red rice regions are similar in their F_{IS} values (Tx = 0.896, La = 0.858 and Ar = 0.851). Red rice in southern China (F_{IS} = 0.848) is not significantly ($p < 0.05$) different from US red rice populations in all three areas, suggesting lower levels of outcrossing in these populations.

Relationships among red rice populations and rice taxa

Principal coordinate analysis (PCA) was first performed on the two regions independently in order to establish a baseline relationship with the red rice populations from both regions (US and China) and the cultivated and wild taxa (Figure 3.4). Both PCA analyses explained over 50% of the genetic variation in the first two axes. Chinese populations form two distinct, cohesive groups; one comprised of the northern Liaoning population and the other comprised of the southern populations (Jiangsu,

Guangdong and Hainan). Surprisingly, neither of the two sub-groups overlaps either wild *O. rufipogon* or cultivated *O. sativa* samples. US red rice populations show a different pattern, forming two diffuse sub-groups that, although distinct from one another, both overlap portions of the diversity seen in *O. rufipogon* and *O. sativa*.

Principal coordinate analysis was also performed on the entire data set, including all US and Chinese red rice populations, *O. sativa indica* and *O. sativa japonica* cultivars and *O. rufipogon* individuals using all 16 microsatellite loci (Figure 3.5). The PCA analysis explained 67.83% of the variation between the rice groups within the first three axes. The first two dimensions accounted for 53.2% of the variation while the third axes explained 14.63% of the additional variation between groups. Strikingly there is no overlap between and of the Chinese red rice individuals and the US red rice individuals, as would be expected if the Chinese populations were the source populations. In fact, the Chinese and US populations could not be more different from one another, defining the opposite ends of the primary axes of the PCA.

Analysis of molecular variance (AMOVA) results confirm the PCA results (Table 3.5) by showing that much of the variation seen in these global red rice samples is partitioned at the regional level (between the US and China). Sixty percent of the variation occurs among populations, 47% held among regions and 13% held among populations within the same region. Within population variation accounted for the remaining 40% of the genetic variation seen in the populations.

The neighbor-joining dendrogram suggests the genetic simmilarity seen in the Chinese red rice populations, placing them in a weakly supported clade (42.8% bootstrap support) (Figure 3.6). *O. sativa indica* is sister to the Chinese population clade (41.8% bootstrap support). US red rice populations fall into three groups, one of which contains *O. sativa japonica* which is placed sister to the Tx13 population with moderate bootstrap support, 57.7%.. Weak bootstrap support separates the remaining US red rice populations into two groups. *O. rufipogon* is nested within the clade along with La 12, Tx 10 and ArRR;

while the Chinese red rice/*indica* clade is sister to the other US red rice clade comprised to La11, MsRR, Ar3, 1, 5, 8, 9. Again we see a similar pattern of cohesiveness among the Chinese red rice populations as compared to the more differentiated US red rice populations. It is also interesting to note that the US populations show limited regionalism within the US, as we do not see three US clades each comprised of populations from one US area (i.e. separate Texas, Louisiana, and Arkansas clades); instead, populations from each US area are found across all three sub-groups. There is no convincing support for de-domestication in either region. There is generally very weak bootstrap support across much of the tree, suggesting a complex evolutionary relationship between the taxa.

InStruct results

InStruct analysis confirms the clear differentiation between Chinese and US red rice presented in the PCA analysis. Analysis performed exclusively on all red rice from both the US and China (using no population of origin information and assuming no Admixture) predicted and optimal K, or population number, of 16 (Figure 3.7). Chinese and US red rice populations share no genetic overlap with one another relative to the 16 genetic clusters assigned using InStruct.

The four Chinese red rice populations were placed in four corresponding genetic clusters (groups 4. 6, 12 and 14), with each population comprising the majority portion of one of the four groups. The Liaoning population is the most distinct Chinese population, having 100% assignment to group 6, which was not associated with any of the southern Chinese populations. Although all 4 populations are very distinct from one another, a small amount of overlap in population identity assignments occurred among the southern Chinese red rice populations suggesting connectivity. Individuals from Guangdong are assigned primarily to genetic cluster 4 (98%) but also have a portion of their population identity in genetic cluster 14, the same genetic cluster to which a majority of Hainan's population is assigned. Hainan also shares a small amount of its population assignment with genetic cluster 12, the genetic group to which Jiangsu is exclusively assigned. The shared genetic variation between Guangdong and Hainan, and

Hainan and Jiangsu; suggests that there is some amount of gene flow occurring between red rice populations in southern China.

Unlike Chinese populations, red rice from the US all share partial assignments in multiple genetic clusters. US populations share diversity across most of the other 12 genetic clusters they are associated with genetically. Arkansas, Texas and Louisiana all share a portion of their population assignment in 4 genetic clusters (3, 7, 8 and 10). Texas is associated with the fewest population clusters as compared to the other US regions. Arkansas is the most subdivided region with regards to population assignment, having partial assignments in 11 of the 12 groups. Of all the genetic groups, Arkansas individuals are associated with largest number of genetic clusters, 5 of which are exclusively associated with red rice individuals from Arkansas. As identified by K genetic clusters, there are more genetic subdivisions shared among regions in the US then there are between Chinese red rice regions, suggesting higher levels of genetic exchange among US regions.

InStruct analysis on the entire data set, including; all red rice populations from the US and China, *O. rufipogon* accessions and all *O. sativa* cultivars subdivided the data into K= 19 populations (Figure 3.8). Not unexpectedly, *O. rufipogon* individuals were assigned to the largest number of populations, or genetic clusters, having individuals entirely or fractionally placed in 14 of the 19 groups. Of the 14 genetic clusters only one group was exclusive to *O. rufipogon* (cluster 3). Associations in genetic variation between the cultivars and *O. rufipogon* is apparent, with 8 genetic clusters (2, 4, 5, 6, 11, 13, 14, 15) comprised of only *O. sativa japonica* and *O. rufipogon* individuals.

There is a clear genetic link between US red rice and *O. rufipogon*. Of the 8 genetic groups that US red rice populations are associated with 2 of them are held in common with all three US regions and *O. rufipogon* (Figure 3.8). Virtually no population membership was held in common between US and Chinese red rice, again emphasizing that lack of genetic similarity between the two regions. US red rice did show associations with both *indica* and *japonica* cultivars. Arkansas populations were strongly linked

with both *O. sativa* sub-species, by associations with both genetic cluster 17, accounting for roughly 89% of *indica* cultivars, and genetic cluster 8, accounting for roughly 88% of the *japonica* cultivars assessed. Interestingly, one genetic cluster (16) was unique to US red rice, comprised of individuals from Louisiana and Arkansas.

Chinese red rice unlike US populations, were each assigned to one genetic cluster. Interestingly we do not see the same population associations between Chinese red rice and *O. rufipogon* that we see in the US populations. This is contrary to expectation given that unlike US red rice, Chinese red rice populations are found in the native range of *O. rufipogon*. Liaoning red rice still remains distinct from the other southern populations, with 100% population association with genetic cluster 10, which also accounts for a portion (7%) of the population assignment across all *O. sativa japonica* cultivars. Jiangsu red rice is placed in a unique cluster (1), sharing very limited (0.01%) genetic association with individuals from Hainan. Populations from Guangdong and Hainan were grouped in the same genetic cluster (18) in this analysis. This genetic cluster (18) accounts for a very small fraction of the genetic variation seen in Arkansas individuals (0.3%) and about 24% of *O. sativa indica* cultivars. The limited overlap between US and Chinese red rice in this genetic cluster is most likely the result of indica specific alleles shared with both red rice populations independently.

Private alleles

O. sativa indica vs. *O. sativa japonica*

Comparisons between the two cultivar sub-species, *O. sativa japonica* and *O. sativa indica* reveal private alleles held between the two groups (Table 3.6). *O. sativa indica* has 71 private alleles occurring in all 16 SSRs screened. *O. sativa japonica* has slightly fewer private alleles, with 61 alleles occurring in 14 of the 16 loci screened. We would expect cultivated alleles in red rice regardless of a hybrid or de-domestication evolutionary history, but the pattern of allele sharing should be different in the two scenarios of weed origin. In a de-domestication scenario we expect to see only one sub-species represented in the red rice populations. If red rice is found in an agricultural area growing *indica* cultivars

we would see only *indica* alleles in the weedy populations, and conversely if only *japonica* cultivars are grown in a particular area (as is the case in northern China and the US) we would expect to find only *japonica* specific alleles. This is not the case. Of these sub-species specific alleles we see the occurrence of both *O. sativa japonica* and *O. sativa indica* alleles in both regions, suggesting a complex evolutionary history involving both cultivated types in both regions. Of the 71 *indica* specific alleles, 17 were found only in US populations, 12 were only found in Chinese populations, 10 alleles were found in both regions, while 32 alleles were not found in either population. The *japonica* specific alleles show a similar distribution with 25 occurring in only US populations, 2 alleles found in Chinese populations only, 5 found in both regions and 29 occurring in neither region.

Cultivar vs. *O. rufipogon*

All microsatellite loci contained at least one cultivar and one *O. rufipogon* private allele. A total of 88 cultivar (*O. sativa japonica* and *indica* were grouped together in this analysis) and 160 *O. rufipogon* specific alleles were found between the two groups. Cultivar specific alleles were found in both US and Chinese red rice populations. Of the private alleles found in the cultivar (absent in *O. rufipogon*) 8 alleles (in 7 SSRs) were found in both US and Chinese red rice populations, 15 (in 11 SSRs) were found in US red rice populations only, and 13 (in 8 SSRs) were present in Chinese red rice populations exclusively. The remaining 52 cultivar alleles were absent from red rice individuals in both regions. Alleles private to *O. rufipogon* relative to the cultivars were found in both US and Chinese red rice, suggesting that both red rice regions have a similar hybrid (*O. rufipogon* x *O. sativa*) evolutionary history. Of the *O. rufipogon* private alleles found in red rice, a majority was found in the US populations, with 48 alleles from all 16 SSRs represented. Four alleles (from 2 SSRs) were found in both US and Chinese red rice, and 9 alleles (in 6 SSRs) were found exclusively in Chinese red rice.

When looking more closely at the cultivar specific private alleles (*O. sativa indica* and *O. sativa japonica*) we find that 15 alleles in 10 SSRs are private to *O. sativa indica* (relative to both *O. sativa japonica* and *O. rufipogon*). Two of these alleles (in Rm230 and Rm316) are found in US red rice, 9

alleles (in Rm13, Rm261, Rm251, Rm206, Rm109, Rm247, Rm258, Rm316 and Rm481) are found only in Chinese populations and 4 alleles (in Rm13, Rm316 and Rm481) are found in both US and Chinese red rice. Eleven *O. sativa japonica* private alleles (not present in either *O. rufipogon* of *O. sativa indica*) in 8 SSRs were present in the sample, with 7 of these alleles occurring only in US red rice (in Rm220, Rm241, Rm234, Rm247 and Rm481), 2 alleles occurring only in Chinese red rice (in Rm241 and Rm109) and 2 alleles occurring in both red rice regions (in Rm206 and Rm316).

DISCUSSION

Data do not support a Chinese origin of the US red rice populations

The data show a striking lack of genetic similarly between Chinese and US red rice populations, suggesting no link between the two regions. If US red rice populations, or a subset of founding propagules, originated in China, then we would expect some amount of genetic overlap between the two regions. Instead we find the opposite, with analyses such and PCA, and pairwise F_{ST} analyses indicating that the two groups are more different from one another than either is to the remainder of the rice groups investigated. Both US and Chinese red rice have *O. rufipogon* private alleles, but the US populations have many more high and low frequency alleles, an unexpected result given that there are no *O. rufipogon* populations in North America. The presence of so many *O. rufipogon* private alleles suggests that US red rice populations are the result of multiple hybrid introductions from several distinct geographic locations within *O. rufipogon's* native range. InStruct analysis also clearly shows that US populations do not cluster with any of the potential Chinese source populations, and because there is no indication of shared ancestry with any of the populations, we can only conclude that none of the populations sampled in this study are source populations for US red rice.

In addition to sharing virtually no genetic overlap, Chinese populations show lower diversity levels than US populations. This is unexpected given that Chinese rice cultivation occurs within the native range of *O. rufipogon*, the putative parental species generating weedy red rice biotypes. Despite the

lack of evidence placing a source population for US red rice in China, we may expect that the co-occurrence of *O. sativa* and *O. rufipogon* would provide a biological context in which red rice types would be repeatedly produced *in situ*, potentially leading to a more genetically diverse red rice population than we would expect outside the native range of the parental species.

The phenomenon of increased genetic variability in the introduced range relative to an organism's native range is seen in other organisms with a history of repeated introduction (Kolbe et al., 2004). These invasive and weedy populations are more genetically diverse then native populations due to the fact that introduced populations contain individuals from multiple geographic locations throughout the native range. The admixture of many genetic types from across the native range via multiple introductions can lead to the creation of an introduced population that is more diverse then any single native population (Chun et al., 2009; Lachmuth et al., 2010; Pairon et al., 2010). Upon introduction these variable propagules can further generate novel genetic variability via gene flow between groups or biotypes that would otherwise be geographically isolated (Brown and Stepien, 2010; Chun et al., 2009; Marrs et al., 2008). Species level comparisons corroborate this hypothesis, as US red rice populations are the most allelicly rich group, housing more total alleles than *O. rufipogon*, indicating that US red rice populations are potentially forming a genetic amalgamation incorporating *O. rufipogon* alleles from multiple geographic localities and numerous distinct cultivars. The inclusion of cultivar and *O. rufipogon* alleles from numerous disparate and genetically distinct areas, followed by continued gene flow with local US cultivars may explain the high levels of diversity in these populations.

Evolutionary consequences of high levels of genetic variation in populations of red rice in the US may explain their ability to persist despite strong selective pressure against them in the agricultural environment. All aspects of rice cultivation act to limit the success of these weedy plants, yet they persist, in the face of various control measures; such as: dry seeding, weeding, and herbicide usage. Gene flow with local cultivars would explain *japonica* specific alleles seen in US red rice individuals with predominantly *indica* x *O. rufipogon* genetic backgrounds. Introgression of local cultivar alleles would

increase diversity in these populations while allowing them to acquire traits suited to local agricultural fields. Effective weeds blend into their environment, escaping removal until they set seed. Crop mimics, or weeds that share most of their phenotypic characteristics with cultivated forms, have the ability to persist in agricultural fields much longer than overtly 'wild' phenotypes. The agricultural dynamic seen in the US; strong selection against weedy types, high levels of genetic diversity, and maintenance of important weed characteristics (such as shattering) have the potential to generate super-weeds, which are virtually impossible to detect in the agricultural system. Conversely, high levels of genetic diversity may increase the likelihood that weedy rice populations are able to persist in marginal areas around agricultural fields, eventually invading local habitats.

Genetic variation in Chinese red rice suggest multiple regional origins

Chinese red rice populations form a monophyletic group in our analysis, suggesting regional cohesiveness among these populations in a global context. Short divergence times, bottlenecks, un-sampled populations, and admixture between taxa can make it difficult to identify the true topologies of a population level tree. Simulations preformed by Estoup and Guilmond suggest that in situations with short divergence times, resulting in low levels of differentiation by drift and mutation, tree topologies will only be correctly reconstructed 40% of the time (Estoup and Guillemaud, 2010). Despite our limited ability to accurately reconstruct evolutionary relationships between these intertwined taxa it is clear that Chinese red rice populations are more similar to each other than they are to any other taxonomic group included in the study. The populations are most genetically differentiated from *O. sativa japonica* cultivars, challenging the hypothesis that these populations are collectively the product of an *O. sativa japonica* de-domestication event (Gross et al., 2009; Gross et al., 2010; Reagon et al., 2010). Variation exists between these populations, separating northern and southern red rice populations into two groups, with northern populations having many more *japonica*-like traits and southern populations being much more *indica*-like. This difference makes sense in light of the specific cultivars grown in each region, *japonica* in the north and *indica* in the south.

