SCIENTIFIC CAPACITY BUILDING: ENHANCING OUR UNDERSTANDING OF WILD RICE GERMPLASM AND HUMAN RESOURCE DEVELOPMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Janelle Kang Hui Jung

August 2016

SCIENTIFIC CAPACITY BUILDING:
ENHANCING OUR UNDERSTANDING OF WILD RICE GERMPLASM
AND HUMAN RESOURCE DEVELOPMENT

Janelle Kang Hui Jung, Ph. D.

Cornell University 2016

The scientific research process is complex and multifaceted, involving the integrated management of background information, experimental design, physical samples, and analytical tools through a research team's ability for planning, execution, analysis, and interpretation. Capacity building in the sciences encompasses the development of technological and experimental resources as well as human elements – a scientific mind for hypothesis generation and testing, broad exposure to various fields and people to acquire interdisciplinary perspective, and collaborations to expedite research progress and impact. This dissertation encompasses all these aspects of scientific capacity building, from the technical and material to the cultural and collaborative, within the microcosm of cultivated Asian rice, *Oryza sativa.*

Much is already known about the ecology, physiology, genetics and breeding of rice, an important staple cereal crop and key model organism. In contrast, knowledge about its wild ancestor, the *Oryza rufipogon* species complex (*ORSC*), remains elusive and inconclusive, despite the vast potential and urgent need of such knowledge to revolutionize the breeding of cultivated rice.

Chapter 1 (and manuscript in Appendix 1) reviews existing literature and explores the evolutionary history, habitat, anatomy, population structure and reproductive habits of the *ORSC*,

as well as the taxonomic confusion surrounding it. Results of this study identify five phylogeographically and genetically distinct subpopulations based on analysis of 286 *ORSC* accessions genotyped with 49 SSR, 41 MITE, and 29 SINE markers. Chapter 2 focuses on genome wide association studies (GWAS) on a wild rice diversity panel of 95 *ORSC* accessions, screened for morphological, reproductive, stress tolerant and root system architecture traits, while Chapter 3 is an indepth review of the genes and hormones involved in root system architecture.

Chapters 4 and 5 detail my work on capacity building. Chapter 4 provides a process and impact evaluation of the first five years of the *Rice:Research to Production* (R2P) course, developed to provide graduate students and young scientists with an intercultural, multidisciplinary, hands-on experience in rice cultivation and new perspectives on global food security. Chapter 5 provides a case study detailing the impact of similar international experiential learning opportunities in advancing cultural knowledge among science and health graduate students.

BIOGRAPHICAL SKETCH

Janelle Jung was born on April 20th, 1982 in Honolulu, Hawai'i, the daughter of William Jung and Lleander Choo Jung. She completed her undergraduate degree at Mount Holyoke College in South Hadley, MA, graduating in 2004 with a B.A. in Biology and English Literature. She started her graduate degree in Plant Breeding and Genetics at Cornell University in August 2007.

To my dad for planting the seed, and my mom for tending it.
To Amy, for coaxing it into bloom, and to Susan, for encouraging shoot and root growth in new directions. To all my friends and supporters, for being the sunshine and rain that sustains me.

I ulu no ka lālā i ke kumu.
The branches grow because of the trunk.

E kolo ana no ke ēwe i ke ēwe.
The rootlet will creep toward the rootlets.
Of the same origin, kinfolk will seek and love each other.

He poʻo ulu ko na mea kanu.
Plants have heads that grow again.
An assurance that if you break off the top of a plant, it will put forth a new one.

From "'Olelo No'eau: Hawaiian proverbs & poetical sayings"
Collected by Mary Kawena Pukui

ACKNOWLEDGMENTS

To my graduate advisor, Susan McCouch, for her support, guidance, and understanding, always

My graduate committee members, Margaret Smith, Tom Brutnell, Leon Kolchian, Rosemary Caffarella, and Noel Magor, for their support, direction, and patience

The funding agencies and PIs who have supported my graduate degree and research:

- CSREES - USDA, Agreement No. 2007-38420-17748 – PI: Elizabeth Earle
- NSF-PGRP awards #0606461 and #1026555 – PI: Susan McCouch
- NSF-EAPSI – Graduate summer internship for field research in China
- USDA – Hatch and Smith Lever Act Formula Funds – PI: Margaret Smith

My key research collaborators:

- Hyunjung Kim (*ORSC* population structure and genetic diversity)
- Randy Clark and James Jones-Rounds (3D-RSA and aluminum tolerance screens)
- Teresa Hancock, Daniel Wood, and Georgia Eizenga (DBNRRC wild panel growout and phenotyping)
- Ruaraidh Sackville-Hamilton, Ken McNally, and Ma. Elizabeth Naredo – (IRRI germplasm provision, wild panel growout and phenotyping)
- Juan-David Arbelaez (Wild diversity panel aluminum tolerance screen)
- Song Ge, Rong Liu, and Lei Huang (China wild diversity panel field growout and phenotyping)
- Rosemary Caffarella and Noel Magor (R2P course evaluation and cultural understanding case study)

Significant lab and analytical support and contributions: Hyunjung Kim, Anthony Greenberg, Mark Wright, Namrata Singh, Chih-Wei Tung, Yuxin Shi, Francisco Agosto-Perez, Pavel Korniliev, Jennifer Kimball, Wricha Tyagi, Lisa Polewczak, Genevieve DeClerke, Kazi Akther, Diane Wang, Shelina Gautama, Lyza Maron, Eric Craft, and Jennifer Spindel

Greenhouse and seed storage: Sandy Harrington, Fumio Onishi, many undergraduate interns, and the Guterman greenhouse crew

My friends and colleagues, whose research assistance, steady support, and unwavering encouragement have contributed greatly to my personal and professional development, seeing me through to the very end (and new beginnings): Hyunjung Kim, Randy Clark, James Keach, Gaganpreet Sidhu, Iman Zarei, Hsiang-Chun Lin, Maria Dwiyanti, Hedia Tnani, Inabat Seytnazarova, Moni Singh, Jason and Sarah Beebout, and Michael Joyce

With special thanks to Jeffrey Doyle and Elizabeth Earle for their care and good counsel

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 - POPULATION STRUCTURE AND GENETIC DIVERSITY IN THE *ORYZA* RUFIPOGON SPECIES COMPLEX, WILD ANCESTOR OF ASIAN CULTIVATED RICE

Note: A manuscript on which I am co-first author, summarizing much of the data presented in this chapter, was submitted to the journal *Rice* on July 6, 2016. My contributions to that paper include the following: development of the wild rice (*Oryza rufipogon* species complex, *ORSC)* diversity panel, phenotypic and genetic analyses for functional mutations at *BH4* and *Rc* genes, haplotype analysis for *Rc*, and significant contributions to the interpretation of the data and manuscript preparation. The submitted manuscript can be found in Appendix 1.

Introduction

Cultivated Asian rice, *Oryza sativa* L., is arguably the world's most important grain crop, as the staple starch for over half of the world's population. While advances in rice breeding, genetics, and now genomics over the past 60 years have enabled breeders to understand and utilize the broad genetic diversity of rice to significantly advance grain yield, quality, and tolerance to biotic and abiotic stress, the wild ancestor, referred to here as the *Oryza rufipogon* species complex (*ORSC*), remains a vast, undefined, and largely underutilized potential source of agronomically valuable traits and alleles. This chapter presents an overview of the taxonomic characteristics of the *ORSC* and hypotheses on its genetic, morphological, and phylogeographic structure and relation to *O. sativa*. A subset of 96 diverse *ORSC* accessions was selected from a larger collection to constitute a "wild *(ORSC)* diversity panel" for more in-depth evaluation.

Here, the panel was characterized for phenotypic and genotypic diversity and evaluated for population structure. Data from the panel provide a means for testing hypotheses about taxonomic classification and the relationship between the *ORSC* and domesticated *O. sativa* to inform a revised classification scheme for the wild species.

The *Oryza* genus

The *Oryza* genus, a small genus in the Poaceae family, includes 23 species with ten genomes (A-J) (Vaughan *et al.* 2003). For all *Oryza* species, the haploid chromosome number is 12 (n=12), but the genus includes both diploid (2n=24) and tetraploid (4n=48) species (Tateoka 1964b). The AA genome complex (2n=24), also called the *O. sativa* complex, is considered the primary genepool. It is composed of seven species, all of which are intermatable with varying degrees of cross-compatibility:

O. sativa L., Asian cultivated rice

O. glaberrima Steud., African cultivated rice

O. rufipogon sensu lacto, the ancestral species of *O. sativa*, native to South and Southeast Asia (*O. rufipogon sensu stricto,* Griff. for the perennial form; *O. nivara* Sharma *et* Shastry for the annual form)

O. barthii A. Chev., the ancestral species of *O. glaberrima*, native to West Africa

O. meridionalis Ng, an annual, primarily self-pollinating, species from seasonally inundated areas in North Australia

O. longistaminata, a perennial, rhizomatous, obligately outcrossing species from East Africa

O. glumaepatula, a Latin American self-pollinating species

For a review of these AA genome species and the other *Oryza* species, see Vaughan et al. (2003) and Vaughan (1994). For information on past synonyms used to describe the members of the *O. sativa* complex, see Vaughan (1989).

Morphological differences between *O. rufipogon* and *O. nivara*

O. rufipogon Griff. and sister species, *O. nivara* Sharma *et* Shastry, are the two most closely related wild species to *O. sativa* and are collectively regarded as its progenitor (Oka 1988; Khush 1997), referred to in this thesis as the *ORSC*. The annual and perennial forms are historically distinguished as separate species by differences in morphology, growth habit, reproductive cycle, and growing environment (Tateoka 1964a; Oka and Morishima 1967; Vaughan 1989, 1994). *O. rufipogon* is characterized as perennial, photoperiod insensitive, and highly outcrossing, though tending toward clonal reproduction with its prostrate growth habit and the proliferation of new ramets from stolon outgrowth. In contrast, *O. nivara* is considered to be annual, upright, photoperiod insensitive, and predominantly self-fertilized (Figure 1.1). A list of key traits found in the literature and historically used to differentiate *O. rufipogon* and *O. nivara* may be found in Table 1.1.