Data suggests that red rice populations in the northern rice growing area of Liaoning may be the product of de-domestication from *O. sativa japonica* cultivars grown in the local area (Cao et al., 2006). The evolutionary origin of the southern populations is less clear, but preliminary data suggests it is of hybrid origin generated from a cross between 'ancient' *O. sativa indica* cultivars and wild *O. rufipogon* (Jing et al., 2007). All data point to minimal genetic overlap between northern and southern populations, confirming independent evolutionary histories as hypothesized in the literature, but the presence of *O. rufipogon* private alleles in both regions and the occurrence of *indica* specific alleles in the north and *japonica* specific alleles in the southern area also suggests that the germplasms in both regions have been influenced by all three putative parental taxa. This data suggests that both areas have a more complex evolutionary history then previously assumed. It remains unclear however if northern populations gained *O. rufipogon* alleles due to an independent hybrid origin *in situ* or via gene flow from southern hybrid (*O. sativa indica* x *O. rufipogon*) red rice individuals. The lack of sympatry between cultivated fields and wild *O. rufipogon* populations in Liaoning suggests gene flow from southern red rice populations may have a role in moving these *O. rufipogon* alleles north.

In spite of shared genetic variation among the southern populations the data clearly show that these populations are genetically distinct from one another. We do not observe the same level of population admixture, or co-ancestry seen in the US red rice populations in this region of China. Instead there is very minimal genetic overlap between areas, indicative of local generation of red rice types. The high genetic similarity of the Chinese populations with *O. rufipogon*, the co-occurrence of these cultivated areas with wild rice populations, and the lack of genetic similarity between the populations collectively indicates that the weedy populations are most likely forming recurrently, and generate weedy populations that persist locally.

Diversity in red rice suggests recurrent evolution of weed populations

Patterns of diversity in global red rice suggest recurrent evolution of the weed. Genetic variation across populations of red rice indicate that two different mechanisms, de-domestication and hybridization, contribute to the formation of weedy populations in different areas of the world. The maintenance of *O. rufipogon* alleles across disparate red rice populations suggests that the integration of portions of the wild rice genome may increase the fitness and/or weediness of these populations. Similarly, the integration and maintenance of cultivar alleles, in particular the assimilation of alleles from local rice cultivars may contribute to local adaptation and persistence of the weeds in the agricultural environment. The amalgamation of cultivated and wild rice genomes seen in red rice populations presents an interesting natural experiment in which further investigation into the exact proportions and regions of cultivated and wild genetic contributions will give us insight into the genetic architecture of weediness.

CONCLUSION

China is unlikely to be the source of US red rice populations. Instead, the data presented here suggest that US red rice populations are the result of multiple hybrid introductions (*O. sativa* X *O. rufipogon*) from Southeast Asia, from populations residing outside of mainland China. The next steps in uncovering the geographic origin of US populations should include a more extensive sampling of *O. rufipogon* throughout its entire range, in particular the regions of Papua New Guinea and east India, as *O. rufipogon* alleles found in these regions make up a large percentage of the *O. rufipogon* private alleles found in US red rice. Understanding patterns of diversity seen though the native range of *O. rufipogon* may help in uncovering the geographic origin(s) of US red rice. Similarly, screening global cultivar samples, historic or older land races, in particular may also offer a productive avenue of research, because using SNPs or other cultivar specific markers could be used to trace potential cultivar parents of weedy populations.

Weedy red rice populations located across the globe may have originated in different ways, via different mechanisms, with some populations arising as the result of hybridization, some as the result of de-domestication, while others are likely the product of a combination of mechanisms. The age of populations, amount of gene flow from other regions and the incidence of backcrossing and continual genetic exchange with new/modern cultivars will affect populations of red rice around the globe to varying degrees.

REFERENCES

Bacigalupe L. (2009) Biological invasions and phenotypic evolution: a quantitative genetic perspective. Biological Invasions:2243-2250.

Barbier P., Morishima H., Ishihama A. (1991) Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. TAG Theoretical and Applied Genetics 81:693-702.

Brown J., Stepien C. (2010) Population genetic history of the dreissenid mussel invasions: expansion patterns across North America. Biological Invasions:3687-3710.

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

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67.

Chun Y.J., Nason J.D., Moloney K.A. (2009) Comparison of quantitative and molecular genetic variation of native vs. invasive populations of purple loosestrife (*Lythrum salicaria* L., Lythraceae). Molecular Ecology 18:3020-3035.

Dlugosch K., Hays C. (2008) Genotypes on the move: some things old and some things new shape the genetics of colonization during species invasions. Molecular Ecology:4583-4585.

Ellstrand N. (2009a) Evolution of invasiveness in plants following hybridization. Biological Invasions:1089-1091.

Ellstrand N., Schierenbeck K. (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? Euphytica:35-46.

Ellstrand N.C. (2009b) Evolution of invasiveness in plants following hybridization. pp. 1089-1091.

Ellstrand N.C., Prentice H.C., Hancock J.F. (1999) Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 30:539-563.

Estoup A., Guillemaud T. (2010) Reconstructing routes of invasion using genetic data: why, how and so what? Molecular Ecology.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611-2620.

Facon B., Pointier J.P., Jarne P., Sarda V., David P. (2008) High genetic variance in life-history strategies within invasive populations by way of multiple introductions. Current Biology 18:363-367.

Falush D., Stephens M., Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567-1587.

Felsenstein J. (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? Molecular Biology and Evolution 23:691-700.

Felsenstein J. (2007) PHYLIP (Phylogeny Inference Package) version 3.67., Department of Genome Sciences, University of Washington, Seattle., Distributed by the author.

Gao H., Williamson S., Bustamante C.D. (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635-1651.

Gao L. (2004) Population structure and conservation genetics of wild rice *Oryza rufipogon* (Poaceae): a region wide perspective from microsatellite variation. Molecular Ecology 13:1009-1024.

Gao L., Chen W., Jiang W., Ge S., Hong D., Wang X. (2000) Genetic Erosion in Northern Marginal Population of the Common Wild Rice *Oryza Rufipogon* Griff. and its Conservation, Revealed by the Change of Population Genetic cstructure. Hereditas 133:47-53.

Gao L.Z., Zhang C.H. (2005) Comparisons of microsatellite variability and population genetic structure of two endangered wild rice species, *Oryza rufipogon* and *O. officinalis*, and their conservation implications. Biodiversity and Conservation 14:1663-1679.

Gealy D.R., Bryant R.J. (2009) Seed physicochemical characteristics of field-grown US weedy red rice (*Oryza sativa*) biotypes: Contrasts with commercial cultivars. Journal of Cereal Science 49:239-245.

Gealy D.R., Tai T.H., Sneller C.H. (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Science 50:333-339.

Goudet J. (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of Heredity 86:485.

Gross B.L., Skare K.J., Olsen K.M. (2009) Novel Phr1 mutations and the evolution of phenol reaction variation in US weedy rice (*Oryza sativa*), New Phytol, England. pp. 842-50.

Gross B.L., Reagon M., Hsu S.C., Caicedo A.L., Jia Y., Olsen K.M. (2010) Seeing red: the origin of grain pigmentation in US weedy rice, Mol Ecol, England. pp. 3380-93.

Jing W., Zhang W., Jiang L., Chen L., Zhai H., Wan J. (2007) Two novel loci for pollen sterility in hybrids between the weedy strain Ludao and the Japonica variety Akihikari of rice (*Oryza sativa* L.). TAG Theoretical and Applied Genetics 114:915-925.

Kolbe J., Glor R., Schettino L., Lara A., Larson A., Losos J. (2004) Genetic variation increases during biological invasion by a Cuban lizard. Nature:177-181.

Kolbe J.J., Glor R.E., Schettino L.R., Lara A.D.A.C., Larson A., Losos J.B. (2007) Multiple sources, admixture, and genetic variation in introduced Anolis lizard populations. Conservation Biology 21:1612-1625.

Lachmuth S., Durka W., Schurr F.M. (2010) The making of a rapid plant invader: genetic diversity and differentiation in the native and invaded range of *Senecio inaequidens*. Mol Ecol 19:3952-67.

Lavergne S., Molofsky J. (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. Proceedings of the National Academy of Sciences 104:3883.

Leung B., Drake J.M., Lodge D.M. (2004) Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology 85:1651-1660.

Liu X., Lu T., Yu S., Li Y., Huang Y., Huang T., Zhang L., Zhu J., Zhao Q., Fan D. (2007) A collection of 10,096 *indica* rice full-length cDNAs reveals highly expressed sequence divergence between *Oryza sativa indica* and *japonica* subspecies. Plant Molecular Biology 65:403-415.

Londo J.P., Chiang Y.C., Hung K.H., Chiang T.Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. Proceedings of the National Academy of Sciences of the United States of America 103:9578-9583.

Marrs R., Sforza R., Hufbauer R. (2008) When invasion increases population genetic structure: a study with Centaurea diffusa. Biological Invasions:561-572.

Morishima H., Barbier P. (1990) Mating System and Genetic Structure of Natural Populations in Wild Rice *Oryza rufipogon*. Plant Species Biology 5:31-39.

Mousadik A., Petit R. (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. Molecular Ecology 5:547-555.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590.

Oka H.I., Morishima H. (1967) Variations in the breeding systems of a wild rice, *Oryza perennis*. Evolution 21:249-258.

Pairon M., Petitpierre B., Campbell M., Guisan A., Broennimann O., Baret P., Jacquemart A., Besnard G. (2010) Multiple introductions boosted genetic diversity in the invasive range of black cherry (*Prunus serotina*; Rosaceae). Annals of Botany:881-890.

Peakall R., Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Pritchard J., Falush D., Stephens M. (2002) Inference of population structure in recently admixed populations. pp. 67.

Reagon M., Thurber C.S., Gross B.L., Olsen K.M., Jia Y., Caicedo A.L. (2010) Genomic patterns of nucleotide diversity in divergent populations of U.S. weedy rice, BMC Evol Biol, England. pp. 180.

Rosenberg N.A. (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4:137-138.

Sax D.F., Stachowicz J.J., Gaines S.D. (2005) Species invasions: insights into ecology, evolution and biogeography Sinauer Associates Incorporated.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234.

Song Z., Xu X., Wang B., Chen J., Lu B.R. (2003a) Genetic diversity in the northernmost *Oryza rufipogon* populations estimated by SSR markers. TAG Theoretical and Applied Genetics 107:1492-1499.

Song Z., Zhu W., Rong J., Xu X., Chen J., Lu B.R. (2006) Evidences of introgression from cultivated rice to *Oryza rufipogon* (Poaceae) populations based on SSR fingerprinting: implications for wild rice differentiation and conservation. Evolutionary Ecology 20:501-522.

Song Z.P., Lu B.R., Zhu Y.G., Chen J.K. (2003b) Gene flow from cultivated rice to the wild species *Oryza rufipogon* under experimental field conditions. New Phytologist 157:657-665.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11:1441-1452.

Vaughan L.K., Ottis B.V., Prazak-Havey A.M., Bormans C.A., Sneller C., Chandler J.M., Park W.D. (2001) Is all red rice found in commercial rice really *Oryza sativa*? Weed Science 49:468-476.

Xie Z., Lu Y., Ge S., Hong D., Li F. (2001) Clonality in wild rice (*Oryza rufipogon*, Poaceae) and its implications for conservation management. American Journal of Botany 88:1058.

Xu X., LU B.A.O.R., Chen Y.H., Xu M., Rong J., Ye P., Chen J., Song Z. (2006) Inferring population history from fine scale spatial genetic analysis in *Oryza rufipogon* (Poaceae). Molecular Ecology 15:1535-1544.

Yu G., Bao Y., Shi C., Dong C., Ge S. (2005) Genetic diversity and population differentiation of Liaoning weedy rice detected by RAPD and SSR markers. Biochemical Genetics 43:261-270.

Zhou H., Xie Z., Ge S. (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff.) in China. TAG Theoretical and Applied Genetics 107:332-339.

Figure 3.1: Map of *O. rufipogon* native range

Figure 3.2: Map of US and Chinese red rice population locations

Figure 3.3: Comparison of allelic richness (R') and genetic diversity (H_e) for all US and Chinese red rice regions.

Figure 3.4: Panel A-Principal Coordinate Analysis of Chinese Red Rice populations, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals. In the Chinese analysis the two-dimensional PCA explained 61% of the variation present in the data (the third axis accounts for an additional 18%). Symbols correspond to red rice, *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* individuals.

Panel B-Principal Coordinate Analysis of US Red Rice populations, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals. In the US analysis the two-dimensional PCA explained 50% of the variation present in the data (the third axis accounts for an additional 17%). Symbols correspond to red rice, *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* individuals.

Figure 3.5: Principal Coordinate Analysis of all data; including: red rice, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals. Two-dimensional PCA explained 52% of the variation present in the data (the third axis accounts for 68%). Symbols correspond to the population of origin for all red rice, *O. rufipogon*, *O. sativa indica* and *O. sativa japonica* individuals.

Figure 3.6: Neighbor-joining dendrogram based on Nei's (1972) genetic distance values for all red rice populations and rice taxa constructed using population genotypic data from all 16 SSR loci. Numbers along the branches represent bootstrap support greater than 40% after 1000 replicates.

Figure 3.7: Distruct graphical representation of the InStruct results for both US and Chinese red rice populations. An optimal K = 16 was found across both regions. A table of percent association with each genetic cluster identified for each population and a visual representation of all individual assignments is presented.

Figure 3.8: Distruct graphical representation of the InStruct results for all data; including US and Chinese red rice populations, *O. sativa indica* and *japonica* and *O. rufipogon*. An optimal K = 16 was found across taxa. A table of percent association with each genetic cluster identified for each population and a visual representation of all individual assignments is presented.

Sample	IRRI #	Species	Status	Location of Origin
tx11	n/a	*O. sativa*	Weedy	N29° 33.8281 W95°.8040
tx10	n/a	*O. sativa*	Weedy	N26°.127006 W98°23.198
tx8	n/a	*O. sativa*	Weedy	N29°33.750 W96°31.421
tx13	n/a	*O. sativa*	Weedy	N29°12.314 W96°20.997
tx3	n/a	*O. sativa*	Weedy	N29°11.859 W96°20.726
tx9b	n/a	*O. sativa*	Weedy	N29°11.859 W96°20.726
la5	n/a	*O. sativa*	Weedy	N32°38.898 W91°16.309
la6	n/a	*O. sativa*	Weedy	N31°49.195 W91°23.198
la7	n/a	*O. sativa*	Weedy	N31°33.750 W91°31.421
la11	n/a	*O. sativa*	Weedy	N30°12.314 W92°20.997
la12	n/a	*O. sativa*	Weedy	N28°11.859 W92°20.726
la13	n/a	*O. sativa*	Weedy	N26°10.716 W92°19.543
ar5	n/a	*O. sativa*	Weedy	N34°28.571 W91°26.61
ar6	n/a	*O. sativa*	Weedy	N34°28.580 W91°26.296
ar8	n/a	*O. sativa*	Weedy	N34°28.261 W91°24.225
ar9	n/a	*O. sativa*	Weedy	N34°28.652 W91°24.199
ar1	n/a	*O. sativa*	Weedy	N34°07.626 W91°15.991
ar2	n/a	*O. sativa*	Weedy	N34°05.011 W91°22.349
ar3	n/a	*O. sativa*	Weedy	N34°14.339 W91°17.681
Orufi24-3	81984	*O. rufipogon*	wild	Laos
Orufi28-3	81997	*O. rufipogon*	wild	Papua New Guinea
Orufi30-3	100189	*O. rufipogon*	wild	Malaysia
Orufi31-3	100211	*O. rufipogon*	wild	India
Orufi32-3	100588	*O. rufipogon*	wild	Taiwan
Orufi38-3	103308	*O. rufipogon*	wild	Taiwan
Orufi40-3	103423	*O. rufipogon*	wild	Sri Lanka
Orufi45-3	104624	*O. rufipogon*	wild	China
Orufi48-3	104714	*O. rufipogon*	wild	Thailand
Orufi53-3	104804	*O. rufipogon*	wild	Thailand
Orufi56-3	104516	*O. rufipogon*	wild	Thailand
Orufi57-3	104618	*O. rufipogon*	wild	Thailand
Orufi59-4	104833	*O. rufipogon*	wild	Thailand
Orufi59-5	104833	*O. rufipogon*	wild	Thailand
Orufi60-3	104857	*O. rufipogon*	wild	Thailand
Orufi62-2	104870	*O. rufipogon*	wild	Thailand
Orufi62-3	104870	*O. rufipogon*	wild	Thailand
Orufi62-5	104870	*O. rufipogon*	wild	Thailand
Orufi63-2	104871	*O. rufipogon*	wild	Thailand
Orufi63-3	104871	*O. rufipogon*	wild	Thailand
Orufi63-4	104871	*O. rufipogon*	wild	Thailand
Orufi64-2	104875	*O. rufipogon*	wild	Thailand

Orufi69-1		105491	*O. rufipogon*	wild	Malaysia
Orufi74-2		105711	*O. rufipogon*	wild	India
Orufi77-4		105843	*O. rufipogon*	wild	Thailand
Orufi77-5		105843	*O. rufipogon*	wild	Thailand
Orufi78-5		105855	*O. rufipogon*	wild	Thailand
Orufi81-2		105942	*O. rufipogon*	wild	Thailand
Orufi81-3		105942	*O. rufipogon*	wild	Thailand
Orufi83-3		106036	*O. rufipogon*	wild	Malaysia
Orufi86-4		106086	*O. rufipogon*	wild	India
Orufi91-4		106134	*O. rufipogon*	wild	India
Orufi93-3		106166	*O. rufipogon*	wild	Vietnam
Orufi101-1		106321	*O. rufipogon*	wild	Cambodia
Orufi103-1		106346	*O. rufipogon*	wild	Mynamar
Onivara-f	n/a		*O. rufipogon*	wild	China
Onivara-I	n/a		*O. rufipogon*	wild	China
tx2-st	n/a		*O.sativa japonica*	Cultivated	United States
tx3-st	n/a		*O.sativa japonica*	Cultivated	United States
tx5-st	n/a		*O.sativa japonica*	Cultivated	United States
tx6-st	n/a		*O.sativa japonica*	Cultivated	United States
tx7-st	n/a		*O.sativa japonica*	Cultivated	United States
tx8-st	n/a		*O.sativa japonica*	Cultivated	United States
tx9-st	n/a		*O.sativa japonica*	Cultivated	United States
tx10-st	n/a		*O.sativa japonica*	Cultivated	United States
tx11-st	n/a		*O.sativa japonica*	Cultivated	United States
tx12-st	n/a		*O.sativa japonica*	Cultivated	United States
tx13-st	n/a		*O.sativa japonica*	Cultivated	United States
la3-st	n/a		*O.sativa japonica*	Cultivated	United States
la4-st	n/a		*O.sativa japonica*	Cultivated	United States
la5-st	n/a		*O.sativa japonica*	Cultivated	United States
la6-st	n/a		*O.sativa japonica*	Cultivated	United States
la7-st	n/a		*O.sativa japonica*	Cultivated	United States
la9-st	n/a		*O.sativa japonica*	Cultivated	United States
la12-st	n/a		*O.sativa japonica*	Cultivated	United States
ar2-st	n/a		*O.sativa japonica*	Cultivated	United States
ar4-st	n/a		*O.sativa japonica*	Cultivated	United States
ar5-st	n/a		*O.sativa japonica*	Cultivated	United States
ar6-st	n/a		*O.sativa japonica*	Cultivated	United States
ar7-st	n/a		*O.sativa japonica*	Cultivated	United States
ar8-st	n/a		*O.sativa japonica*	Cultivated	United States
ar9-st	n/a		*O.sativa japonica*	Cultivated	United States
Cl111	n/a		*O.sativa japonica*	Cultivated	United States
CL161	n/a		*O.sativa japonica*	Cultivated	United States

Cl181	n/a	*O.sativa japonica*	Cultivated	United States
CarolinaGold	n/a	*O.sativa japonica*	Cultivated	United States
Labelle	n/a	*O.sativa japonica*	Cultivated	Asian
Lemont	n/a	*O.sativa japonica*	Cultivated	Asian
Calrose	n/a	*O.sativa japonica*	Cultivated	Asian
nipponbare	n/a	*O.sativa japonica*	Cultivated	Asian
gimbozo	n/a	*O.sativa japonica*	Cultivated	Asian
EarlyWantanabune	n/a	*O.sativa japonica*	Cultivated	Asian
Rexoro	n/a	*O.sativa japonica*	Cultivated	Asian
Kitake	n/a	*O.sativa japonica*	Cultivated	Asian
Mzoz	n/a	*O.sativa japonica*	Cultivated	Asian
morobrekan	n/a	*O.sativa japonica*	Cultivated	Asian
azucena	n/a	*O.sativa japonica*	Cultivated	Asian
Wab58	n/a	*O.sativa japonica*	Cultivated	Asian
XiNan175	n/a	*O.sativa japonica*	Cultivated	Asian
TaiNong67	n/a	*O.sativa japonica*	Cultivated	Asian
K59	n/a	*O.sativa japonica*	Cultivated	Asian
NongLin8	n/a	*O.sativa japonica*	Cultivated	Asian
co39	n/a	*O.sativa indica*	Cultivated	Asian
Tetep	n/a	*O.sativa indica*	Cultivated	Asian
Ir46-a	n/a	*O.sativa indica*	Cultivated	Asian
IRRB-21	n/a	*O.sativa indica*	Cultivated	Asian
IR-72	n/a	*O.sativa indica*	Cultivated	Asian
C64	n/a	*O.sativa indica*	Cultivated	Asian
MiYang46	n/a	*O.sativa indica*	Cultivated	Asian
Q2436	n/a	*O.sativa indica*	Cultivated	Asian
FuHui838	n/a	*O.sativa indica*	Cultivated	Asian
MianHui501	n/a	*O.sativa indica*	Cultivated	Asian
MianHui725	n/a	*O.sativa indica*	Cultivated	Asian
MingHui63	n/a	*O.sativa indica*	Cultivated	Asian
MIngHui70	n/a	*O.sativa indica*	Cultivated	Asian
MingHui77	n/a	*O.sativa indica*	Cultivated	Asian
MIngHui86	n/a	*O.sativa indica*	Cultivated	Asian
ChengHui047	n/a	*O.sativa indica*	Cultivated	Asian
ChengHui448	n/a	*O.sativa indica*	Cultivated	Asian
EnHui58	n/a	*O.sativa indica*	Cultivated	Asian
ZhenHui084	n/a	*O.sativa indica*	Cultivated	Asian
YanHui559	n/a	*O.sativa indica*	Cultivated	Asian
ShuHui527	n/a	*O.sativa indica*	Cultivated	Asian
Ce64-7	n/a	*O.sativa indica*	Cultivated	Asian
DuoXi1	n/a	*O.sativa indica*	Cultivated	Asian
DiJiaoWuJian	n/a	*O.sativa indica*	Cultivated	Asian

R458	n/a	*O.sativa indica*	Cultivated	Asian
R818	n/a	*O.sativa indica*	Cultivated	Asian
CDR22	n/a	*O.sativa indica*	Cultivated	Asian
ShengLiXian	n/a	*O.sativa indica*	Cultivated	Asian
93-11	n/a	*O.sativa indica*	Cultivated	Asian
GuMei2	n/a	*O.sativa indica*	Cultivated	Asian
Pa64	n/a	*O.sativa indica*	Cultivated	Asian
AiZiZhan	n/a	*O.sativa indica*	Cultivated	Asian
IR24	n/a	*O.sativa indica*	Cultivated	Asian
IR26	n/a	*O.sativa indica*	Cultivated	Asian
IR30	n/a	*O.sativa indica*	Cultivated	Asian
IR36	n/a	*O.sativa indica*	Cultivated	Asian
Tadukan	n/a	*O.sativa indica*	Cultivated	Asian
Liaoning	WRLN001	*O. sativa*	Weedy	41°44'N, 123°14'E
Jiangsu	WRJS035	*O. sativa*	Weedy	32°01'N, 120°16'E
Guangdong	WRGS001	*O. sativa*	Weedy	20°53'N, 110°05'E
Hainan	WRHA020	*O. sativa*	Weedy	19°48'N, 110°31'E

Table 3.1: Sample Information. Name, identification number, location of collection, taxon affiliation, and improvement status.

Taxonomic group	N	P_p(%)	A_p	AP_p	A_e	H_e	H_o
All red rice	410	100%	338	21.13	3.998	0.726	0.086
US red rice	322	100%	285	17.8	3.147	0.658*	0.084
Chinese red rice	88	81.25%	96	7.39	3.303	0.477*	0.08
O.sativa japonica	45	100%	96	6	2.353	0.453	0.03
O.sativa indica	38	100%	106	6.63	3.25	0.64	0.069
O. rufipogon	37	100%	239	14.94	8.498	0.858	0.221

Table 3.2: Species level genetic diversity comparisons. N = sample size; P_p = percent polymorphic loci; A_p = total number of alleles per taxon (including monomorphic loci); AP_p = mean number of alleles per polymorphic locus; A_e = mean effective number of alleles per polymorphic locus; H_e = genetic diversity (expected heterozygosity); H_o = observed heterozygosity. *significantly different ($p < 0.0001$)

Pairwise F_{ST} (via Frequency) for all Rice Groups

	US red rice	Chinese red rice	*O.rufipogon*	*O. sativa japonica*	*O. sativa indica*
US red rice	0.000				
Chinese red rice	0.254	0.000			
O. rufipogon	0.065	0.183	0.000		
O. sativa japonica	0.203	0.297	0.167	0.000	
O. sativa indica	0.150	0.234	0.107	0.272	0.000

Table 3.3: Pairwise F_{ST} values between rice groups. Low F_{ST} values indicate a high level of similarity between US red rice and *O. rufipogon*. High F_{ST} values between US and Chinese red rice suggest that these groups are highly differentiated from one another.

Summary AMOVA Table

Source	df	Est. Var.	%
Among Regions	1	51952.724	47%
Among Pops/Regions	21	14099.667	13%
Within Pops	741	44789.679	40%
Total	763	110842.071	100%

Table 3.4: AMOVA: US and Chinese red rice populations

Analysis of molecular variance shows that most variation in red rice is present among regions (47%). Within population variation accounts for a large portion of the genetic variation (40%), while among population variation within a region accounts for the least amount of variation (13%).

Region	Ho	He	R'	Fis
Tx	0.063	0.606*	4.908	0.897
La	0.079	0.550	4.937	0.858
Ar	0.098	0.647*	5.795	0.851
N. China	0.135	0.201*^	2.463	0.347
S. China	0.055	0.359*^	3.227	0.848
O. sativa japonica	0.030	0.453	4.978	0.936
O. sativa indica	0.069	0.640	5.752	0.896
O. rufipogon	0.221	0.858	11.569	0.750

Table 3.5: Rarefaction generated measures of diversity

Analysis of regional diversity as measured by H_o (observed heterozygosity), H_e (expected heterozygosity), F_{IS} (inbreeding coefficient) and R' (allelic richness) show variation in diversity across the regions. *O. rufipogon* is the most diverse group, while Chinese red rice groups are the least using the same criteria. ^ no significant difference *significant ($p < 0.05$)

Private Allele Summary

Cultivar (*indica* and *japonica*) vs. *O. rufipogon*					
	Cultivar specific	US red rice	Chinese red rice	US and China	Neither
Number loci with private alleles	16	11	8	7	0
Percent SSR with private alleles	100%	68.75%	50%	43.75%	0%
Total number private alleles	88	15	13	8	52
	O. rufipogon specific	US red rice	Chinese red rice	US and China	Neither
Number loci with private alleles	16	16	6	2	0
Percent SSR with private alleles	100%	100%	37.75%	12.50%	0%
Total number private alleles	160	48	9	4	99

O. sativa indica* vs. *O.sativa japonica					
	O. sativa indica specific	US red rice	Chinese red rice	US and China	Neither
Number loci with private alleles	16	10	8	7	1
Percent SSR with private alleles	100%	62.50%	50%	43.75%	6.25%
Total number private alleles	71	17	12	10	32
	O. sativa japonica specific	US red rice	Chinese red rice	US and China	Neither
Number loci with private alleles	14	11	2	5	3
Percent SSR with private alleles	87.50%	68.75%	12.50%	31.25%	18.75%
Total number private alleles	61	25	2	5	29

Table 3.6: Private Alleles between rice groups

Analysis of private allele sharing between groups reveals that both US and Chinese red rice have *O. sativa indica, O. sativa japonica* and *O. rufipogon* private alleles. US red rice populations surprisingly contain more *O. rufipogon* alleles then Chinese populations.

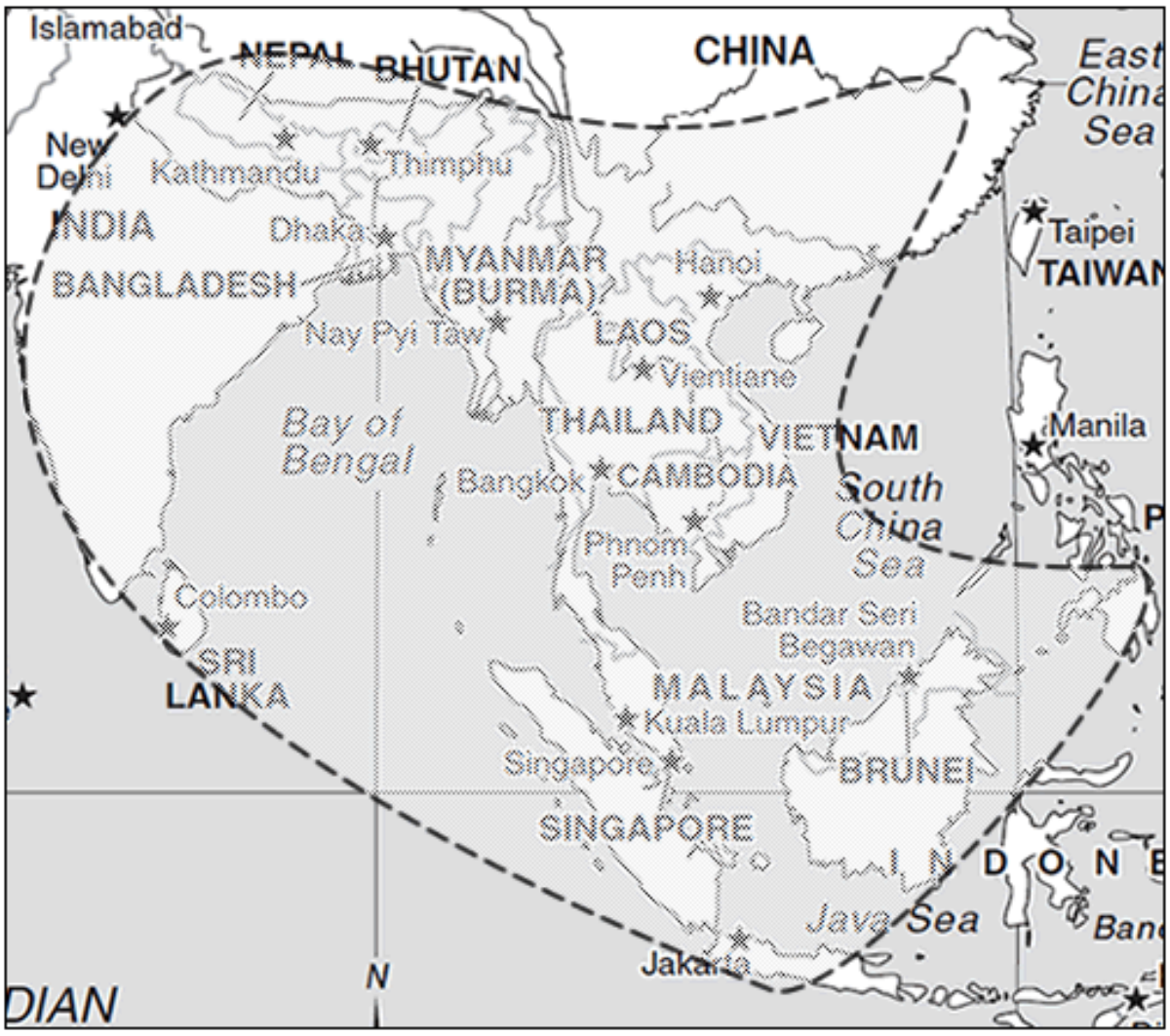


Figure 3.1: Map of *Oryza rufipogon's* native range.