Figure 1.1 ***Oryza rufipogon*** **and** ***O. nivara*** **plants growing in natural environments** **A**, (L-R) *O. rufipogon* stand in a lake in Papua New Guinea, in a swamp in Bangladesh, and a representative panicle. **B**, (L-R), *O. nivara* in a dry pond in Cambodia, alongside a cultivated rice paddy in Nepal, and a representative panicle. (Photo credits: Top left, clockwise: Vaughan, 1994; http://www.knowledgebank.irri.org/extension/*Oryza*-rufipogon-griff.html; Vaughan, 1994; http://www.knowledgebank.irri.org/extension/*Oryza*-nivara-sharma-et-shastry.html; –bottom two right: http://www.gene.affrc.go.jp/databases-plant_images_detail_en.php?plno=3110390026)

Table 1.1 Distinguishing morphological and ecological traits between *O. rufipogon* (sensu stricto) and *O. nivara*

Species	*O. sativa*	*O. nivara*	*O. rufipogon* (*sensu stricto*)
Life habit[1]	Annual	Annual	Perennial
Habitat[2]	Dry and wetland fields, deepwater up to 4m, floating >4m	Seasonally dry, swamps, pond/stream/field banks	Deepwater/aquatic, swamps/marshes, ricefields, partially deepwater, prefers clay/loam soil and black soil
Geographic range[2]	Worldwide	Drier regions of S/SE Asia	Tropical Asia to Australia[2]
Photoperiod response	Sensitive	Usually insensitive	Sensitive
Plant type[2]		Semi-erect to decumbent	Decumbent or floating, tufted and spreading/scrambling
Lateral meristem formation/nodal tillering	Absent		Present
Horizontal stems	Absent	Absent	Present
Regen. Ability of stem segments[6,7]	Low-moderately high	Mostly low; low (0.07-0.33 avg. in 0-3 scale)	Mod-high; mod-high (1.51-2.50 avg. in 0-3 scale)
Plant height		Short to intermediate (usu. <2m)[2]; short (Avg. 84cm)[5]; Culm length: 127-151.7cm[7]	Tall, ~150cm Avg[5]; Culm length: 234-293cm[7]
Internodes			long
Ligule	1.71cm avg. Lgth[5]	Long; 1.19 cm avg. L[5]	2.07cm Avg. L[5]
Characteristics at end of growing season		All tillers are productive, all dried	Dried productive tillers, green tillers present that will flower next season
Rhizomes	None	None	
Roots			Perennial root stock, adventitious roots
Days to heading[7]		Shorter (112-145d Avg)	Longer (137-146d Avg.)
Panicle number[7]		Higher (10.4-14.5 Avg)	Lower (3.3-8.5 Avg)
Panicle length[5]	avg. 21.8cm	avg. 13.3cm	avg. 21.3cm
Panicle exsertion		Inserted or not well exserted/ partially exserted	Well exserted

Panicle shape	Erect, compact	Semi-open	Spreading, open
Panicle branching[5]	Secondary branching; Avg. 10.2 primary branches/panicle	Few secondary and tertiary branches; Avg. 5.06 primary branches/panicle	Avg. 7.2 primary branches/panicle
Spikelet dimensions	Usu. 4-8.5mm L, 2-4 mm W[2] ; Avg. 8.03mm L, 3.05mm W[5]	large - 6-8.4mm L, 1.9-3.0 mm W, 1.2-2.0mm thick, Avg.[2]; 8.14L, 2.56W Avg[5]	Usu. 8-9mm L[2]; Avg 8.13mm L, 2.27mmW[5]
Spikelets/ panicle	Avg.113.95	Avg.39.35	Avg.63.45
Spikelet fertility		High	May be low
Time between spikelet opening and pollen emission[6]	Short: Immediately-30s	Short: ~1-2min	Longer: ~2-6min
Awns	Short-none[2]; Avg. 0.72cm[5]	Long/strong (4-10cm)[2]; 6.91cm Avg[5]	Long - 5-11cm [2]; Avg.5.87[5]
Anthers	usu. <2.1mm L[2]; 2.51mm Avg.[5]	<2.5mm; immd. dehiscent, upright; 2.82mm Avg. L[5]; 2.10-2.21cm Avg.[7]	>3mm L to >7; indehiscent, pendant; 4.88mm Avg.L[5]; 4.79-5.07cm Avg.[7]
Embryo size	Usu <2.1mm long	Usu. 1-1.5mm long	Usu. 1-1.5mm long
Synchronicity of seed maturation	Synchronous	Asynchronous	Asynchronous
Shattering[2]	Non-shattering	Highly shattering	Highly shattering
Seed production	High	High	Low
Seed dormancy	Low[6]	Mod-high[6], strong[2]	Mod-mod high [6]

[1] Vaughan, DA, Morishima, H., and Kadowaki, K. (2003). Diversity in the *Oryza* genus. Current Opinion in Plant Biology 6, 139–146.

[2] Vaughan DA. 1994. The Wild Relatives of Rice: A Genetic Resources Handbook, IRRI, Philippines

[3] Grillo et al. 2009. Genetic Architecture For the Adaptive Origin of Annual Wild Rice, *O. nivara* (Individuals in study chosen based on characteristics displayed under greenhouse growing conditions.)

[4] Li, Zhou, and Sang. 2005. Genetic analysis of rice domestication syndrome with the wild annual species, *O. nivara*..

[5] Morishima, H, Oka HI, & Chang, WT, 1961. Directions of differentiation in populations of wild rice, *Oryza perennis* and *O. sativa f. spontanea.* Evolution 15: 326-339.

O. perennis* traits entered here as *O. rufipogon* and *O. sativa f. spontanea* as *O. nivara

[6] Morishima and Oka, 1965. Variations in the breeding system of a wild rice *O. perennis*

O. perennis* (Asian race); *perennis* type entered as *O. rufipogon* and *O. perennis* Asian race, *spontanea* type entered as *O. nivara

[7] Barbier, 1989. Genetic variation and ecotypic differentiation in the wild rice species *Oryza* rufipogon. I. Population differentiation in life-history traits and isozymic loci.

Habitat preference and geographic range

The two species also prefer different growth habitats: *O. rufipogon* is predominantly aquatic and found in areas with year-round standing water, such as lakes, swamps, river beds, and marshes, while *O. nivara* is found in seasonally wet habitats such as lake shores and river banks that may be dry most of the year but undergo flooding with monsoon rains, as well as in disturbed areas like roadside ditches and in or around the edges of paddy fields. The two species may therefore coexist in the same area with populations of both occurring along a hydrologic cline (ex. *O. nivara* on river banks and *O. rufipogon*-extending into deeper water). On a larger geographic scale, both the species share the same geographic range across Continental South and Southeast Asia, though the range of *O. rufipogon* also extends further south into the Indonesian archipelago, Papua New Guinea, and possibly also the tips of northern Australia (Figure 1.2).

The current geographic ranges of the two species were shaped by the evolunary forces of climate change imposed on the prehistoric populations of their common ancestor species during the past 10,000-20,000 years. From archeobotanical data-based reconstructions of vegetation maps, Fuller ((Fuller *et al.* 2010) hypothesizes that during the Last Glacial Maximum (LGM), the cold, dry glacial period of the Pleoistocene (~20,000 years before present (bp)), the ranges of ancestral wild rice populations were limited to the southernmost part of the Indian subcontinent, including Sri Lanka, and a large swathe of continental Southeast Asia, extending down to the then-interconncted northern Indonesian archipelagic region. The gradual shift in climate to a warmer, wetter cycle during the Early Holocene (~9000 years bp) allowed wild rice populations to expand northwards, but rising sea levels also cut off the southernmost ranges by inundating landbridges, creating islands of reproductively isolated populations (reviewed in Fuller et al., 2010; Figure 1.2).

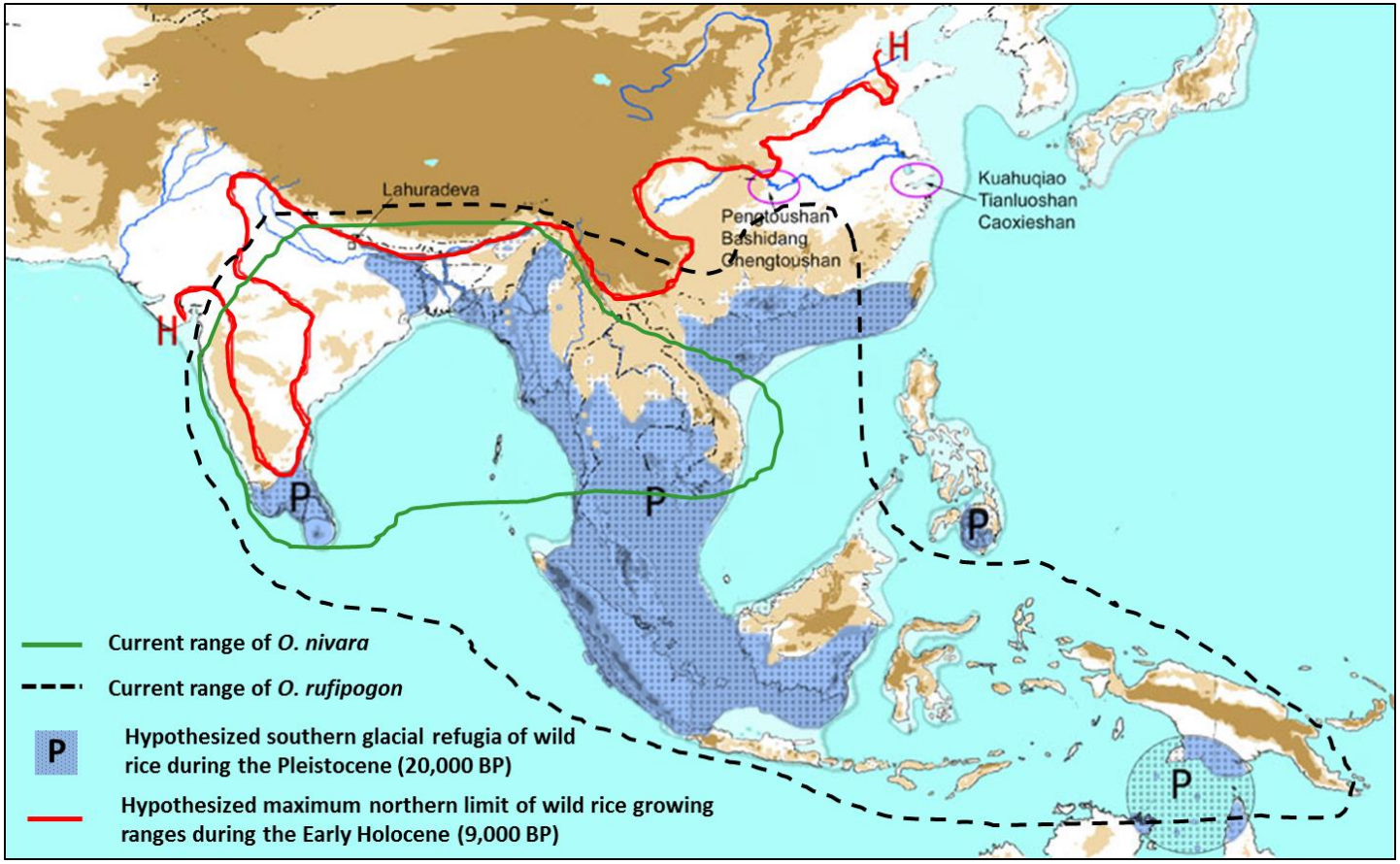


Figure 1.2 Map of the prehistoric and current ranges of *ORSC* populations. During the glacial period of the Pleistocene ancestral wild rice populations were limited to a southern ecological range. A gradual shift in climate to a warmer, wetter cycle during the Holocene allowed wild rice populations to expand northwards, but rising sea levels also isolated southernmost ranges by inundating landbridges (Fuller *et al.* 2010) The current estimated range of *O. nivara* is limited to continental South and Southeast Asia, whereas the currant extant range of *O. rufipogon* still extends into archipelagic Southeast Asia (Vaughan *et al.* 2003) (Modified from Fuller et al., 2010).

Molecular genetic studies on *O. rufipogon* and *O. nivara*

Many molecular genetic studies have been conducted on *O. rufipogon* and *O. nivara* to try and elucidate their relationship to each other and to *O. sativa* (Supplementary Table 1.1). These studies have utilized isozymes (Second 1982; Cai *et al.* 2004), RFLPs (Cai and Morishima 2002; Lu *et al.* 2002; Cai *et al.* 2004), RAPDs (Ge *et al.* 1999; Ren *et al.* 2003), a variety of repetitive element markers including SINEs (Cheng *et al.* 2003), MITEs (Park *et al.* 2003; Ge *et al.* 2005), and SSRs (Ren *et al.* 2003; Shishido *et al.* 2006; Li *et al.* 2006b; Lu *et al.* 2008), intron sequences(Barbier *et al.* 1991; Zhao *et al.* 2009), chloroplast markers (Dally and Second 1990; Ishii *et al.* 2001; Ge *et al.* 2002; Guo and Ge 2005; Xu *et al.* 2010), and most recently, SNPs from whole genome resequencing data ((Xu *et al.* 2012; Huang *et al.* 2012c). The broad applicability, if not the veracity, of the results of most studies on population structure and genetic diversity among the two wild species has been limited by germplasm bias (single species/ecotype, small sample number, single country/limited geographic range) and often also genetic bias (low marker number, short genetic region, single/multi gene study); however, most studies show little or no interspecific genetic differentiation (Supplementary Table 1.1).

Persistent confusion on species names

A clear understanding of what exactly distinguishes *O. rufipogon* from *O. nivara* as distinct species remains elusive. There are evidently two groups (plus a range of intermediate morphotypes) that differ in life habit, reproductive habit, and ecological adaptation, but they share partially overlapping habitats and geographic ranges, are cross-compatible, and genetically undifferentiatable. The reason for this confusion is partly historic, partly technological, and wholly semiotic. Nomenclature, particularly species-specific binomial nomenclature is a human

attempt to connect two partially-overlapping groups— (1) a biologically-defined "species" (organisms which can intermate and have fertile progeny), and (2) a taxonomic cluster based on 1) morphological traits compared to type specimens (alpha taxonomy), 2) genetic diversity (molecular systematics), or 3) genetic-identity-based evolutionary lineages of morphologically-defined taxa (cladistics/phylogenetics), and 4) the improper consideration of non-standard species characteristics such as ecological niches or reproductive habits, as in the case of *O. rufipogon* and *O. nivara*.

When much of the *Oryza* genus was being characterized and accessions collected for genebank deposition in the 1950s and 1960s, different species names were conferred on populations of *Oryza* spp. around the world, based on nonstandardized, morphological, geographical, or ecologically-based species definitions and boundaries that varied between individual researchers and collectors. *O. perennis* was the common species name used by some researchers to refer to wild rice populations across Asia, Africa, Australia, Oceania, and South America that were largely perennial and share several different morphological or developmental characteristics (Morishima *et al.* 1961; Oka and Morishima 1967; Morishima 1969); however, *O. perennis* has since been divided into four separate AA genome species based on geography and supported bygenetic differences and : Asian (*O. rufipogon/O. nivara*), African (*O. longistaminata* Chev *et* Roehr), Oceanian/Australian (*O. meridionalis* Ng), and South American (*O. glumaepatula* (for review, see Vaughan et al., 2003). In a 1974 study on ecological and morphological traits which may distinguish various *Oryza sativa* complex species, long after much of the initial germplasm collection and characterization had been completed in the 1960s, Oka estimates that over 60 different Latin names have been given to specimens in this complex (Oka 1974)

Life habit—the annual or perennial designation used as a principle means of differentiating individuals of *O. nivara* from *O. rufipogon*—is a key example of the information that was likely not noted directly at the time of seed collection in natural environments. A true score of life habit would theoretically require at least two, if not several, separate observations over time to determine whether an individual was able to survive, grow, and reproduce for more than one growing season or reproductive cycle. Life habit is a complex trait dependent on several other developmental and morphological characteristics and evidence of annuality or perenniality in the field was likely deduced through other secondary traits and observations, such as green leaf area at seed maturity, flowering determinancy and synchronicity, presence or absence of stolons, and hydrology of the growing environment.

While some researchers differentiated *O. rufipogon* (*sensu stricto*) as the "perennial" species or form and *O. nivara* as the"annual" species or form, others recognized both annual and perennial ecotypes as part of a common species, *O. rufipogon* (*sensu lacto*). Several scientists have noted the presence of 'intermediate' ecotype (Morishima *et al.* 1961, 1984; Vaughan *et al.* 2003), or an annual-perennial continnum (Sano *et al.* 1980), underscoring the fact that there are populations of plants that defy strict categorization and that it is problematic to categorize all wild individuals as either annual or perennial, and to presume that nomenclature is based on or predictive of life habit (e.g. not all accessions called *O. nivara* are strictly annuals and not all accessions called *O. rufipogon* are strictly perennial).

Weedy rice vs. wild rice relatives

Weedy rice or "red rice" is named for one of its most common and agronomically undesirable traits—a red or dark colored pericarp that is difficult to fully polish off. It is found in close

association with the cultigen, growing in or at the edges of rice fields, as well as in swampy or marshy areas (Oka 1974; Chang 1976), and exhibit many agronomically undesirable traits such as early and non-synchronous maturity, seed shattering and dormancy, awns, and red or dark hull or pericarp color, and possibly also an assortment of domesticated characteristics, such as white or light colored hulls and pericarp and awnless spikelets (Diarra *et al.* 1985; Noldin *et al.* 1999). Weedy rice commonly considered to refer to natural hybrids between wild and cultivated *Oryza* species, likely as a result of cultivated introgression into wild, outcrossing species (Langevin *et al.* 1990; Gealy *et al.* 2003; Cao *et al.* 2006). Weedy rice may sometimes be referred to by its own taxonomic name*: O. sativa* L. F. s*pontanea* Roschev. (formerly *O. sativa* var. *fatua* Prain), in which case the core literature on the *O. sativa* complex considers it an annual of mixed *O. rufipogon/O. nivara* and *O. sativa* ancestry with a natural geographic range limited to Asia (Morishima *et al.* 1961; Chang 1976; Vaughan *et al.* 2001; Sharma 2003).

In several research articles and databases, the subspecific designation has been erroneously dropped, leading to accession designations of *"*O. *spontanea*" (Fuller et al., 2010; Lu et al., 2008; Vaughan, 1994; http://www.irgcis.irri.org:81/grc/IRGCISHome.html). Conflicting literature also identifies as a completely wild *O. spontanea* as perennial and equated with *O. perennis* and *O. rufipogon* (Hill 2010), or annual and synonymous with *O. nivara* (Oka 1974; Sano *et al.* 1980; Vaughan 1989). Genetic studies have shown *O. spontanea* and weedy rice accessions, even those introduced and naturalized in non-native cultivated areas such as in the US, which has no endemic *Oryza* species, to be of varying degrees of mixed *O. rufipogon*/*O. nivara* and *O. sativa* ancestry, and introgression signatures suggested both recent and ancient introgression of *O. sativa* alleles into wild relatives (Tang and Morishima 1997; Vaughan *et al.* 2001; Yu *et al.* 2005; Cao *et al.* 2006; Londo and Schaal 2007; Lawton-Rauh and Burgos 2010;

Jiang *et al.* 2012).

These findings emphasize genetic support for morphological evidence that outcrossing between wild populations and cultivated landraces or varieties is a naturally occurring phenomenon from the domestication process to the present day as well as a driver of genetic diversity within subgroups of the species complex, lending support for the inclusion of accessions labeled as *O. spontanea* or weedy accessions thought to be of interspecific wild and cultivated origin as part of wild species diversity panels.