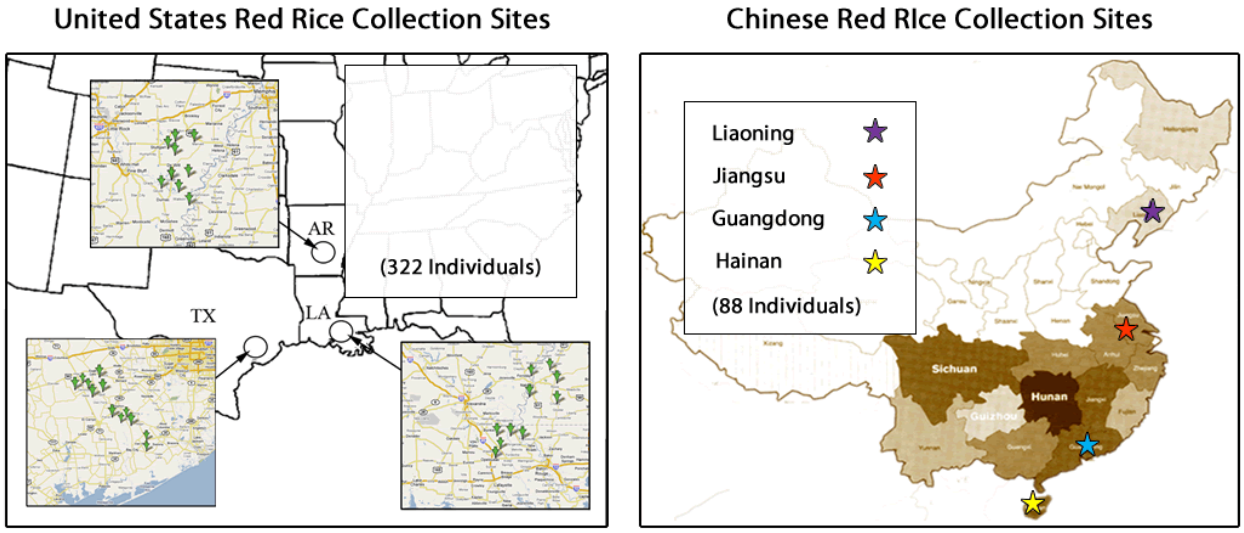


Figure 3.2: Map of collection sites in China and the US

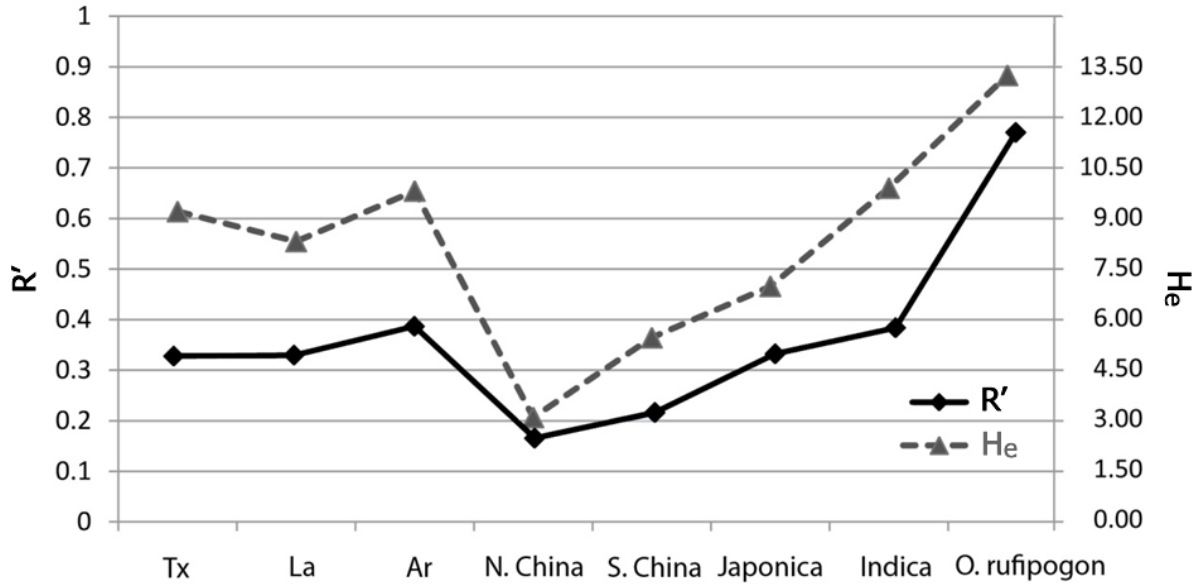
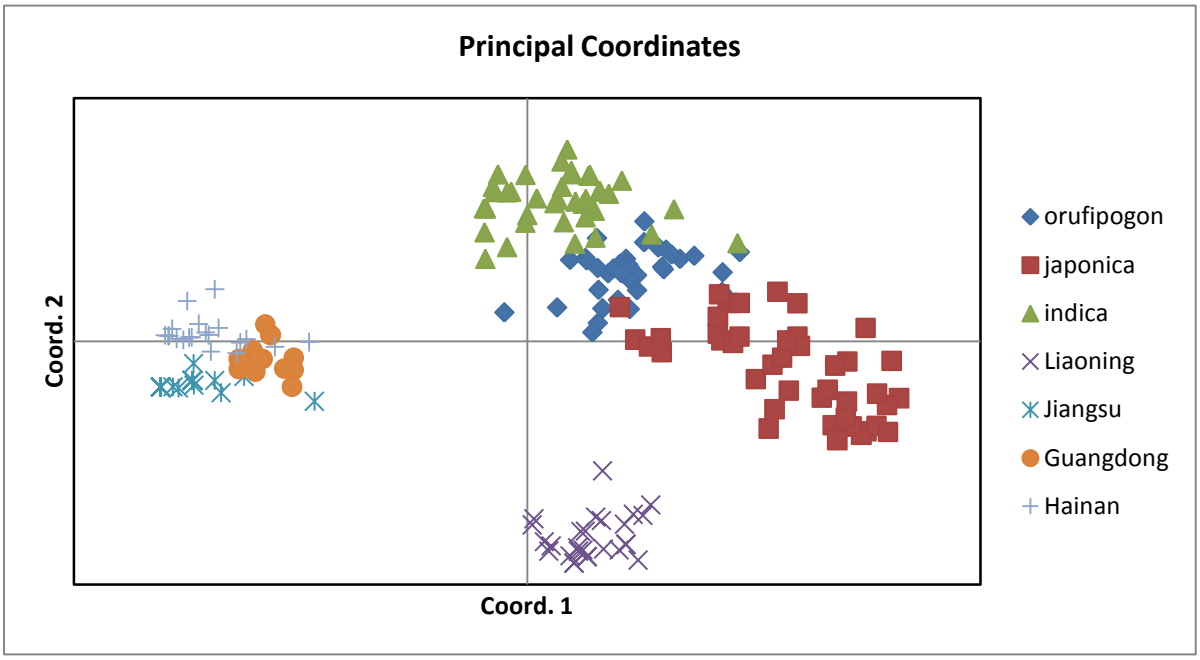


Figure 3.3 Comparison of genetic diversity using allelic diversity (R') and expected heterozygosity (H_e)



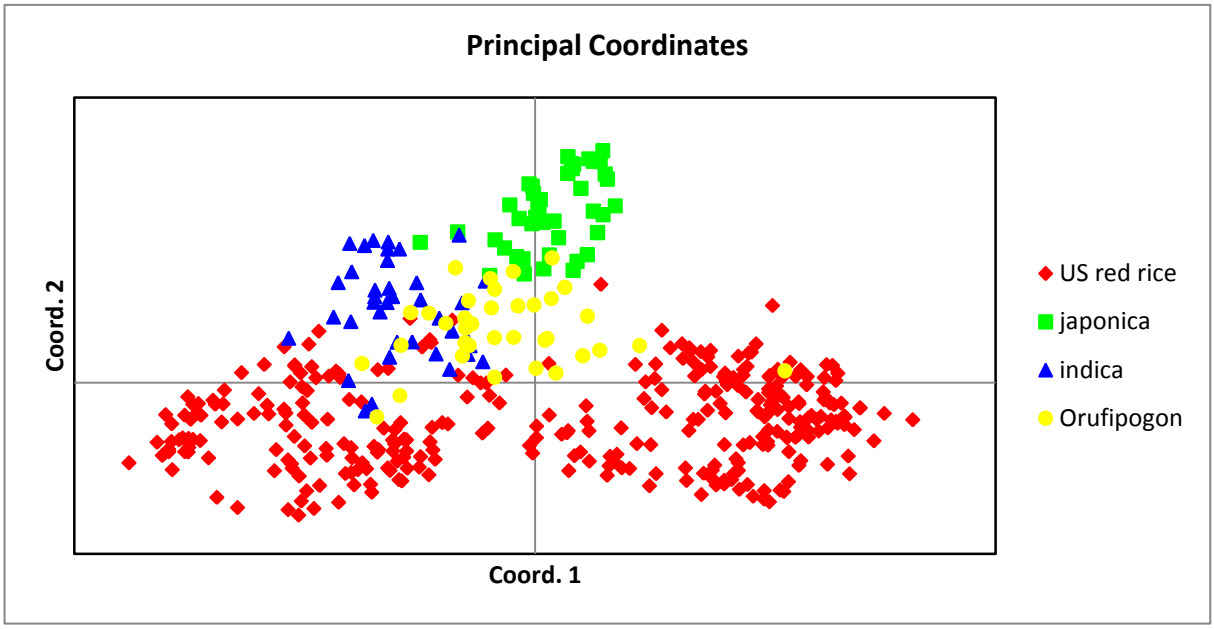


Figure 3.4: Panel a- PCA Chinese red rice populations, *O. rufipogon* and *O. sativa indica* and *japonica* cultivars. Panel b- PCA of US red rice populations, *O. rufipogon* and *O. sativa indica* and *japonica* cultivars.

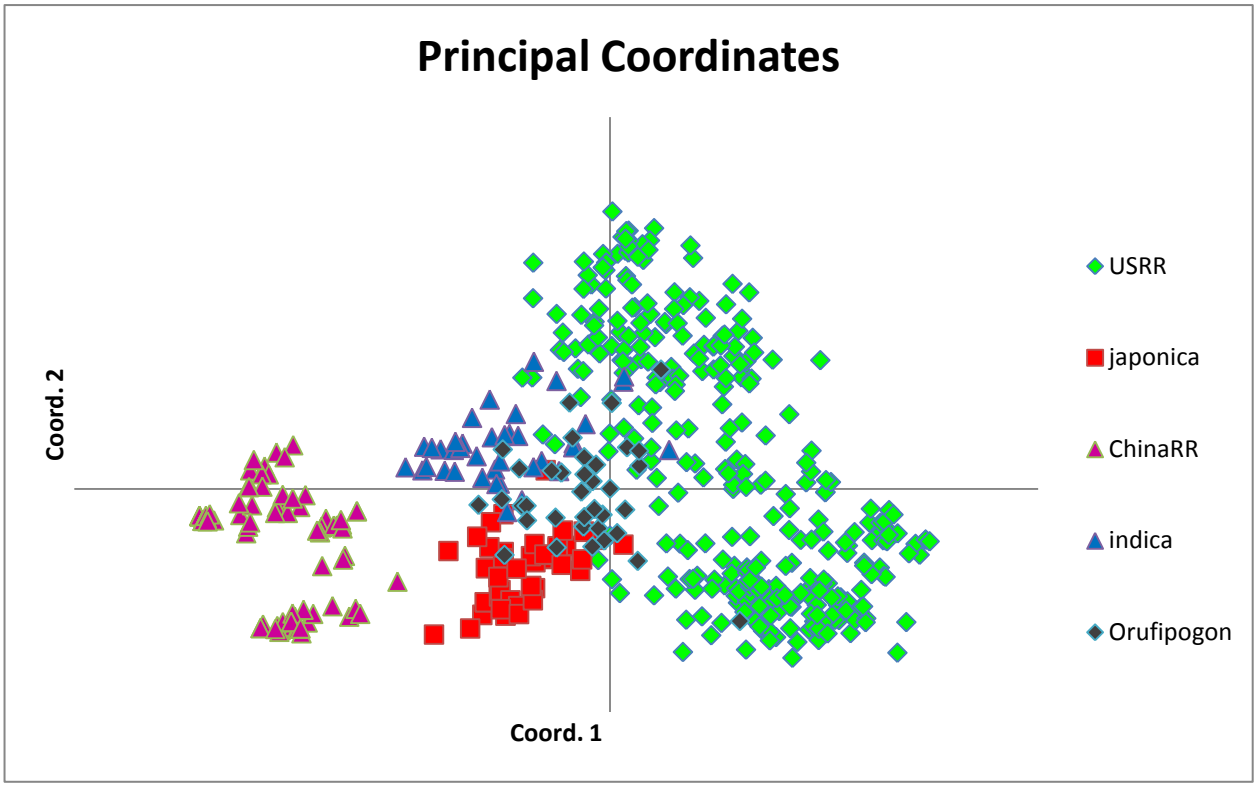


Figure 3.5: Principal Coordinate Analysis of US red rice, Chinese red rice, *O. rufipogon*, *O. sativa japonica* and *O. sativa indica* individuals.