Germplasm collections of *O. rufipogon*

To this day, wide differences in the nomenclature and trait-based definition of *O. rufipogon* persist in rice genebank databases and research labs around the world. Case in point, the two largest, publicly-available germplasm repositories of *O. rufipogon*, which hold in common hundreds of duplicate accessions, have distinctly different naming schema inherited from the accessions and definitions of the researchers whose deposited accessions formed the core collections of that genebank. The online database of the International Rice Genebank Collection Information System (IRGCIS) (http://www.irgcis.irri.org:81/grc/IRGCISHome.html), housed at the International Rice Research Institute (IRRI) has 3098 accessions listed in its database as '*O. rufipogon*,' '*O. nivara,'* '*O. spontanea,'* all possible pairwise combinations thereof, or all possible pairwise combinations of the previous three species names with *O. sativa* (ex. '*O. nivara/O. sativa'* and '*O. sativa/O.nivara'*), with no indication as to how these were originally defined or differ from each other. In contrast, the online *Oryza*base wild strain database of the Japanese National Institute of Genetics acknowledges confusion in the wild rice taxonomy and nomenclature and the presence of an *O. nivara/O.rufipogon sensu stricto* distinction based on

ecology and life habit, but clearly states its use of the *sensu lacto* classification of *O. rufipogon* as a single species with a continuous range of annual, perennial, and intermediate types as held by core collectors Morishima and colleages (Morishima *et al.* 1961). There are 651 accessions in this database designated as '*O. rufipogon'* with additional, incomplete information on former species designations, such as '*O. perennis* (*O. nivara),*' '*O. perennis*,' or '*O. sativa* f. *spontanea,*', and life habit designations, such as annual or perennial.

Genebank accessions are composed of partially overlapping sets of individuals or populations that have been distinguished or grouped based on collector/evaluator perceived similarities or differences in growth environment, habitat, and morphological characteristics, but their nomenclature is inconsistent and incompletely correlated with the genetics, ecology, morphology, and/or reproductive biology of the accessions or populations characterized. While the species names of these accessions have persisted largely unchanged through the decades, there has been a paradigm shift in biology from morphology-based species definition to genetic-identity-based species definition, but no major subsequent attempt at reclassifying the wild ancestor under a single species name, or demoting the groups from species to ecotypes based on the lack of any documented genetic differentiation. This leads to circular arguments and erroneous conclusions when limited panels, or even single accessions characterized in genebanks as *O. rufipogon* or *O. nivara* are assumed to be satisfactory representatives of the species or species complex, and used to interpret phylogenetic studies. Decades of conflicting studies have engendered an underlying and often unacknowledged confusion which permeates and divides the rice community today. Hotly-debated conclusions inferred from genetic and genomic research claim support for either single- or multiple-origin theories of rice domestication, but these theories are often built on a foundation of poorly-examined and often ambiguously classified

germplasm that is used to anchor the study (Supplementary Table 1.2).

Since the relationship between genetics, morphology, ecology, and nomenclature of the wild ancestor of *O. sativa* is yet unclear, in this study we will henceforth refer to *O. rufipogon* and *O. nivara* jointly as the '*O. rufipogon* species complex, or *ORSC*'. Individual species names (i.e. *O. rufipogon* (*sensu stricto*) and *O. nivara*) will be used to refer to original genebank designations.

The purpose of this study was to: (a) evaluate genotypic variation in a large panel of *ORSC* accessions as the basis for characterizing population structure and its relationship to geography, ecology, and morphology, (b) evaluate phenotypic variation as the basis for associating genotype and phenotype, (c) test hypotheses about the ancestral origins of *O. sativa,* and (d) use the information gained on phenotypic and genetic diversity in our *ORSC* to construct an immortal "wild diversity panel" consisting of 95 purified accessions selected to represent the genetic, geographic, and morphological diversity of the species complex across Asia.

Materials and Methods

Germplasm selection

A diverse set of 317 *ORSC* accessions were selected to represent the entire geographic range and the genetic and morphological diversity of the species complex (Supplementary Table 1.3). Seeds of these accessions were imported from the International Rice Genebank Collection (IRGC; IRRI, the Philippines). These accessions included those characterized by the IRGC as: *O. rufipogon, O. nivara, O. spontanea, O. rufipogon/O. nivara, O. rufipogon/O.sativa,* and *O. nivara/O. sativa,* as we believed all the accessions classified by these designations to be representative of the *ORSC*. Initial selection of accessions for importation from the 3,098 *ORSC*

accessions publicly available in the IRGC was advised by Ruaraidh Sackville-Hamilton.

To conduct comparative population structure analyses with the cultivated species, 50 *O. sativa* cultivars consisting of ten accessions from each of the five cultivated rice subpopulations: *temperate japonica, tropical japonica, indica, aromatic*, and *aus*, as well as one outgroup *O. officinalis* accession (IRGC 105220) were also included in the genotyping panel.

Of the 317 accessions imported from IRRI, 287 were successfully germinated and grown to seed under greenhouse conditions. Along with 45 of the 50 *O. sativa* cultivars, and a single purified sample of *O. officinalis* a subset of 178 *ORSC* accessions were satisfactorily genotyped using three types of molecular markers, SSR, MITE, and SINE, as documented in Supplementary Table 1.3). A map showing the geographic distribution of all genotyped lines may be found in Figure 1.3.

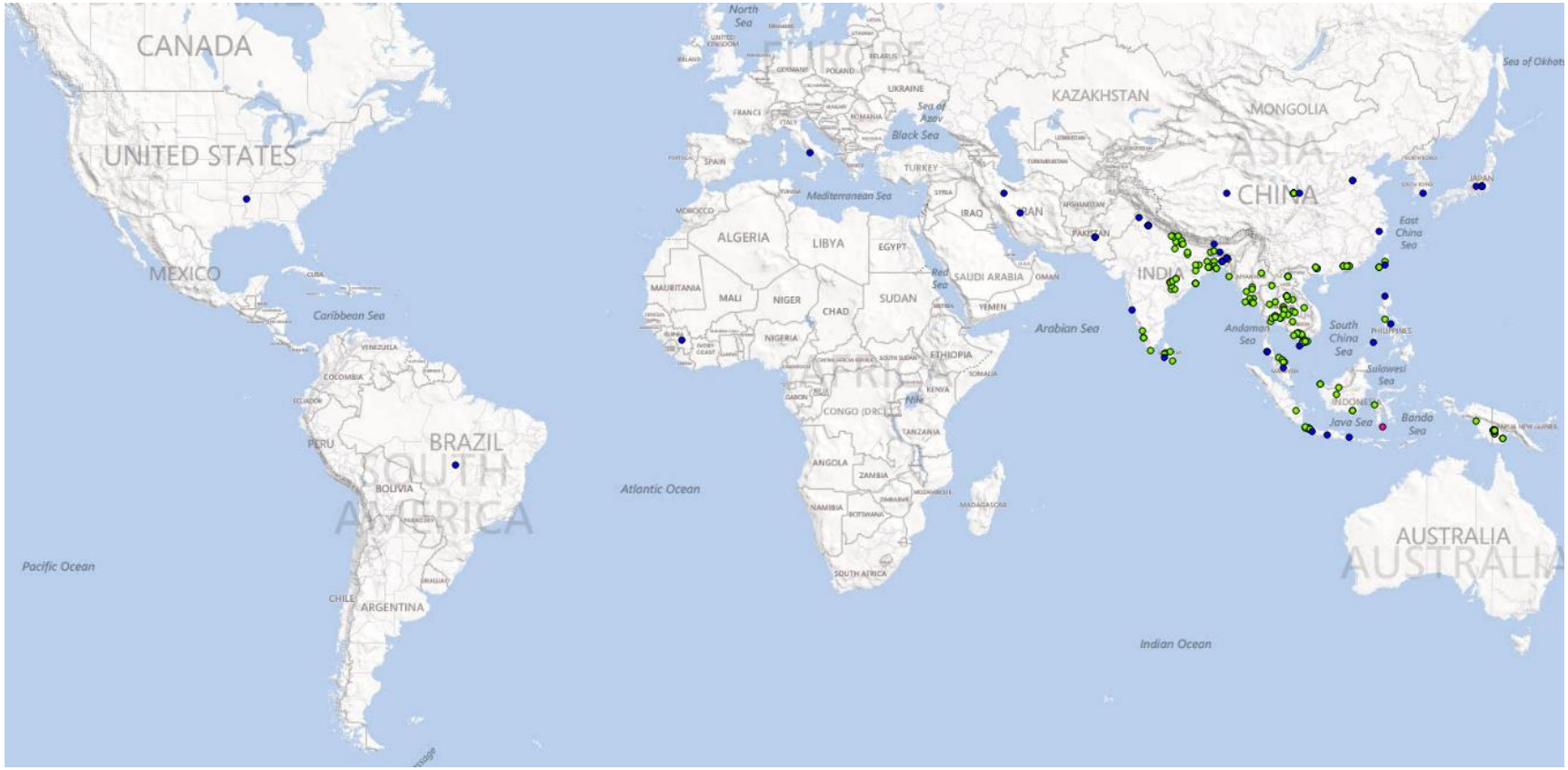


Figure 1.3 Geographic map of the 178 *ORSC* accessions, 45 O. sativa accessions, and one O. officianalis accession genotyped in this study. (*O. sativa* accessions for which genebank coordinates were unavailable (See SupplementaryTable 1.3) were graphed using derived latitude and longitude coordinates from the accession country of origin.)