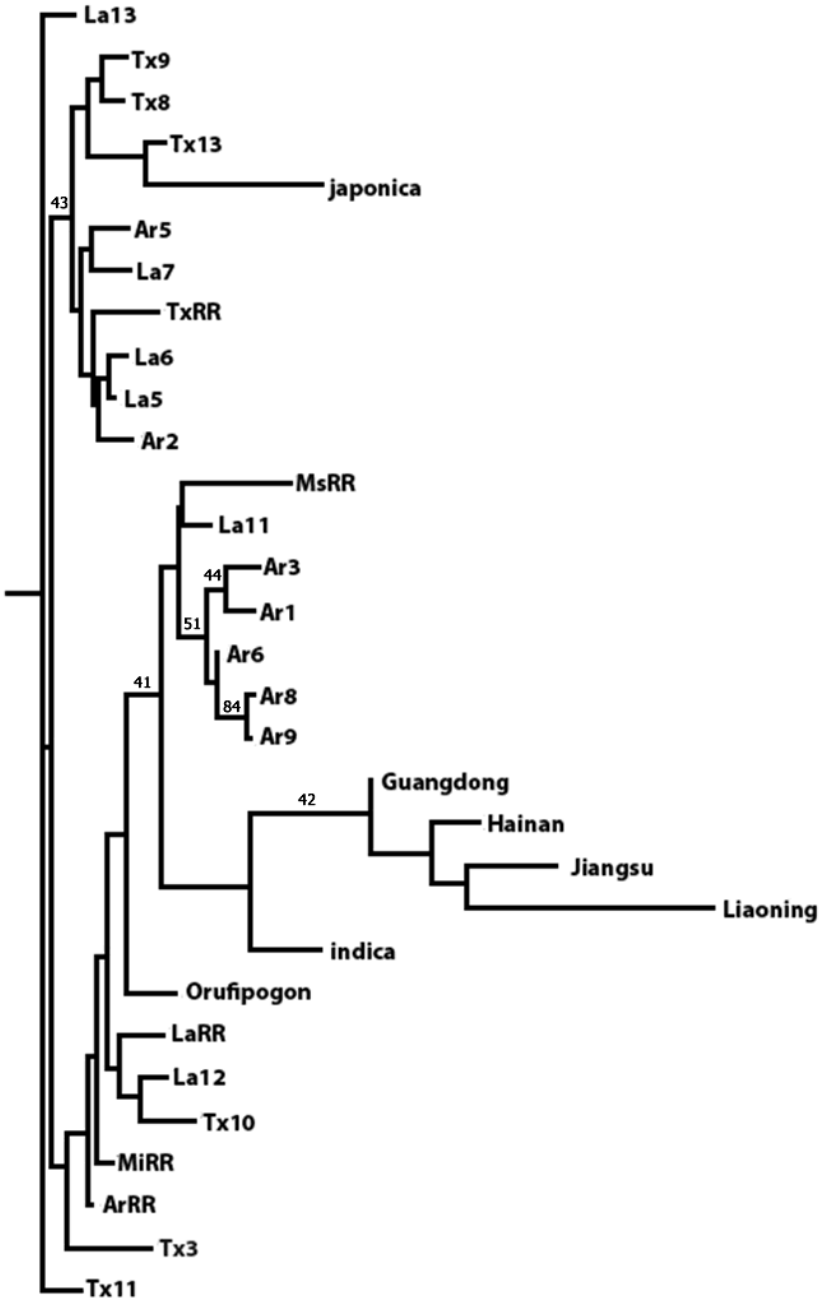


Figure 3.6: Unrooted Neighbor-joining dendogram of all rice groups, *O. sativa indica* and *japonica* cultivars, *O. rufipogon* accessions, US and Chinese red rice individuals.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11	Group 12	Group 13	Group 14	Group 15	Group 16
Tx:	0.0000	0.0000	0.0346	0.0000	0.0000	0.0000	0.4732	0.0025	0.0000	0.2516	0.0000	0.0000	0.0000	0.0000	0.0000	0.2154
La:	0.0000	0.3042	0.4016	0.0000	0.0000	0.0000	0.0885	0.1305	0.0000	0.0001	0.0000	0.0000	0.0490	0.0000	0.0000	0.0003
Ar:	0.0014	0.0177	0.0085	0.0000	0.0019	0.0000	0.0089	0.3127	0.0012	0.0020	0.0013	0.0000	0.4481	0.0000	0.0014	0.0000
Liao:	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jiang:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Guang:	0.0000	0.0000	0.0000	0.9840	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0160	0.0000	0.0000
Hainan:	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0000	0.9983	0.0000	0.0000

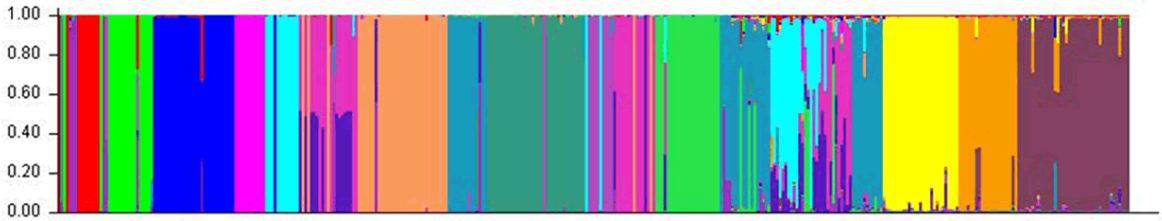


Figure 3.7: Distruct graphical representation of the InStruct results for all US and Chinese red rice populations. An optimal K = 16 as found for the red rice populations.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10	Group 11	Group 12	Group 13	Group 14	Group 15	Group 16	Group 17	Group 18	Group 19
Tx:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5855	0.0000	0.1679	0.0000	0.0000	0.2465	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001
La:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0681	0.0000	0.1730	0.0000	0.0000	0.3736	0.0000	0.0000	0.0000	0.3019	0.0000	0.0000	0.0835
Ar:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0214	0.0086	0.3786	0.0000	0.0000	0.0144	0.0000	0.0000	0.0000	0.0178	0.0061	0.0003	0.5527
Japonica:	0.0000	0.0008	0.0000	0.0010	0.0002	0.0007	0.0404	0.8840	0.0000	0.0704	0.0016	0.0000	0.0001	0.0006	0.0003	0.0000	0.0000	0.0000	0.0000
Liaoning:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Jiangsu:	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Guangdong:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Hainan:	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
Indica:	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9994	0.2368	0.0005
O. rufipogon:	0.0000	0.0187	0.1510	0.0215	0.0157	0.0157	0.0014	0.0010	0.0430	0.0000	0.0224	0.0238	0.4347	0.0179	0.1791	0.0000	0.0000	0.0000	0.0541

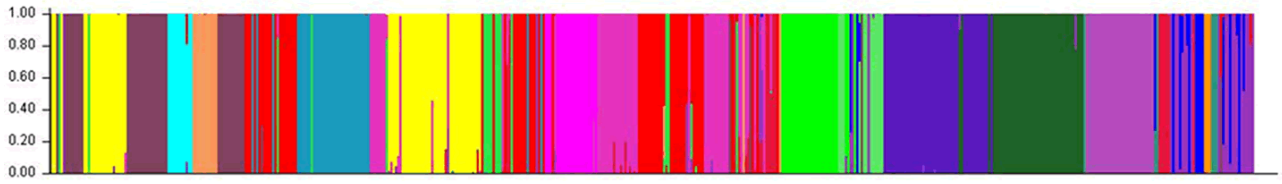


Figure 3.8: Distruct graphical representation of the InStruct results for all data; including: US and Chinese red rice populations, *O. sativa indica*, *O. sativa japonica* and *O. rufipogon*. An optimal K = 19 as found for the entire data set.

Chapter IV:

INTRA- AND INTER-SPECIFIC GENE FLOW IN POPULATIONS OF RED RICE IN THE SOUTHERN UNITED STATES[3]

[3] Kuntz, E. J., Mauricio, R. To be submitted to Heredity.

ABSTRACT

Gene flow is a general term used to describe mechanisms which move genetic information between individuals in the same population and among populations. Occurring through the movement of seeds and pollen, gene flow has the potential to act as a genetic bridge between individuals, populations and species. The balance between selfing and outcrossing, largely characterized by mating system, has a direct effect on the ability individuals to incorporate genetic variability into their populations. Red rice, the major noxious weed of rice cultivation, persists in sympatry with large numbers of conspecific cultivated congeners. Genetic admixture gained as the result of seed introduction and long flowering periods in red rice provide ample opportunity for intra-specific genetic exchange. Inter-specific gene flow in this system additionally opens the prospect of introgression of cultivated genetic material in the red rice, potentially creating novel, highly fit genotypes. In this paper we used paternity analysis to better understand patterns of genetic exchange by directly estimating the occurred and proportion of outcrossing events in four populations of red rice in the southern United States. We find direct evidence for crop to weed gene flow, suggesting that the introgression of cultivated alleles into weedy rice populations may be playing a role in the weeds persistence and survival. Out crossing rates comparable to other wild *Oryza* taxa imply that selection for the maintenance of a mixed-mating system is acting to increase diversity in red rice populations.

INTRODUCTION

Gene flow is the universal bridge that connects individuals, populations and ultimately species through genetic exchange. The power to connect groups is exemplified by the homogenizing effect gene flow has on populations (Slatkin, 1985b). High gene flow will limit divergent population structure, maintaining genetic continuity through allele sharing and ultimately leads to collective evolution of populations (Strasburg and Rieseberg, 2008). Continuously distributed populations have increasing levels of genetic similarity as a byproduct of increased genetic exchange (Slatkin, 1985b). Homogeneity among discrete populations is directly affected by the physical distances separating these populations and for plants the ability of pollen and seeds to overcome the geographic distance (Dunphy and Hamrick, 2007; Jones, 2010; Nason et al., 1996; Slatkin, 1993).

In the absence of gene flow populations will diverge in isolation. This separation, regardless of its source (allopatry, temporal or mechanistic) can ultimately lead to the creation of new species (Hoskin et al., 2005; Lande, 1980; Mayr, 2000). Selection and mutation can act to counter the effects of gene flow, but even minimal levels of genetic exchange maintain continuity across populations (Slatkin, 1985a; Slatkin, 1987). Under some conditions particularly strong selection for local environmental conditions is experienced can drive population differentiation in the face of gene flow. In these populations strong selection can cause local adaptation and population divergence (Brooks and Endler, 2001).

Overall, species vary widely in the extent to which they experience gene flow (Hamrick and Godt, 1996). Episodic in nature; gene flow can vary temporally, among populations and between individuals within the same populations (Kameyama et al., 2001; Watkins and Levin, 1990). Gene flow between plant populations can occur as the result of both pollen and seed movement. In order for gene movement to have an evolutionary impact it must make a contribution to the next generation and actually persist to change allele frequencies. There may be selection against migrants at different stages in this process. For instance, there may be selection against migrant seed or against pollen with different genetic

type, or for or against hybrid seeds. The frequency with which novel types can be introduced into the system will depend on its mode of delivery. Wind-dispersed pollen, as is seen in our species of interest *Oryza sativa*, can travel greater distances than pollen dispersed by animal vectors potentially increasing the incidences of outcrossing in the system (Rieger et al., 2002; Rognli et al., 2000).

Mating system is an exceptionally important life history trait that has direct implications for gene flow, diversity and population structure (Slatkin, 1987; Slatkin, 1993). Where drift and high levels of selfing tend to differentiate populations, gene flow tends to increase genetic variability overall, while decreasing population structure. Changes in mating system can have drastic affects on levels of genetic diversity, the ability of populations to persist and the potential for populations to respond to selection (Hamrick and Godt, 1996; Slatkin, 1985b). When populations shift from primarily selfing to outcrossing levels of heterozygosity and overall diversity increase, but these changes may be countered by the negative effects of outbreeding depression (Ellstrand, 1992; Parker, 1992). Domestication has affected mating system in many crops, usually limiting or eliminating outcrossing in the species as a means to preserve preferred domestication related traits. Selection for temporal isolation is observed in many crop systems, in which the distribution of flowering time in the domesticated species is narrower and offset from wild relatives (Doebley et al., 2006; Halsey et al., 2005). Selfing can act to retain ideal genotypes, locally adapted to their environments (Lande and Schemske, 1985; Parker, 1992).

Since, members of the *Oryza* genus are predominantly selfing, we expect weedy red rice to also be highly selfing, which would result in low to moderate levels of diversity and high levels of population structure in any given rice growing region. This is not the case. Instead US populations have high diversity levels and show little population structure across the southern rice growing region (Kuntz Chapter 2). These findings suggest that red rice may have undergone changes in its mating system, perhaps moving from the highly selfing phenotype seen in cultivated rice to a mating system in which outcrossing events occur with greater frequency. The timing of this change and the geographic context of the change remains unknown.

Multiple introductions of red rice to the United States from diverse geographical locations may explain the genetic and phenotypic diversity throughout the region. Gene flow by seed into red rice populations will increase admixture and augment genetic diversity. The occurrence of multiple divergent lineages of red rice resulting from disparate introduction events will provide opportunities for intra-specific gene flow to act to increase genetic variability. Genetic variability due to outcrossing can include novel genotypic combinations due to recombination and mixing of previousely isolated alleles (Allard, 1996; Ellstrand et al., 2010). This can lead to transgressive phenotypes and heterosis, reinforcing the shift toward outcrossing (Ledig, 1986; Lippman and Zamir, 2007). The incorporation of diversity from multiple sources has the potential to create very diverse, highly evolvable weed populations.

Hybridization between crops and their wild relatives or between divergent lineages of the same species can affect the generation and maintenance of weed populations, often times increasing weediness (Ellstrand et al., 2010; Ellstrand, 2009; Ellstrand and Schierenbeck, 2006). Weed populations that are able to incorporate agronomically important cultivated traits and retain weedy traits may generate the raw evolutionary material needed to generate super weeds (Kling, 1996). Cultivated rice have an abbreviated phenology compared to their wild relatives, usually only flowering for one to two weeks, limiting the potential for genetic exchange. Experimental approaches have confirmed that red rice and various cultivars will readily cross with one another (Shivrain et al., 2008; Shivrain et al., 2009; Zuo et al., 2011). When taken into the field this, along with an extended flowering phenology in red rice populations produces a situation in which the ability and potential to exchange genes with cultivated types exists (Figure 4.1). The extended flowering window experienced by red rice may allow red rice populations to incorporate alleles from multiple cultivars, and potentially multiple plantings in the same year. Gene flow between cultivated and red rice may lead to the introgression of alleles responsible for traits such as herbicide resistance, dwarf stature and various seed morphology traits, all of which have the potential to contribute to the ability of these populations survive in the face of strong selection against them.

In fact, some of the most successful agricultural weeds are crop mimics, having already acquired these agronomically fit traits (Baker, 1974; Barrett, 1983; Ellstrand et al., 2010). The acquisition of cultivated phenotypes helps weeds escape removal, allowing them to persist in the field until they set seed. Gaining crop phenotypic characteristics such as shorter stature, loss of awns and changes in hull pigmentation, will allow red rice to go undetected in the field, whereas maintaining weedy characteristics such as shattering, seed dormancy, high tillering capacity and long flowering times will allow them to persist as successful weeds. The balance of 'wild' and 'cultivar' phenotypic traits which allow red rice to escape removal and increase seed yield has the potential to generate super weeds.

Recognizing patterns and rates of gene flow in red rice populations in the southern United States will help us understand the patterns of genetic diversity of red rice seen across the region. To this end we directly estimated outcrossing rates in four populations using progeny arrays, such that outcrossing rates among plants within the same population and gene flow rates between populations could be directly calculated and compared. We then made inferences about the impacts of intra- and inter-specific gene flow in these red rice populations.

MATERIALS AND METHODS

Sampling

Four populations of red rice across the rice growing area in the southern United States were chosen for this analysis: Louisiana population (La7), Arkansas population (Ar3), Texas populations (Tx8, Tx14) (Figure 4.2). Samples were collected in July of 2006 a few weeks prior to the rice harvest in order to increase the proportion of mature red rice individuals collected. Individual panicles, including the flag leaf were collected at one meter intervals to ensure that clonal ramets were not sampled. The entire panicle, including the flag leaf and seeds, of a single maternal plant was collected and stored in seed

enveloped with silica gel until ready for germination and DNA extraction at the University of Georgia. Seeds of cultivated rice from each field were also collected at this time.

Several criteria were used when choosing populations to uses for this study. To insure that inter-specific gene flow could be assayed, we choose populations were required to come from rice fields in which the cultivar grown was known to have multiple cultivar specific alleles. To insure we could estimate intra-specific gene flow, previous analysis, (Kuntz Chapter2) was used to choose populations that were as geographically isolated and as genetically diverged from one another as possible. In addition to genetic diversity (as measured at neutral loci) populations were also chosen to represent the morphological diversity seen red rice across the south (Table 4.1).