Growout of accessions

Of the 317 *ORSC* accessions, 287 were successfully grown out under greenhouse conditions (85°F day/75°F night, 14-hour light cycle; Guterman Bioclimatic Laboratory and Greenhouse Complex, Ithaca, NY). Source seed (S0) was planted out using the following method: three seeds of each accession were sterilized with a 30-second soak in 70% EtOH, followed by 15-minute soak in a 20% bleach solution, and three washes with ddH2O. Seeds were planted individually in 6-inch clay pots, 1 cm below soil surface in damp Cornell mix. Accesions were ordered into three replicate and were positioned in growth tanks with uniform spacing and in random order with 35 pots per tank. Throughout the grow-out period, standing water in tanks was maintained at a constant depth of 8-10cm.

Because seed imports from the IRGC came in two shipments and non-germinating accessions had to be repeatedly planted, there were two overlapping growouts of material with multiple planting dates. The first growout of 182 unique accessions was planted on October 11, 2006; an additional 5 accesions were planted on May 1, 2007; 14 accessions (4 repeated) were planted on June 5, 2007, and the final 91 (1 repeated) accessions on September 25, 2007. Panicles from each individual plant were covered with glassine or waxed paper pollination bags prior to stigma exsertion to prevent cross-pollination and facilitate the collection of selfed-pollinated (S1) seed. To facilitate to creation of homogeneous single seed descent (SSD) lines of the wild accessions, selfed seed from each individual source plant was kept separated, as opposed to being bulked for all representatives of a single accession, as source genebank seed was heterogeneous and showed phenotypic segregation.

Phenotyping of accessions

All surviving individuals of the three replicates representing each of the 287 accessions were phenotyped for 12 developmental or morphological traits: days to heading (DTH), plant type – a categorical score combining culm angle and stolon presence, plant height, tiller number, panicle number, panicle length, primary branches on panicles, flag leaf length, flag leaf width, awn presence and length, hull color, and pericarp color. Stolon absence or presence was later derived from plant type scores, for a total of 13 phenotypic traits. However, data on all 13 traits is incomplete for most individuals. Trait ontology and phenotyping methodology may be found in Supplementary Table 2.2.

Initial growout for DNA isolation and genotyping

Initial growout of all accessions for genotyping was conducted by members of the McCouch lab, Jennifer Kimball, Wricha Tyagi, and Lisa Polewczak in the Guterman Greenhouse.

DNA isolation was conducted by Jennifer Kimball and Wricha Tyagi. Young leaf tissue was harvested from a single plant of each line and DNA was extracted using a modified potassium acetate-SDS protocol (Dellaporta *et al.* 1983).

Genetic marker selection and genotyping

Three markers sets consisting of 49 SSR markers, 41 MITE markers, and 29 SINE markers were chosen for genetic diversity analysis.

SSR marker identification and primer design were done as previously described in research by Coburn et al (Coburn *et al.* 2002). SSR genotyping by PCR and electrophoresis was performed as per Garris et al (2005). SSR marker and primer information may be found in Supplementary Table 1.4.

MITE markers were developed for *mPing* elements identified by BLAST searching the *Nipponbare* and *93-11* TIGR psuedomolecule (V 2.0) rice genomic sequence. Most markers were based on complete MITE insertions into the *Nipponbare* genome, with additional MITE-associated indels found by manually searching for MITEs in the *93-11* sequence, followed by back alignment to the *Nipponbare* reference genome. Primer 3 was used to design primers which were then validated *in silico*. A total of 41 MITEs were selected based on satisfactory PCR amplification and electrophoresis on 2% agarose gels for optimal segregation analysis on a reference panel of 48 *O. sativa* accessions (Diane Wang, Cornell University, personal communication). MITE primer information can be found in Supplemental Table 1.5.

SINE markers used in this study were chosen from a set of 47 members of *pSINE1* previously identified in the *O. sativa* complex (Cheng et al., 2003; Motohashi et al., 1997, personal communication, Dr. Suguru Tsuchimoto, University of Tokyo). Primer information on the 29 SINE markers with satisfactory results on all the accessions genotyped in this analysis can be found in Supplemental Table 1.6. SINE genotyping was performed as per Cheng et al (2003).

SSR marker development and primer design was as described in Coburn et al. (2002). MITE marker primers were designed by Greg Wilson and MITE genotyping was conducted by Wricha Tyagi and Diane Wang. SINE marker primer validation and genotyping was conducted by Hyunjung Kim. SSR genotyping was conducted by Wricha Tyagi, Jennifer Kimball, and Jung-Wook Chung, SINE marker genotyping was conducted by Shelina Gautama and Hyunjung Kim, and MITE genotyping was conducted by Diane Wang.

Population structure estimation

Population structure in the *ORSC* was investigated using the Bayesian, model-based clustering

algorithm implemented in the software package STRUCTURE (Pritchard et al. 2000). Individuals were assigned to K population genetic clusters based on their combined multi-locus SSR, MITE, and SINE genotypes, and their membership coefficients in each cluster for each individual accession were estimated.

For each analysis of K = 1-10, 10 runs were completed with 500,000 MCMC iterations preceeded by an intitial burn-in of 100,000 MCMC iterations. The DeltaK method, an ad hoc statistic described in Evanno et al. (2005), was used to identify the most significant number of clusters through the online program, Structure Harvester (Earl and VonHoldt 2012). Using the downstream program CLUMPP (Jakobsson and Rosenberg 2007), the probability values for each cluster were averaged across 10 runs. In order to determine how population structure may differ in the *ORSC* when grouped independently or together with *O. sativa*, STRUCTURE analysis as described above was conducted on the 178 *ORSC* accessions alone, and on the complete set of 178 *ORSC*, 50 *O. sativa*, and single outgroup accessions. STRUCTURE and CLUMPP analyses were completed as multiple tasks through Computational Biology Service Unit (CBSU) at Cornell University. All results from Structure were visualized using Distruct (Rosenberg 2003).

An initial population structure estimation using the set of 287 *ORSC* accessions and 50 *O. sativa* described previously, genotyped with the 49 SSR markers was performed by Wricha Tyagi, Jung Wook Chung, and Keyan Zhao. Population structure analysis detailed in this chapter was done by Hyunjung Kim and Janelle Jung.

Genetic distance estimation

The tree-based method using 3 marker characters was used to estimate genetic distance using an unrooted neighbor-joining (NJ) algorithm (Saitou and Nei 1987) implemented in PowerMarker

V3.23 and visualized in Mega5 (Tamura *et al.* 2011).

Phenotypic data analysis

Raw line means for all traits were calculated and the statistical software program JMP Pro 10 (2012) was used to calculate means, standard deviations and oneway ANOVA analyses for traits grouped according to species, at cluster membership at K=4 (*ORSC* only analysis), and geographic region. Geographic region designation was assigned according to accession country of origin, with India, Sri Lanka, and Nepal assigned as West Asia; Myanmar, Vietnam, Cambodia, Laos, and Thailand as continental Southeast Asia; China and Taiwan as East Asia; and Malaysia, Indonesia, the Philippines, and Papua New Guinea as Archipelagic Southeast Asia.

In order to simplify the analyses testing whether *O. rufipogon* and *O. nivara*-type individuals could be differentiated by variation in morphological and developmental trait ranges, only accessions designated as either *O. rufipogon* (n=174) or *O. nivara* (n=66) were included in these analyses. Excluded accessions included 19 characterized as *O. spontanea*, 17 *O. rufipogon/O.nivara*, and 3 *O. sativa/O. rufipogon* accessions. The number of accessions represented in each analysis differed according to the availability of phenotypic and or genotypic data for that particular analysis. Phenotypic trait data highlighted in this study was also limited to the seven observed traits that matched traits shown in the literature to differentiate *O. nivara* from *O. rufipogon* (*sensu stricto*) individuals. These seven traits were: days to heading, panicle length, panicle branch number, panicle length, plant height, culm angle, and stolon absence or presence. Line means for all accessions and all traits are provided in Supplementary Table 1.7.

To allow for missing phenotype data and to maximize marginal likelihoods based upon error

contrasts, the restricted maximum likelihood (REML) method was used to estimate correlations between subpopulations designated at K=4 (60% membership admixture cutoff) and the seven aforementioned traits associated in the literature with distinguishing *O. rufipogon* from *O. nivara*.

Results

The 283 *ORSC* accessions that successfully germinated and produced seed under greenhouse conditions in Ithaca were grown out and phenotyped for 13 morphological and developmental traits. Of these, a subset of 178 *ORSC* accessions were genotyped with 41 MITE markers, 49 SSR markers, and 29 SINE markers. Using the combined marker set, accessions were clustered by multilocus genotype, with and without the 45 diverse *O. sativa* varieties and one *O. officianalis* outgroup. All had been genotyped with the same markers, providing a way to determine intraspecific and interspecific levels of population structure and genetic diversity.

Population structure in *ORSC*

Bayesian clustering analysis of 178 *ORSC* accessions alone suggested either two or four genetic groups as the most significant number of clusters according to the DeltaK values (Figure 1.4c). At K=2, a small cluster of accessions, designated as 'W1' for 'Wild 1' (purple), separated from the majority group of accessions (W2, black) (Figure 1.4a). The vast majority (17/19) of these W1 accessions were from Papua New Guinea and all were genebank-designated as *O. rufipogon*, indicative of a geographically-isolated, genetically distinct Papua New Guinea subpopulation (Figure 1.5). At K=4, a third, almost-genetically pure group of accessions (W3, green), mostly *O. nivara*-designated and from Nepal and India (13/17) was distinguished, as well as a fourth group (W4, blue) that separated from the W2 genetic background. Eleven of the fourteen

accessions with over 60% membership in the W4 cluster were originally collected in China or Taiwan, and all had significant levels of genetic admixture with W2 and W3.

When clustering analysis was performed on the 178 O rufipogon complex together with the 45 *O. sativa* accessions, K=2 was identified as the number of clusters with the most significance (Figure 1.4d). At this level, the two clusters correspond to the deeply differentiated subpopulation groups within *O. sativa*, *aus* and *indica,* grouping almost all of the *ORSC* accessions with the *aus*/*indica*-like group (yellow) and distinguishing those accessions from the cluster of *aromatic/temperate japonica/tropical japonica* accessions (blue, Figure 1.4b), At K=3, an independent *O. rufipogon* group (black) is distinguished, while at K=4 and K=5, the W1 and W3 clusters, respectively, are resolved in the wild germplasm.