DNA extraction

Thirty to fifty seeds per maternal plant were germinated on Watman filter paper in Petri dishes and grown in a growth chamber at 30°C and germination rates were scored. Seedlings were collected after 10-20 days and the entire plantlets of 8-35 sibling plants were snap frozen in liquid nitrogen. Flag leaf tissue was collected and pulverized using liquid nitrogen in order to extract maternal DNA. Once frozen and homogenized, total genomic DNA for all tissues were extracted using Qiagen's DNAeasy mini kit (Valencia, CA). All extracted DNA was quantified using a spectrophotometer (NanoDrop/ Wilmington, DE) and standardized to approximately 30 ng DNA per µl.

Microsatellite genotyping

Eleven microsatellite loci (SSRs) known to be polymorphic within these populations and among the cultivar and red rice populations were chosen from previously identified microsatellite regions (Temnykh et al., 2001). Polymerase chain reaction (PCR) amplifications were performed using three primers: a SSR sequence specific forward primer with an M13 tail at its 5' end (5'-CACGACGTTGTAAAACGACA-3'), a SSR sequence specific reverse primer and a fluorescent-labeled

M13 primer (Schuelke, 2000). Thirteen microliter reactions contained 1.5µl of DNA sample, 0.1µl 10 mM forward SSR primer, 0.5µl 10 mM reverse SSR primer, 0.5µl 10 mM M13 primer, 1.0 µl 10 mM dNTPs, 1.0µl 10 mM MgCl2, 1.0µl 10XPCR buffer, and one unit Taq polymerase. Samples were amplified using the following PCR profile: initial denaturation at 94°C for 3min; followed by 35 cycles of: 94°C for 45 sec, 55°C annealing for 45sec, 72°C extension for 45sec; and a final extension at 72°C for 15 min. Post PCR, samples were held at 4°C. The resulting PCR fragments were diluted 1:10 and genotyped on a capillary electrophoresis sequencer ABI3700 (Applied Biosystems; Foster City, Ca) using MapMarker1000 ROX (BioVentures Inc; Murfreesboro, TN) as an internal size standard.

Estimates of genetic diversity

All microsatellites loci were scored using GeneMarker, version 1.70 (SoftGenetics; State College, PA). Genetic diversity statistics were calculated for all four populations using between 21-27 maternal plants per population (Table 4.2). Patterns of genetic diversity were calculated using GENALEX (Peakall and Smouse, 2006); estimating the of percentage of polymorphic loci (P_p), the mean number of alleles per locus (A), the mean number of alleles per polymorphic loci (AP), and the effective number of alleles per locus (A_e). Genetic diversity, as measured by expected heterozygosity (H_e) and observed genetic diversity were calculated with the FSTAT software package (Goudet, 1995) and 1000 bootstraps were used to generate standard deviation around these values (Table 4.2).

Direct measures of gene flow

Estimates of multi-locus (t_m) and single-locus (t_s) outcrossing rates were estimated using the methods of Ritland (2002) as implemented by the software program MLTR (Ritland, 2002; Ritland and Jain, 1981). Paternity exclusion analysis was used to estimate levels of intra-specific gene flow between red rice individuals within each population. MLTR used a two step process to generate outcrossing rates. First, the program assigns the most likely maternal genotype for each family, but because we have maternal tissue, in most cases (99%) the maternal genotype was known definitively and assigned to the

family. In the few cases where the maternal DNA quality was poor or the genotype could not be ascertained the maternal genotype was inferred from the progeny genotype arrays. Once the maternal genotype is assigned to each family, the progeny array is used to generate estimates of outcrossing based on deviations in sibling genotypes as compared to the maternal genotype. Estimates of confidence intervals around there outcrossing estimates were generated by re-sampling progeny within families across 1000 bootstraps. If mating occurred between relatives (or individuals with the same genotype at the loci scored), then some outcrossing events will be missed and scored as selfing events. This cryptic gene flow can be estimated using the difference between the multi-locus and single-locus outcrossing rates, as this provides a minimal estimate of the fraction of apparent selfing events due to biparental inbreeding. The uses of loci with private cultivar alleles, or alleles found in cultivated rice, which are absent from red rice in these populations, allowed us to track inter-specific outcrossing events.

RESULTS

Populations of red rice were found to be highly phenotypically different, both within and between populations and differed in population size. The red rice population found in Ar3 was collected from a field growing Clearfield 151 (CL-151), which is an herbicide resistant cultivar. These seeds had the lowest germination rate (5.6%) of all the populations, possibly due to negative effects on seed development and viability as the result of herbicide application. All seeds from Ar3, a moderately sized population (>100) were red, although both straw hull, awn-less and black hull, awned types were found (Table 4.1). Red rice from La7, a large population of over 1000 plants, was collected from a Coco cultivar field. All red rice individuals collected in this field had red pericarps and were straw hulled, but both awned and awn-less biotypes were found. Germination rates were highest in this population, with 79% of the seeds germinating. Tx14, the smallest population, with approximately 50 individuals was found in another Clearfield (CL-161) field. All apparent red rice individuals were collected. Although this was the

smallest population, seed diversity in this population was greatest, with straw hull, awned individuals with either white or red seeds, or straw hull awn-less individuals with either white or red seeds found. This population had the second lowest germination rate, again suggesting that herbicide application in these Clearfield cultivar fields may have affected seed viability. Tx8 was also a large population, with >1000 red rice individuals found in a Chenier rice field. All red rice individuals sampled from this population contained red seeds and had black, awned hulls. Germination rates in this population were low (27%) but were higher than either of the red rice populations collected in Clearfield cultivar fields.

Genetic diversity

The maternal plants were used to calculate diversity across the 4 populations. Of the populations assayed all 11 loci screened were polymorphic in at least one population, with an average of 3.39 alleles per polymorphic locus (Table 4.2). The mean number of alleles per locus ranged from 3.27-4.09, with 1.85 mean effective alleles per locus. In all populations the observed heterozygosity (H_o) was lower than the expected heterozygosity (H_e), with a mean H_o of 0.075 and a mean H_e of 0.358, suggesting high levels of selfing or bi-parental inbreeding.

Genetic diversity estimates between red rice populations found in non-herbicide treated fields were not significantly different from red rice populations found in Clearfield herbicide resistant fields. Surprisingly, both red rice populations collected in Clearfield fields (Ar3 and Tx14) had slightly higher, although not significantly, observed heterozygosity levels (0.110 and 0.082 respectively) than the red rice populations found in standard cultivar fields. This result was unexpected, given that these red rice populations were collected in the first year that the Clearfield cultivars were grown, and we would expect that the application of herbicide to these fields would kill many susceptible red rice biotypes, thus lowering overall genetic diversity.

Mating system and direct estimates of gene flow

Direct estimates of outcrossing indicated higher levels of outcrossing than estimated measures in all populations (Table 4.3). The mean multilocus estimate of outcrossing rate (t_m) across all four populations was 0.204, while the mean single-locus estimate was 0.076, resulting in an estimated biparental inbreeding (t_m-t_s) of 0.129, or 13%. This estimate of biparental inbreeding is a conservative estimate given the high levels of allele sharing amongst the maternal plants. Allele frequencies were not significantly different between pollen and ovule pools. Because there is no difference between the pollen and ovule allele frequencies we can assume that pollination from red rice genotypes present in the field but not sampled was minimal.

Estimates of individual maternal outcrossing proportion ranged from 0-86%, indicating a wide range in the number of progeny formed due to outcrossing (versus selfing) among individual maternal plants. The average proportion of outcrossed progeny varied across population, with La7 having the lowest average number of outcrossing events per maternal plant (0.080). Tx8 mothers outcrossed less frequently on average, but outcrossed a higher percentage of their progeny (0.420) than maternal plants in other populations. The average proportion of outcrossing across all mothers in a given population did not correlate one-to- one with direct estimates of outcrossing, indicating that population level statistics do not take into account the internal population variation in mating system between individual plants. When looking specifically at this dynamic we see that in some populations only a few mothers outcross, but outcross a high percentage of their progeny; while in other populations many mothers outcross, but only contribute a few seeds per panicle. These different patterns of outcrossing rates among mothers can have differing effects on population levels of diversity, and may be directly correlated to flowering timing. All estimates confirm that red rice has a mixed mating system, unlike their selfing cultivar counterparts.

Inter- specific gene flow was also observed in the La 7 population. One of the screened maternal plants produced two progeny with cultivated rice. Parental exclusion in this population allowed for the

assignment of the Coco cultivar as the paternal parent of these two red rice progeny with 100% exclusion of all other possible red rice parents sampled. It is interesting to note that genetic diversity as measured by H_e was lowest in this population (0.190) as compared to the other populations. The multi-locus estimates of outcrossing ranged from 0-0.717 for mothers in this population, with an average of 0.142. The estimate of biparental inbreeding in the population was 13.4%, virtually identical to the across population average

The Tx8 red rice population was found in an uncontrolled Chenier field. The red rice infestation was so bad in this field that the farmer 'gave up' on the possibility of harvesting and left the plants completely unmanaged. The multilocus outcrossing rate in this population averaged 0.162, with a range of 0-0.836 (Table 4.4). Biparental inbreeding in this population was estimated to be the lowest of all four populations investigated, with an estimate of 9%. Although this population did not have the highest outcrossing rate, it was the most diverse population (He = 0.475), possibly due to relaxed selection against these weeds due to the lack of weed management in the field.

Tx14 and Ar3 populations of red rice were found in fields growing Clearfield varieties for the first time. It was surprising to find populations of red rice growing in Clearfield cultivar fields, as we would suspect the use of herbicide would exclude all red rice individuals. The use of herbicide in these populations may be the reason that these two populations had such small population sizes (with approximately 100 plants in Ar3 and about 50 plants in Tx14), and low germination rates (Ar3 = 5.6% and Tx14 = 17.9%) (Table 4.1). Multilocus estimates of outcrossing were similar between the two populations with an average of 0.239 in Ar3 and 0.272 in Tx14. The highest percent outcrossing for a single mother was seen in the Tx14 population at 86% of its progeny resulting from outcrossing. It wa surprising that both of these populations had high diversity levels, Ar3 (H_e = 0.390) and Tx14 (H_e = 0.379), since we would anticipate these populations to have undergone diversity reduction as the result of herbicide application and the death of many red rice individuals.

DISCUSSION

Our estimates of outcrossing confirm that red rice exhibits a mixed mating system, with variation in the proportion of outcrossing among populations and maternal plants within populations. Direct estimates of gene flow in these four populations resulted in an estimated average outcrossing rate of 20% in red rice. This estimate is much closer to outcrossing levels seen in *O. rufipogon* populations. Morishima estimated population level outcrossing rates of 5-60%, with a striking division between annual highly selfing morphs and perennial mixed mating *O. rufipogon* plants (Morishima and Barbier, 1990). Data suggests that the annual selfing populations of *O. rufipogon* are the progenitor populations of cultivated rice (Cheng et al., 2003; Londo et al., 2006; Morishima and Barbier, 1990; Rakshit et al., 2007). It is possibly that the hybrid evolutionary history of red rice populations in the US is the result of genetic exchange between cultivated rice and perennial populations of *O. rufipogon.* Outcrossing rates varied greatly among maternal plants within populations, ranging from 0-86%.

Red rice is generally thought of as a weedy variant of the cultivated *Oryza sativa*, and as such it has always been assumed that these plants are predominantly selfing (Vaughan et al., 2001). This study shows that despite its close evolutionary relationship with the cultivated rice, changes in mating system have occurred, significantly increasing outcrossing levels from virtually 0% as seen in cultivated rice to levels greater than 80%. Evolutionary implications of moving from a virtually exclusively selfing mating system to a mixed mating system are great, with repercussions on levels of diversity and the ability of these populations to respond to selection (Hamrick and Godt, 1996; Mariac et al., 2006; Sakai et al., 2001). In wild populations of barley and barley landraces outcrossing rates have been shown to be influenced by maternal genotype and environmental conditions, with rates of outcrossing varying in different seasons (Abdel-Ghani et al., 2004). Our results are similar, with variation in the outcrossing rate of maternal plants.

Variation in outcrossing rates is profoundly influenced by flowering time in both the red rice population and the cultivar grown in the field. Cultivars generally have very narrow flowering window of only one to two weeks, and vary widely in flowering time, with some flowering early in the season while others flower mid season (Gealy et al., 2002). This is further complicated by the agricultural practice of sequential plantings to stagger seed maturity in the field, which effectively widens the flowering window. Red rice populations with longer flowering periods will experience more opportunities for outcrossing, potentially overlapping with the flowering window of multiple cultivars and increasing the chances that inter-specific outcrossing will occur. Although only a single cultivar is generally planted in a given field each year, farmers often switch cultivars season to season leading to temporal exposure of red rice populations to multiple cultivars.

Finding evidence of inter-specific outcrossing between red rice and cultivated rice in the La7 population confirms the potential for introgression of cultivated alleles into red rice populations, and validates that the hypothesis that local US cultivar alleles are moving into red rice populations. The agricultural system strongly selects against weeds and as such the survival of many agricultural pests is predicated on the ability to escape notice until setting seed (Harlan, 1992). The introgression of cultivar traits such as herbicide resistance and dwarf stature will inevitably increase the ability of red rice populations to survive in the agricultural system (Campbell et al., 2006; Linder et al., 1998; Snow et al., 2003). The genetic diversity seen in red rice populations makes it difficult to track inter-specific gene flow because there are so few alleles that are exclusive to the cultivars. As such we are probably greatly underestimating gene flow between cultivated and weedy types due to a lack of power to detect these outcrossing events. There are many common cultivar alleles in red rice populations at low levels which are most likely the result of red rice X cultivar outcrossing events in the past, but we cannot use these to directly monitor inter-specific gene flow. The presence of these common cultivar alleles in red rice populations suggests that gene flow between red rice and cultivars has occurred in the past and that these alleles are maintained in the population. Although outcrossing can be episodic the incorporation of

beneficial alleles such as herbicide resistance will inevitably have a profound impact on red rice population persistence and survival.

The occurrence of red rice in fields using herbicide resistant cultivars suggests that there is some native level of either herbicide resistance or extreme tolerance to imidazolinone herbicides in red rice populations in the south. If there is variation for herbicide resistance in these populations, then given high levels of outcrossing and strong selection against individuals without resistance, the utility of these herbicide resistant cultivars will be short lived. Conversely it could be agricultural practices, specifically the misuse of herbicide regimens, or neglecting/ missing marginal areas of the field when applying herbicide treatments, that allow red rice to persist. This situation, although evolutionarily different, is also potentially dangerous considering that red rice readily outcrosses, which would afford these populations the opportunity for genetic exchange with Clearfield herbicide-resistant cultivars. Poorly managed Clearfield red rice populations have the potential to gain herbicide resistance, possibly leading to the creation of super-weed populations (Gressel, 1999; Kling, 1996). Once gained, there will be strong selection for the maintenance of herbicide resistance as long as herbicides are used in these fields.

CONCLUSION

Crops and their weeds are inextricably linked. De Wet and Harlan ask the question in their classic 1975 paper "Weeds and Domesticates: Evolution in the man-made habitat": if humans act to select crops that grow densely in monospecific stands, are those plants only a few allele changes away from becoming invasive or weedy? Empirical evidence suggests that many of agriculture's most noxious weeds are the direct result of direct of 'crops going wild' either via *in situ* feralization, or by hybridization with closely related wild species. The shared evolutionary history seen in so many crop/weed complexes suggests that these populations are able to generate divergent phenotypes and evolve in sympatry.

Grown in the same tightly planted, disturbed agricultural environment, crops and weeds experience vastly different selection pressures in the field. Genetic exchange between crops, their wild relatives and weeds provide a complex web of genetic interactions which if detangled will allow researchers to better understand the differences in genetic architecture leading to weedy and cultivated types. Weeds which strike a balance between crop-like and weedy traits will be able to respond to the strong selection pressures exerted against them and therefore persist in the agricultural environment. Monitoring levels of gene flow, especially in agricultural fields employing genetically modified crops will be important in controlling the generation of super weeds. The introgression of 'super traits' such as herbicide resistance into weed populations will likely have devastating effects on agricultural production and the ability to successfully manage weed populations. A laissez faire approach to weed management in many areas may be contributing to the generation of hyper aggressive, virtually undetectable weeds.

REFERENCES

Abdel-Ghani A.H., Parzies H.K., Omary A., Geiger H.H. (2004) Estimating the outcrossing rate of barley landraces and wild barley populations collected from ecologically different regions of Jordan. TAG Theoretical and Applied Genetics 109:588-595.

Allard R. (1996) Genetic basis of the evolution of adaptedness in plants. Euphytica 92:1-11.

Baker H.G. (1974) The evolution of weeds. Annual Review of Ecology and Systematics 5:1-24.

Barrett S.H. (1983) Crop mimicry in weeds. Economic Botany 37:255-282.

Brooks R., Endler J.A. (2001) Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). Evolution 55:1002-1015.

Campbell L.G., Snow A.A., Ridley C.E. (2006) Weed evolution after crop gene introgression: greater survival and fecundity of hybrids in a new environment. Ecology Letters 9:1198-1209.

Cheng C., Motohashi R., Tsuchimoto S., Fukuta Y., Ohtsubo H., Ohtsubo E. (2003) Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Molecular Biology and Evolution 20:67.

Doebley J.F., Gaut B.S., Smith B.D. (2006) The molecular genetics of crop domestication. Cell 127:1309-1321.

Dunphy B.K., Hamrick J.L. (2007) Estimation of gene flow into fragmented populations of Bursera simaruba (Burseraceae) in the dry-forest life zone of Puerto Rico. American Journal of Botany 94:1786-1794.

Ellstrand N., Heredia S., Leak-Garcia J., Heraty J., Burger J., Yao L., Nohzadeh-Malakshah S., Ridley C. (2010) Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evolutionary Applications:494-504.

Ellstrand N.C. (1992) Gene flow by pollen- implaications for plant conservation genetics. Oikos 63:77-86.

Ellstrand N.C. (2009) Evolution of invasiveness in plants following hybridization. pp. 1089-1091.

Ellstrand N.C., Schierenbeck K.A. (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? Euphytica 148:35-46.

Gealy D.R., Tai T.H., Sneller C.H. (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Science 50:333-339.

Goudet J. (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of Heredity 86:485.

Gressel J. (1999) Tandem constructs: preventing the rise of superweeds. Trends in Biotechnology 17:361-366.

Halsey M.E.R., Davis K.M., Qualls C.A., Eppard M., Berberich P.J., Sharon A. (2005) Isolation of maize from pollen-mediated gene flow by time and distance. Crop Science 45:2172.

Hamrick J., Godt M. (1996) Effects of life history traits on genetic diversity in plant species. Philosophical transactions: biological sciences 351:1291-1298.

Harlan J.R. (1992) Crops and Man, American Society of Agronomy, Madison, WI.

Hoskin C.J., Higgie M., McDonald K.R., Moritz C. (2005) Reinforcement drives rapid allopatric speciation. Nature 437:1353-1356.

Jones A. (2010) Reconciling field observations of dispersal with estimates of gene flow. Mol Ecol 19:4379-82.

Kameyama Y., Isagi Y., Nakagoshi N. (2001) Patterns and levels of gene flow in *Rhododendron metternichii var hondoense* revealed by microsatellite analysis. Molecular Ecology 10:205-216.

Kling J. (1996) Agricultural Ecology: Could Transgenic Supercrops One Day Breed Superweeds? Science 274:180.

Lande R. (1980) Genetic variation and phenotypic evolution during allopatric speciation. The American Naturalist 116:463-479.

Lande R., Schemske D.W. (1985) The evolution of self-fertilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24-40.

Ledig F.T. (1986) Heterozygosity, heterosis, and fitness in outbreeding plants. in M.E.Soule (ed.)Conservation Biology: The Science of Scarcity and Diversity. Sinauer Associates, Inc. Sunderland, Ma.

Linder C.R., Taha I., Seiler G.J., Snow A.A., Rieseberg L.H. (1998) Long-term introgression of crop genes into wild sunflower populations. Theoretical and Applied Genetics 96:339-347.

Lippman Z.B., Zamir D. (2007) Heterosis: revisiting the magic. Trends in Genetics 23:60-66.

Londo J.P., Chiang Y.C., Hung K.H., Chiang T.Y., Schaal B.A. (2006) Phylogeography of Asian wild rice, Oryza rufipogon, reveals multiple independent domestications of cultivated rice, Oryza sativa. Proceedings of the National Academy of Sciences of the United States of America 103:9578-9583.

Mariac C., Robert T., Allinne C., Remigereau M.S., Luxereau A., Tidjani M., Seyni O., Bezancon G., Pham J.L., Sarr A. (2006) Genetic diversity and gene flow among pearl millet crop/weed complex: a case study. Theoretical and Applied Genetics 113:1003-1014.

Mayr E. (2000) The biological species concept. Species concepts and phylogenetic theory. A debate:17-29.

Morishima H., Barbier P. (1990) Mating System and Genetic Structure of Natural Populations in Wild Rice *Oryza rufipogon*. Plant Species Biology 5:31-39.

Nason J.D., Herre E.A., Hamrick J.L. (1996) Paternity analysis of the breeding structure of strangler fig populations: evidence for substantial long distance wasp dispersal. Journal of Biogeography 23:501-512.

Parker M.A. (1992) Outbreeding depression in a selfing annual. Evolution 46:837-841.

Peakall R., Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6:288-295.

Rakshit S., Rakshit A., Matsumura H., Takahashi Y., Hasegawa Y., Ito A., Ishii T., Miyashita N.T., Terauchi R. (2007) Large-scale DNA polymorphism study of *Oryza sativa* and *O. rufipogon* reveals the origin and divergence of Asian rice. TAG Theoretical and Applied Genetics 114:731-743.

Rieger M.A., Lamond M., Preston C., Powles S.B., Roush R.T. (2002) Pollen-mediated movement of herbicide resistance between commercial canola fields. Science 296:2386.

Ritland K. (2002) Extensions of models for the estimation of mating systems using n independent loci. Heredity 88:221-228.

Ritland K., Jain S. (1981) A model for the estimation of outcrossing rate and gene frequencies using n independent loci. Heredity 47:35-52.

Rognli O.A., Nilsson N.O., Nurminiemi M. (2000) Effects of distance and pollen competition on gene flow in the wind pollinated grass Festuca pratensis Huds. Heredity 85:550-560.

Sakai A.K., Allendorf F.W., Holt J.S., Lodge D.M., Molofsky J., With K.A., Baughman S., Cabin R.J., Cohen J.E., Ellstrand N.C., McCauley D.E., O'Neil P., Parker I.M., Thompson J.N., Weller S.G. (2001) The population biology of invasive species. Annual Review of Ecology and Systematics 32:305-332.

Schuelke M. (2000) An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18:233-234.

Shivrain V.K., Burgos N.R., Gealy D.R., Moldenhauer K.A.K., Baquireza C.J. (2008) Maximum Outcrossing Rate and Genetic Compatibility between Red Rice (*Oryza sativa*) Biotypes and Clearfield (TM) Rice. Weed Science 56:807-813.

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

Slatkin M. (1985a) Rare alleles as indicators of gene flow. Evolution 39:53-65.

Slatkin M. (1985b) Gene flow in natural populations. Annual Review of Ecology and Systematics 16:393-430.

Slatkin M. (1987) Gene flow and the geographic structure of natural populations. Science 236:787.

Slatkin M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47:264-279.

Snow A.A., Pilson D., Rieseberg L.H., Paulsen M.J., Pleskac N., Reagon M.R., Wolf D.E., Selbo S.M. (2003) A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. Ecological Applications 13:279-286.

Strasburg J.L., Rieseberg L.H. (2008) Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris*- large effective population sizes and rates of long term gene flow. Evolution 62:1936-1950.

Temnykh S., DeClerck G., Lukashova A., Lipovich L., Cartinhour S., McCouch S. (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Research 11:1441-1452.

Vaughan L.K., Ottis B.V., Prazak-Havey A.M., Bormans C.A., Sneller C., Chandler J.M., Park W.D. (2001) Is all red rice found in commercial rice really *Oryza sativa*? Weed Science 49:468-476.

Watkins L., Levin D.A. (1990) Outcrossing rates as related to plant density in *Phlox drummondii*. Heredity 65:81-89.

Zuo J., Zhang L., Song X., Dai W., Qiang S. (2011) Innate factors causing differences in gene flow frequency from transgenic rice to different weedy rice biotypes. Pest Manag Sci.

Table 4.1: Red rice population characteristics

Population	Cultivar	Estimated population size	Germination rate	seed color	Hull color/Awn status
Ar3	Cl-151	>100	0.056	R	Straw hull/no awn
				R	Bk hull/awn
La7	Coco	>1000	0.786	R	Straw hull/no awn
				R	Straw hull/awn
Tx8	Chenier	>>1000	0.270	R	Bk hull /awn
Tx14	Cl-161	~50	0.179	R & W	Straw hull/awn
				R & W	Straw hull/ no awn

Four populations from the southern United States were directly assayed for estimates of outcrossing. The *O. sativa japonica* cultivar grown in the field where the red rice populations were collected; an estimate of red rice population size; germination rate; and seed morphological characteristics (including: awn presence of absence, seed and hull color) are assigned.

Table 4.2: Genetic diversity statistics describing four red rice populations in the United States.

Population	N	Pp	AP	A	Ae	Ho (SD)	He (SD)
Ar3	24	100.00%	4.091	4.091	1.856	0.110 (0.039)	0.390 (0.066)
La7	25	90.91%	2.978	3.273	1.293	0.044 (0.013)	0.190 (0.049)
Tx8	27	90.91%	3.392	3.727	2.099	0.064 (0.024)	0.475 (0.059)
Tx14	21	81.82%	4.656	3.818	2.171	0.082 (0.020)	0.379 (0.091)
Mean	24	90.91%	3.392	3.727	1.855	0.075 (0.013)	0.358 (0.036)

Genetic diversity statistics were calculated, including: Pp, percent polymorphic loci; AP, mean number of alleles per polymorphic locus; A, mean number of alleles per locus; Ae, effective number of alleles; Ho, observed heterozygosity; and He, expected heterozygosity (gene diversity). Standard deviations (SD) were calculated by bootstrapping.

Table 4.3: Population level estimates of outcrossing

Population	Cultivar	Outcrossing		Av. maternal outcrossing rate
		Estimated	Direct	
Ar3	Cl-151	0.164	0.239	0.140
La7	Coco	0.131	0.142	0.080
Tx8	Cheniere	0.072	0.162	0.420
Tx14	Cl-161	0.121	0.272	0.370
mean		0.122	0.204	0.253

Population level estimates of outcrossing rates were estimated using genetic diversity statistics generated from all mothers in each population. Direct estimates of average population outcrossing rate and individual estimates of average maternal outcrossing rate were calculated using progeny array data in MLTR.

Table 4.4: Family level estimates of outcrossing

Population	Family (Nsibs)	t_m(SD)	t_s(SD)	t_m- ts(SD)
Ar3	a(12)	0.251 (0.121)	0.071 (0.036)	0.180 (0.086)
	b(12)	0.169 (0.081)	0.109 (0.053)	0.060 (0.030)
	c(13)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	d(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	e(11)	0.640 (0.307)	0.337 (0.169)	0.303 (0.143)
	f(10)	0.503 (0.238)	0.146 (0.074)	0.357 (0.166)
	g(12)	0.334 (0.158)	0.101 (0.051)	0.233 (0.109)
	h(11)	0.647 (0.312)	0.286 (0.147)	0.361 (0.170)
	i(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	j(12)	0.084 (0.039)	0.027 (0.013)	0.057 (0.026)
	k(12)	0.167 (0.079)	0.046 (0.022)	0.121 (0.059)
	l(12)	0.337 (0.160)	0.150 (0.076)	0.187 (0.086)
	m(12)	0.422 (0.202)	0.165 (0.082)	0.258 (0.122)
	n(13)	0.154 (0.074)	0.059 (0.030)	0.095 (0.046)
	o(12)	0.084 (0.039)	0.048 (0.024)	0.035 (0.017)
	p(15)	0.335 (0.160)	0.156 (0.078)	0.179 (0.085)
	q(10)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	r(10)	0.200 (0.096)	0.038 (0.018)	0.162 (0.078)
	s(12)	0.083 (0.040)	0.020 (0.009)	0.064 (0.031)
	t(11)	0.273 (0.133)	0.067 (0.033)	0.246 (0.100)
	u(13)	0.308 (0.146)	0.049 (0.024)	0.258 (0.122)
	v(12)	0.167 (0.080)	0.029 (0.014)	0.137 (0.066)
	w(11)	0.182 (0.085)	0.043 (0.020)	0.139 (0.065)
	x(16)	0.376 (0.182)	0.107 (0.054)	0.269 (0.129)
		0.239	0.078	0.163
La7	b(12)	0.182 (0.087)	0.083 (0.042)	0.099 (0.046)
	c(12)	0.091 (0.044)	0.041 (0.021)	0.050 (0.024)
	d(12)	0.091 (0.045)	0.041 (0.022	0.050 (0.024)
	e(12)	0.183 (0.088)	0.165 (0.086)	0.018 (0.019)
	f(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	g(12)	0.717 (0.340)	0.457 (0.218)	0.26 (0.130)
	h(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	i(12)	0.274 (0.136)	0.124 (0.067)	0.150 (0.072)
	j(13)	0.077 (0.037)	0.052 (0.026)	0.025 (0.013)
	k(12)	0.367 (0.178)	0.212 (0.113)	0.154 (0.072)
	l(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	m(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	n(12)	0.183 (0.089)	0.209 (0.108)	-0.026 (0.027)

	o(12)	0.364 (0.180)	0.206 (0.110)	0.158 (0.075)
	p(15)	0.134 (0.064)	0.070 (0.035)	0.064 (0.030)
	q(13)	0.154 (0.072)	0.049 (0.023)	0.105 (0.049)
	r(12)	0.087 (0.041)	0.036 (0.017)	0.051 (0.024)
	s(11)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	t(9)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	u(13)	0.155 (0.074)	0.052 (0.025)	0.103 (0.049)
	v(8)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	w(11)	0.091 (0.043)	0.022 (0.010)	0.069 (0.033)
	x(12)	0.167 (0.078)	0.052 (0.025)	0.115 (0.053)
	y(14)	0.072 (0.033)	0.023 (0.011)	0.048 (0.022)
	z(14)	0.143 (0.068)	0.044 (0.022)	0.099 (0.047)
	mean	0.142	0.008	0.134
Tx8	a(11)	0.093 (0.045)	0.064 (0.032)	0.029 (0.014)
	b(8)	0.260 (0.129)	0.098 (0.057)	0.161 (0.076)
	c(14)	0.172 (0.086)	0.096 (0.056)	0.076 (0.037)
	d(13)	0.836 (0.407)	0.285 (0.161)	0.551 (0.257)
	e(12)	0.258 (0.127)	0.096 (0.055)	0.162 (0.077)
	f(8)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	g(12)	0.421 (0.205)	0.169 (0.101)	0.251 (0.116)
	h(12)	0.450 (0.280)	0.242 (0.191)	0.208 (0.106)
	i(12)	0.175 (0.085)	0.107 (0.060)	0.068 (0.034)
	j(12)	0.083 (0.039)	0.064 (0.031)	0.019 (0.012)
	k(12)	0.417 (0.199)	0.143 (0.074)	0.274 (0.129)
	l(12)	0.167 (0.081)	0.072 (0.036)	0.095 (0.046)
	m(12)	0.433 (0.222)	0.208 (0.137)	0.225 (0.107)
	n(12)	0.421 (0.205)	0.162 (0.099)	0.259 (0.118)
	o(12)	0.172 (0.085)	0.096 (0.056)	0.076 (0.037)
	p(9)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	q(8)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	r(8)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	s(9)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	t(10)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	u(9)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	v(11)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	w(11)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	x(8)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	y(14)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	z(13)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	aa(14)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
	mean	0.162	0.071	0.091
Tx14	b(11)	0.364 (0.174)	0.116 (0.060)	0.248 (0.117)

c(13)	0.555 (0.269)	0.368 (0.191)	0.186 (0.086)
d(11)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
e(13)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
f(11)	0.364 (0.176)	0.112 (0.058)	0.252 (0.119)
g(12)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
h(13)	0.470 (0.224)	0.389 (0.189)	0.081 (0.046)
i(14)	0.167 (0.080)	0.088 (0.045)	0.079 (0.038)
j(11)	0.001 (0.000)	0.001 (0.000)	0.001 (0.000)
k(9)	0.448 (0.217)	0.156 (0.081)	0.291 (0.138)
l(19)	0.435 (0.212)	0.222 (0.113)	0.213 (0.101)
m(16)	0.764 (0.369)	0.545 (0.274)	0.219 (0.104)
n(12)	0.855 (0.412)	0.530 (0.279)	0.324 (0.148)
o(11)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
p(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
q(15)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
r(12)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
s(8)	0.375 (0.180)	0.169 (0.084)	0.206 (0.097)
t(19)	0.501 (0.237)	0.211 (0.106)	0.290 (0.134)
u(20)	0.401 (0.190)	0.124 (0.061)	0.277 (0.129)
v(9)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)
mean	0.272	0.145	0.127

Estimates of multi-locus outcrossing rate(t_m), the average single locus outcrossing rate (t_s) among families in all four populations, and t_m- t_s the minimal fraction of apparent selfing events due to biparental inbreeding were calculated. Standard deviation obtained from 1000 bootstraps.