The unrooted dendrogram showing the relationship by genetic distance between the *ORSC* and *O. sativa* accessions underscores the shallow population structure apparent in the wild ancestor complex as compared with the highly segregated population substructure in the cultigen (Figure 1.5). As also shown in the clustering analysis with *O. sativa* at K=2 (Figure 1.4b),, the majority of the *ORSC* accessions in the panel cluster with the *O. sativa indica* or *aus* subpopulations. *O. sativa aus* and *indica* varieties clearly branch out from two independent clusters of *O. rufipogon*; however, only a single *ORSC* accession, originating in Taiwan, clusters closely with the *O. sativa temperate japonica* subpopulation. These results suggest: 1) that there are groups of *ORSC* accessions present that have a close genetic relationship to different subpopulations of *O. sativa,* 2) that these groups of the *ORSC* were likely predifferentiated into *indica*-like, and *japonica*-like genetic groups prior to the beginning of *O. sativa* domestication.

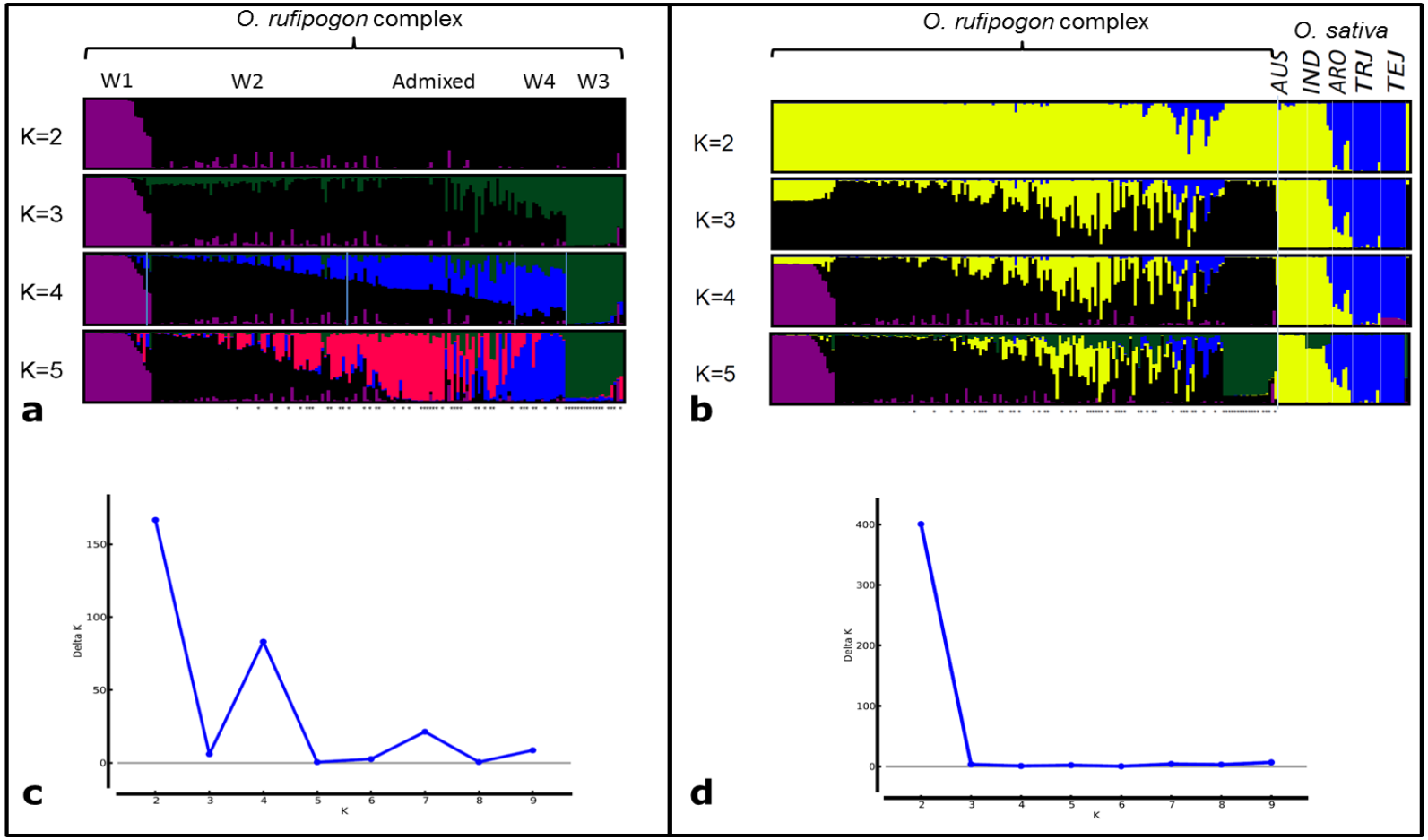


Figure 1.4 Estimates of population structure in the *ORSC* clustered alone (**a, c**) or with *O. sativa* (**b, d)**, using combined MITE, SINE, and SSR genotypes. In the bar plots (**a, b**), the genetic identity of each accession is represented by a vertical line colored according to the proportion of the mean cluster membership coefficients for that accession. W1-W4 and admixed cluster designations pertain to 60% admixture cutoff (light blue lines) at the K=4 level. *ORSC* accessions are in the same order for both **a** and **b.** Graphs of mean ΔK values (**c, d**) show the assignment of individual genotypes to different *k* specified sub-populations using the clustering method of the program STRUCTURE. Peaks in ΔK values indicate the most significant number of clusters in the *ORSC* are at K=2 and K=4, or at K=2 when the *ORSC* and *O. sativa* accessions are analyzed together.

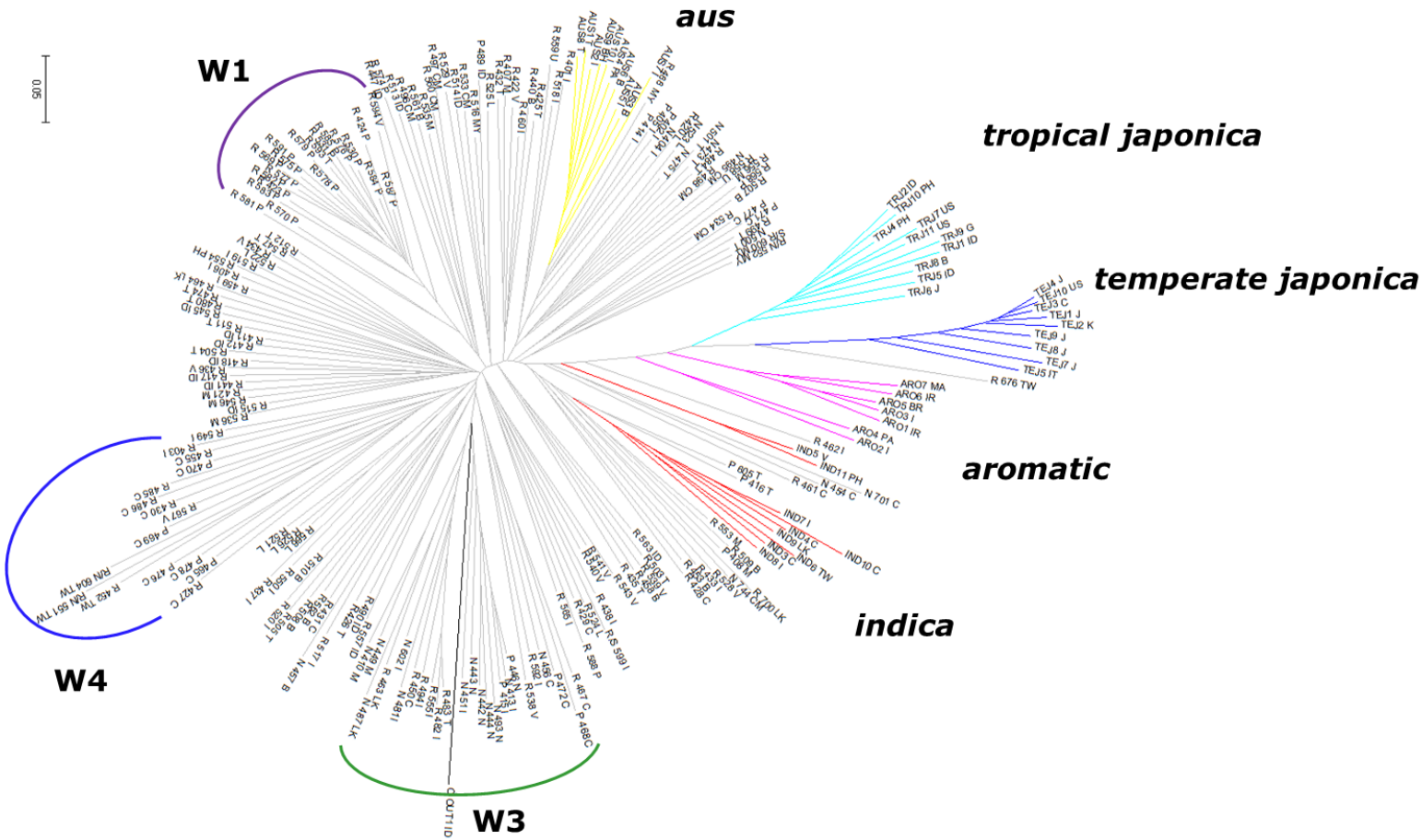


Figure 1.5 Unrooted dendrogram of 178 *ORSC* and 45 O. sativa accessions - *ORSC* (black), *O. sativa* accessions (colored according to subpopulation), and one *O. officinalis* accession (black, bold line), based on combined SSR, MITE, and SINE marker data. *ORSC* accessions are labeled with a single letter species identifier, a 3-digit internal accession number, and a country of origin abbreviation. *ORSC* W1, W3, and W4 clusters, identified at the K=4 level, are also shown. *O. sativa* accessions are labeled with a three-letter subpopulation abbreviation, a single-digit internal accession number, and a country of origin abbreviation. Genetic distance was estimated based on the chord distance of Cavalli-Sforza and Edwards (1967).