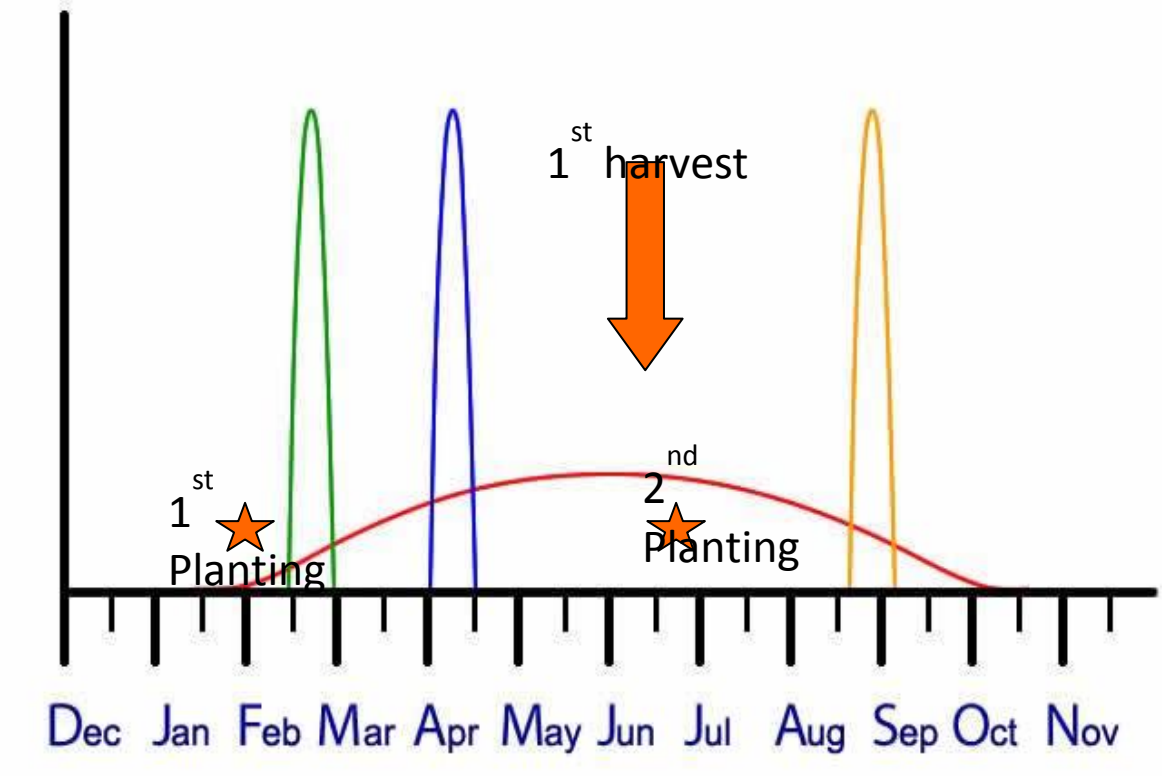


Figure 4.1: Diagram of flowering phenology seen in cultivated and red rice. Cultivated rice, as indicated by the green and blue peaks, flowers in a short, discrete window usually one to two weeks in length. The exact timing of flowering depends on the cultivar and date of planting. Red rice, as indicated by the red line, has a much longer phenology, with the potential for synchronous flowering with multiple cultivars.

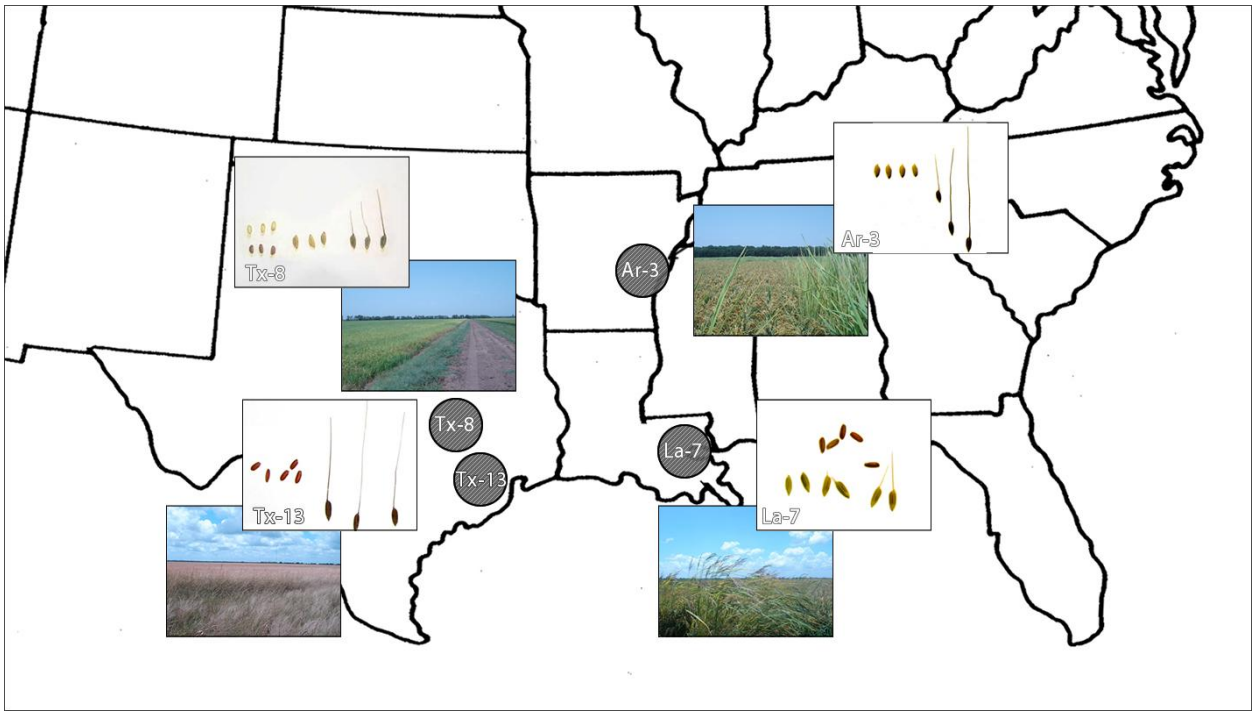


Figure 4.2: Map of collection sites in the southern United States with pictures of the field sites and seen morphologies found.

Chapter V: Summary and Conclusions

At the outset of this study, little was known regarding the evolutionary history, population genetics, or genetic structure in populations of red rice in the United States. Red rice, the major weed of rice agriculture worldwide, causes yield loss by outcompeting rice in the field and lowering grain value if harvested (Gealy et al., 2000). The evolutionary origins of red rice populations in many areas across the globe remain unknown. The generation of weeds from cultivated types is generally thought to occur via two major mechanisms, de-domestication (endo-ferality) and crop to wild relative hybridization (exo-ferality) (De Wet and Harlan, 1975). Data from various localities worldwide provide mixed and often conflicting results regarding the mechanisms and processes involved in generating and maintaining red rice populations (Gao et al., 2000; Gealy et al., 2002; Morishima and Barbier, 1990). Our approach to understanding this problem in the United States was to employ population level sampling of red rice throughout the southern rice growing region in an attempt to clarify the evolutionary history of the weeds. Red rice populations along with cultivar (both *Oryza indica* and *japonica*) and *O. rufipogon* accessions were utilized to assay genetic diversity and population structure in the United States. This knowledge was then used as a springboard from which to explore additional dynamics; including, introduction history and gene flow within and among red rice populations. These results will collectively help researchers understand the generation, maintenance and potentially the evolutionary stability and trajectory of these weed populations.

We first sought to examine diversity in red rice populations from the southern United States using microsatellite markers. Field collected red rice populations are highly morphologically and genetically diverse, forming a distinct taxonomic group. Variation in plant morphology among red rice individuals is extreme, with plants ranging from crop mimics to plants that look like wild rice, *Oryza rufipogon*.

Bayesian clustering analysis in conjunction with private allele-based analysis reveals evidence for mixed ancestry, with four distinct genetic groups co-occurring across the region.

Contrary to findings suggesting low diversity levels in red rice (Gross et al., 2009; Londo and Schaal, 2007), we find high levels of diversity in US red rice populations. Measures of genetic diversity, as measured by H_e, and allelic richness provide data indicating that US red rice is more diverse then *O. rufipogon*. This unexpected result suggests that red rice populations are acting as a genetic sponge incorporating alleles from multiple taxa (wild and cultivated), and potentially drawing diversity from multiple geographic locations. Private allele analysis corroborates this hypothesis, indicating that red rice populations contain *O. sativa indica*, *O. sativa japonica* and *O. rufipogon* private alleles.

There is no evidence of de-domestication in US population of red rice. *O. sativa japonica* cultivars are grown exclusively in the southern US rice growing region. If de-domestication were generating weedy rice biotypes, then we would expect paired cultivar/red rice groupings in our data. The data confirm genetic associations between red rice, *O. rufipogon* and *O. sativa*, but provides no evidence for the tight association of red rice and local US *O. sativa japonica* cultivars that we would expect to see under a de-domestication scenario. Our results provide strong support for a hybrid origin of red rice in Asia with subsequent introduction to the United States and no support for an exclusively de-domesticated origin.

With data placing the origin of US red rice in Asia, we looked to areas with known rice crop to wild rice gene flow for potential source populations. *Oryza rufipogon* populations in southern China are known to experience inter-specific gene flow with cultivated rice. The close proximity of *O. rufipogon* populations to areas of rice cultivation, the existence of red rice populations throughout the region, and a well documented history of biological exchange between the two countries offered a promising location to query for potential source populations.

The data show no direct link between Chinese and US red rice populations. The expectation that US populations, or a sub-set of individuals in these populations, are the result of a Chinese introduction would result in genetic overlap between the two regions; no such overlap is seen in the data. In fact the data show a striking lack of genetic similarity. Both Chinese and US populations have *O. rufipogon* private alleles, suggesting that hybrid origins play a role in the formation of red rice in both regions, but the US populations have many more *O. rufipogon* alleles than Chinese populations. This is unexpected, as we would assume recurrent hybrid red rice generation *in situ* would result in higher allele diversity than populations that form as the result of an introduced history as is the case in US red rice populations. A higher level of diversity in US populations suggests a role for multiple introductions from disparate geographic locations (Kolbe et al., 2007). Difference in diversity and the genomic lineages involved in the formation of global red rice populations suggest that the dynamics of weed generation and maintenance may differ by geographic region.

An intertwined evolutionary history complicates the management of these weedy pests. Modern eradication measures such as the application of chemical herbicides does not work in this system, because chemicals that kill red rice also kill cultivated types. The recent advent of herbicide resistant cultivars (Clearfield ^{TM}) offers the promise of rice cultivation free from red rice, but crop to weed gene flow in this system has the potential to destroy this utopic vision by generating herbicide resistant populations and potentially super-weeds.

Changes in mating system have the potential to drastically affect levels of diversity in red rice populations. High genetic diversity in red rice populations will contribute to their ability to respond to environmental change. Direct estimates of outcrossing rates in four populations of red rice, using paternity analysis, confirm that red rice individuals are exchanging genes in the field. Inter-specific gene flow was seen in one population, suggesting that the acquisition of agronomically important trait through crop to weed hybridization is a reality in this system. Outcrossing rates varied across populations and between maternal plants within the same population.

Gene flow among red rice individuals (intra-specific) is acting to directly increase genetic variability, potentially generating novel genetic types better suited to surviving the strong selective pressure against the weeds. Management strategies that act to limit genetic exchange between red rice individuals and between red rice and their crop congeners are critical in preventing the evolution of super-weedy red rice. Farmers in this area must be encouraged to eradicate heavily infested fields before red rice populations flower and set seed. Although a simple suggestion, this strategy flies in the face of the economic constraints driving these decisions, as heavily infested fields are often left 'to their own devices' until the end of the season when they are used as duck hunting grounds in an attempt to regain monetary losses.

The intertwined evolutionary history of red rice exemplifies the inextricable link between crops and their weeds. Clarifying the history and modern dynamics of red rice populations in the United States and around the globe will indicate that US populations are more diverse then their Chinese counterparts and China help us to deepen our understanding of the processes that influence weed evolution. Weeds and weedy systems provide a testing ground for the process of rapid and recurrent evolution, providing natural experiments which may provide insight into the roles of diversity, selection and plasticity in the generation and maintenance of specific phenotypes.

REFERENCES

De Wet J.M.J., Harlan J.R. (1975) Weeds and domesticates: evolution in the man-made habitat. Economic Botany 29:99-108.

Gao L., Chen W., Jiang W., Ge S., Hong D., Wang X. (2000) Genetic Erosion in Northern Marginal Population of the Common Wild Rice *Oryza Rufipogon* Griff. and its Conservation, Revealed by the Change of Population Genetic cstructure. Hereditas 133:47-53.

Gealy D.R., Saldain N.E., Talbert R.E. (2000) Emergence of red rice (*Oryza sativa*) ecotypes under dry-seeded rice (*Oryza sativa*) culture. Weed Technology 14:406-412.

Gealy D.R., Tai T.H., Sneller C.H. (2002) Identification of red rice, rice, and hybrid populations using microsatellite markers. Weed Science 50:333-339.

Gross B.L., Skare K.J., Olsen K.M. (2009) Novel Phr1 mutations and the evolution of phenol reaction variation in US weedy rice (*Oryza sativa*), New Phytol, England. pp. 842-50.

Kolbe J.J., Glor R.E., Schettino L.R., Lara A.D.A.C., Larson A., Losos J.B. (2007) Multiple sources, admixture, and genetic variation in introduced Anolis lizard populations. Conservation Biology 21:1612-1625.

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

Morishima H., Barbier P. (1990) Mating System and Genetic Structure of Natural Populations in Wild Rice *Oryza rufipogon*. Plant Species Biology 5:31-39.