Morphological variation, correlation with nomenclature, geography, and genetics

Greenhouse growout of the 283 accessions in this study shows a wide range of variation in morphological characteristics at both the whole-plant gross morphology level (Figure 1.6), and with respect to individually measured traits (Figure 1.7). Thirteen morphological and developmental traits were measured on each line. For the 7 traits measured which were reported in the literature to distinguish *O. rufipogon* from *O. nivara* type plants, only three: plant height, days to heading, and stolon absence or presence, show significant variation at $p<0.05$ by species designation for the *O. rufipogon* (n=174) and *O. nivara* (n=66) accessions in this study (Figure 1.7). The variation observed on greenhouse grown plants is largely inconsistent with the trait values and the range of variation reported in the literature as distinguishing *O. rufipogon* from *O. nivara* (Table 1.1); however this may be due to differences in growth environment, as phenotypic data from the literature was measured either on plants from natural populations in their original habitat, or paddy-grown accessions. The absence of statistically significant trait variation between *O. rufipogon*-designated and *O. nivara* designated accessions supports previous literature which also documented a continuous range of morphological and developmental trait variation amongst Asian populations of what was then-called *O. perennis*, synonymous with *O. rufipogon* (*sensu lacto*) (Morishima *et al.* 1961, 1984; Sano *et al.* 1980).

Significant morphological trait association ($p<0.05$) for panicle length, plant height, and culm angle was observed in accessions grouped both by genetic cluster and geographic range (Figure 1.7). Differences in days to heading and stolon absence or presence were also significantly associated with subpopulation structure ($p<0.05$) (Figure 1.7). REML correlation estimates between subpopulation designations indicate high negative correlations between W1 membership and both panicle number (PNNB; -0.6249) and panicle length (PNLG; -0.6132),

indicating that the W1 accessions are characterized by a low panicle number and shorter panicle lengths. In contrast, W4 membership had a modest positive correlation with panicle length (0.5701) (Table 1.2).

These trait associations corresponded to the phylogeographic association of accessions, with phenotypes at the periphery of the ranges being very different from the cluster of germplasm found in mainland SE Asia. Accessions in the W1 cluster were mainly from archipelagic Southeast Asia, and largely limited to Papua New Guinea. According to the REML correlations, these W1 accessions were characterized by an upright growth habit and extended culms with very few, short, panicles, and tended to not produce stolons, though interestingly enough, accessions from elsewhere in archipelagic SE Asia tended to be stoloniferous. Although panicle branch number was only measured on a single accession from Papua New Guinea, the other accessions from archipelagic Southeast Asia all tended to have a smaller number of panicle branches than accessions from other regions. Accessions from the W2 cluster are largely restricted to continental Southeast Asia and tend to share several traits with accessions from that region, including a long culm length and high panicle production, which contrasted with a wide culm angle, and a tendency to produce stolons. Those from the W3 group were largely from India Sri Lanka and Nepal and shared many traits with the West Asian accessions of which they were a subset. Both W3 and West Asian accessions tended to be early flowering with a very short stature, wide culm angle, no stolons and many panicles per plant, though W3 accessions tended to have panicles with fewer branches than the overall West Asian group. Plants in the W4 cluster localized to East Asia (China and Taiwan) tended to be shorter, with large numbers of very long panicles, few tillers, no stolons and were very late flowering, in general congruence with the larger West Asian group of accessions. (Figure 1.7).

Trait complexes and correlations

REML correlation estimates between the morphological and developmental traits identified in the literature as differentiating *O. rufipogon* and *O. nivara*-type plants reveal interesting correlations in our set of *ORSC* accessions. Plants that took longer to flower had a slight positive correlation with culm number (0.0810), plant height (0.0933) and stolon presence (0.0862), but a moderate negative correlation with panicle branch number (-0.1867) and slight negative correlations with panicle length (0.0815), panicle number (-0.0862), and culm angle (-0.1033). Panicle number was highly correlated with panicle length (0.4438), and weakly correlated with panicle branch number (0.0721), indicating plants with greater numbers of panicles tended to also have much longer panicles with slightly more branches, all traits which are likely related to seed-based reproductive habit. A moderate negative correlation (-0.2138) between culm number and panicle number indicated that panicle production increased as culm number decreased, raising the possibility, though not measured in this study, that the wild ancestral accessions may have multiple panicles per culm – a trait that would be in direct contrast to the cultivated *O. sativa*, which produces only a single terminal panicle on each reproductive tiller (panicle-bearing tiller). Moderate to strong positive correlations between stolon presence and both plant height (0.2810) and culm angle (0.2488), indicated that stoloniferous plants tended to produce taller or longer culms and had a more open or prostrate growth habit. Oddly, there was no correlation between stolon presence, a trait representative of clonal reproduction, and panicle number, a trait representative of seed-based/sexual reproduction (-0.0172) in this growth environment. This lack of correlation may be due to the fact that stolon number was not counted, only absence or presence, raising the possibility that there are indeed intermediate plant types which are both highly stoloniferous and produce a large number of panicles, thus having an equal sexual and

clonal reproductive capacity, an intriguing possibility which will be explored further in future phenotyping studies.

Table 1.2 Restricted maximum likelihood (REML) correlation-estimates between subpopulation membership and morphological traits associated with the differentiation of *O. rufipogon* vs. *O. nivara*-type plants (N=178) Stronger positive correlations are shaded in increasingly deep shades of blue and stronger negative correlations in increasingly deep shades of red.

	W1	W2	W3	W4	Days to heading (DTHD)	Branches at panicle base (PBRNB)	Panicle length (PNLG)	Panicle number (PNNB)	Culm number (CUNO)	Plant height (PTHT)	CULM ANGLE	Stolon presence /absence (STOLON)
W1	1.0000											
W2	-0.4463	1.0000										
W3	-0.2085	-0.5514	1.0000									
W4	-0.3847	-0.1597	-0.2126	1.0000								
DTHD	0.1218	-0.0365	-0.1957	0.1450	1.0000							
PBRNB	0.2039	-0.0234	-0.1408	-0.0503	-0.1867	1.0000						
PNLG	-0.6249	0.1517	0.0158	0.5701	-0.0815	0.1471	1.0000					
PNNB	-0.6132	0.1180	0.2554	0.2973	-0.0862	0.0721	0.4438	1.0000				
CUNO	0.0888	0.1483	0.0097	-0.3413	0.0810	-0.1506	-0.2838	-0.2138	1.0000			
PTHT	0.1552	0.0194	-0.3254	0.1868	0.0933	0.2589	0.0894	-0.1325	-0.0779	1.0000		
CULM_ANGLE	-0.2763	0.3061	0.0624	-0.1633	-0.1033	-0.2257	-0.0800	0.1439	0.3626	0.0312	1.0000	
STOLON	-0.0605	0.2399	-0.1796	-0.0380	0.0862	0.0762	0.0456	-0.0172	-0.0036	0.2810	0.2488	1.0000

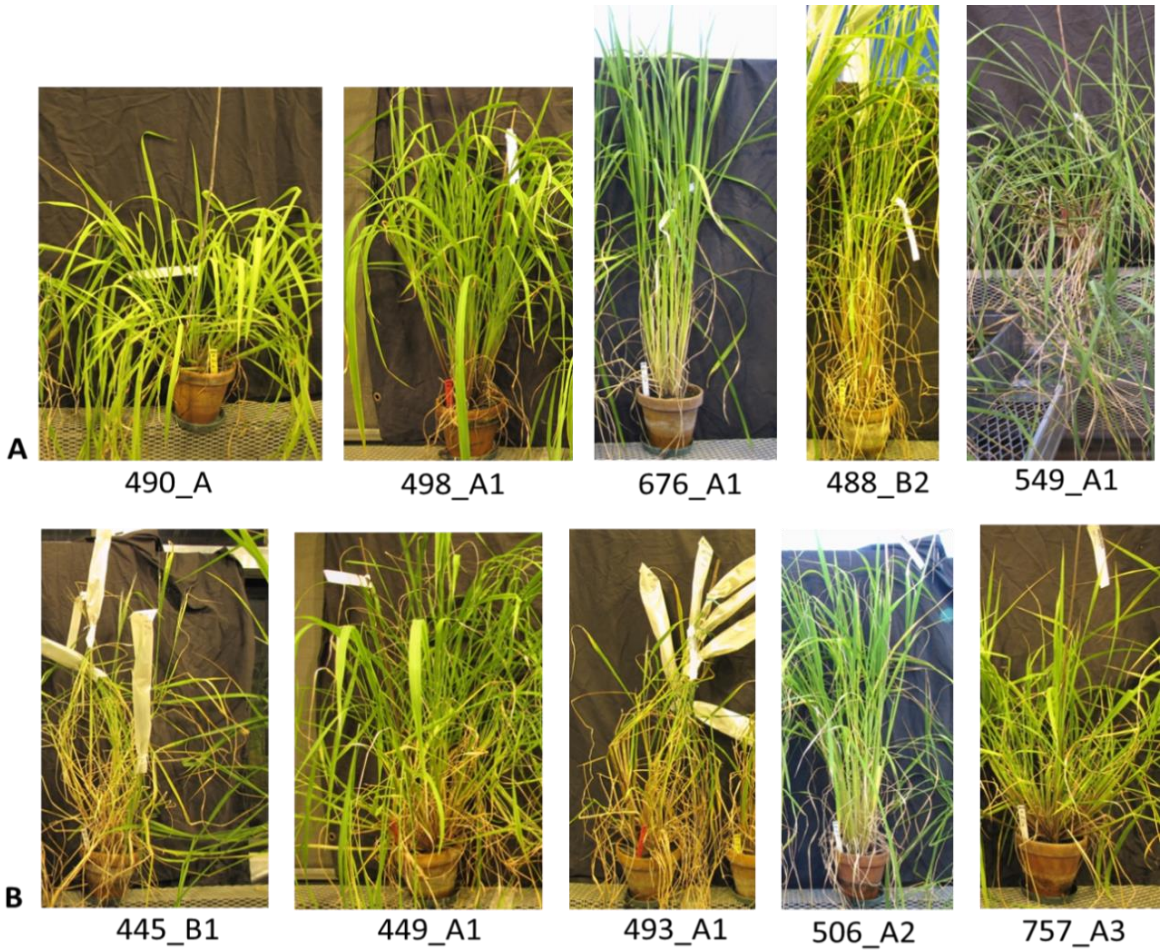
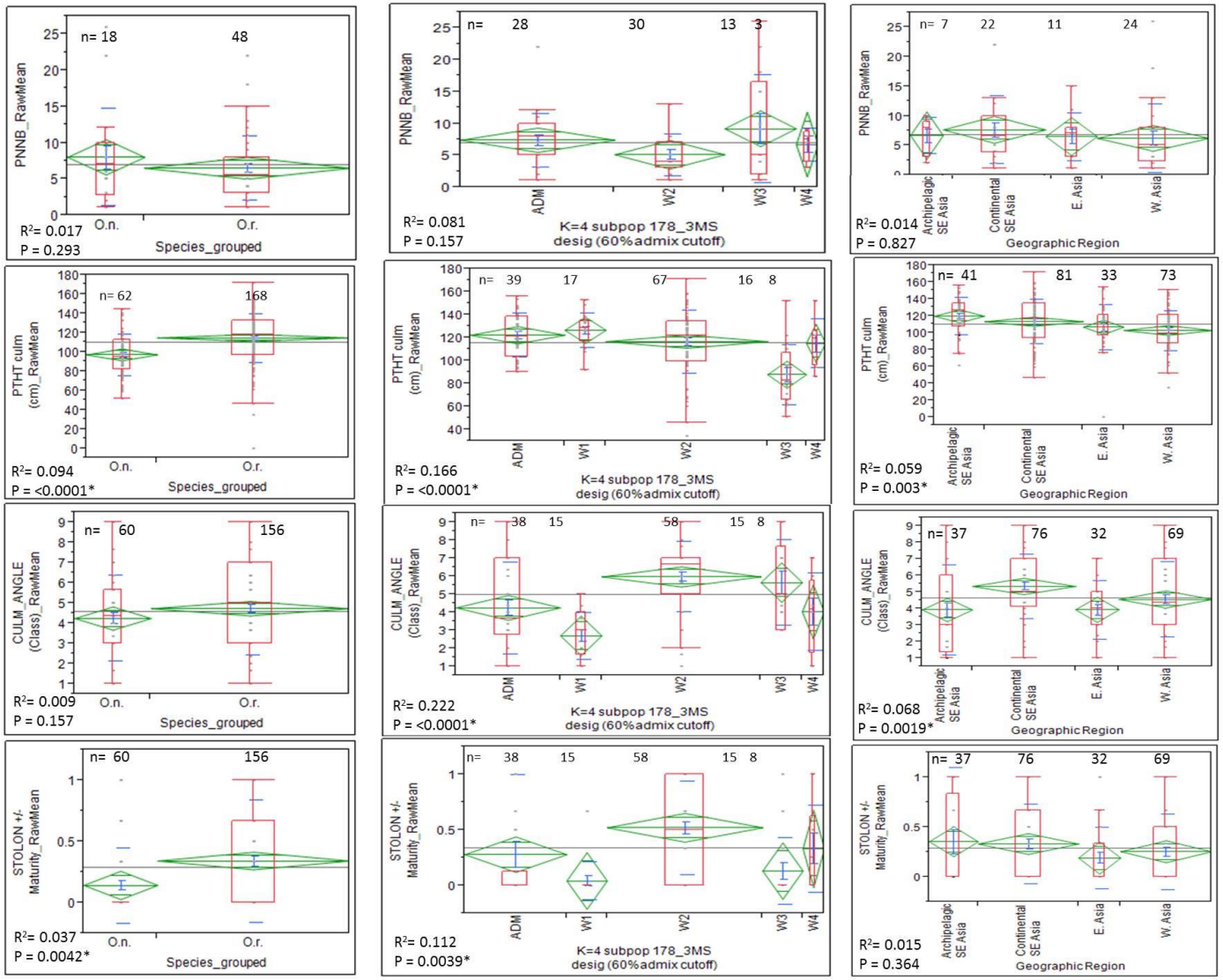


Figure 1.6 Greenhouse-grown accessions of *O. rufipogon* and *O. nivara*. A. *O. rufipogpon* and **B.** *O. nivara* at 12 weeks past germination, showing wide variation in height, culm angle, morphology, developmental stage, and stolon presence. All pictured individuals represent accessions included in the wild diversity panel genotyped and phenotyped in this study.


PNNB_RawMean
n= 18 48
R²= 0.017
P = 0.293
O.n.
O.r.
Species_grouped
n= 28 30 13 3
R²= 0.081
P = 0.157
ADM W2 W3 W4
K=4 subpop 178_3MS desig (60%admix cutoff)
n= 7 22 11 24
R²= 0.014
P = 0.827
Archipelagic SE Asia
Continental SE Asia
E. Asia
W. Asia
Geographic Region
PTHT culm (cm)_RawMean
n= 62 168
R²= 0.094
P = <0.0001*
n= 39 17 67 16 8
R²= 0.166
P = <0.0001*
ADM W1 W2 W3 W4
K=4 subpop 178_3MS desig (60%admix cutoff)
n= 41 81 33 73
R²= 0.059
P = 0.003*
CULM_ANGLE (Class)_RawMean
n= 60 156
R²= 0.009
P = 0.157
n= 38 15 58 15 8
R²= 0.222
P = <0.0001*
n= 37 76 32 69
R²= 0.068
P = 0.0019*
STOLON +/- Maturity_RawMean
n= 60 156
R²= 0.037
P = 0.0042*
n= 38 15 58 15 8
R²= 0.112
P = 0.0039*
n= 37 76 32 69
R²= 0.015
P = 0.364

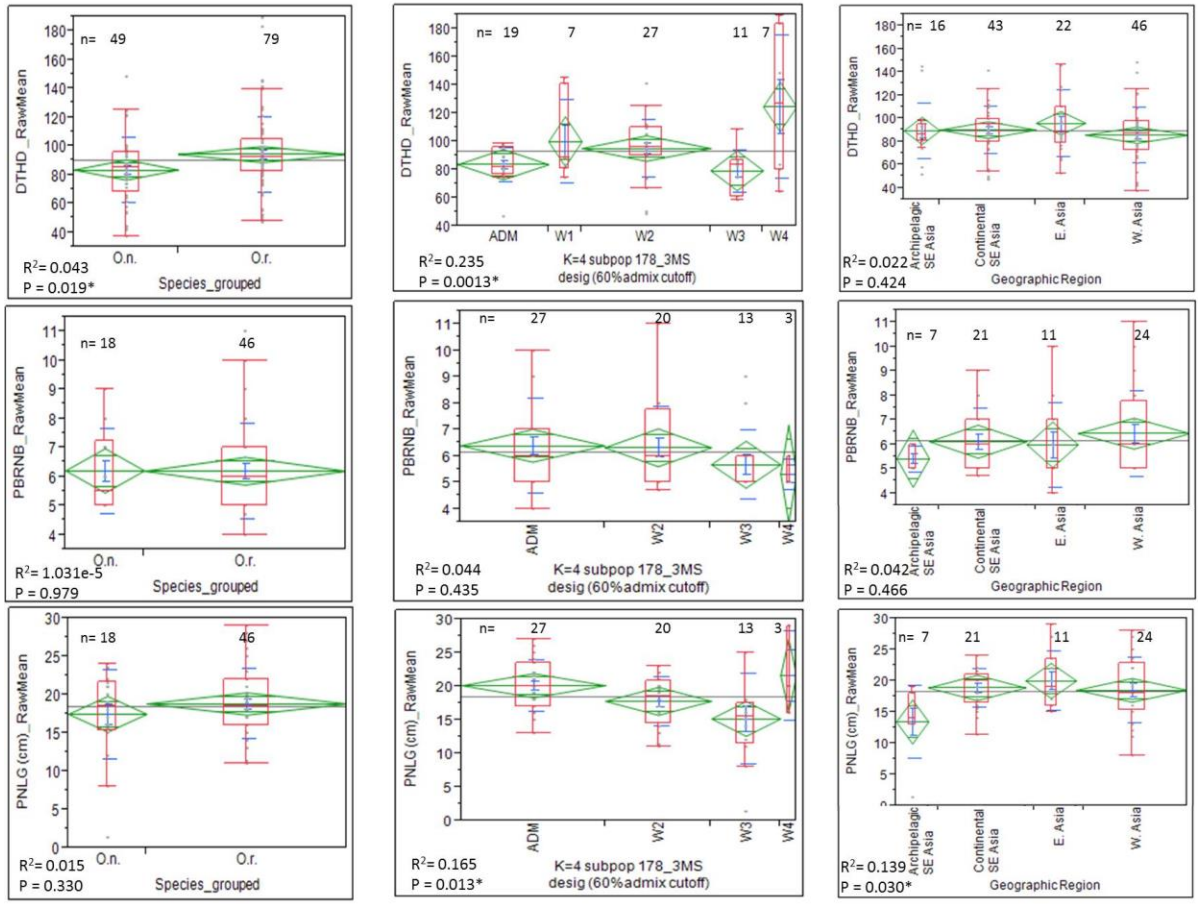


Figure 1.7 Box-plots of seven morphological traits purported to differentiate O.rufipogon from *O. nivara*-type individuals by species designation, subpopulation, and geographic region. The tops and bottoms of the green diamonds indicate the 95% confidence interval, with the width varying by the sample number. Red boxes indicate the 1st -3rd quantile, with whiskers to the limits of 1.5x the interquartile range. The blue bar indicates the mean error, and the flanking disconnected blue lines indicate the standard deviation., *significance at P<0.05.

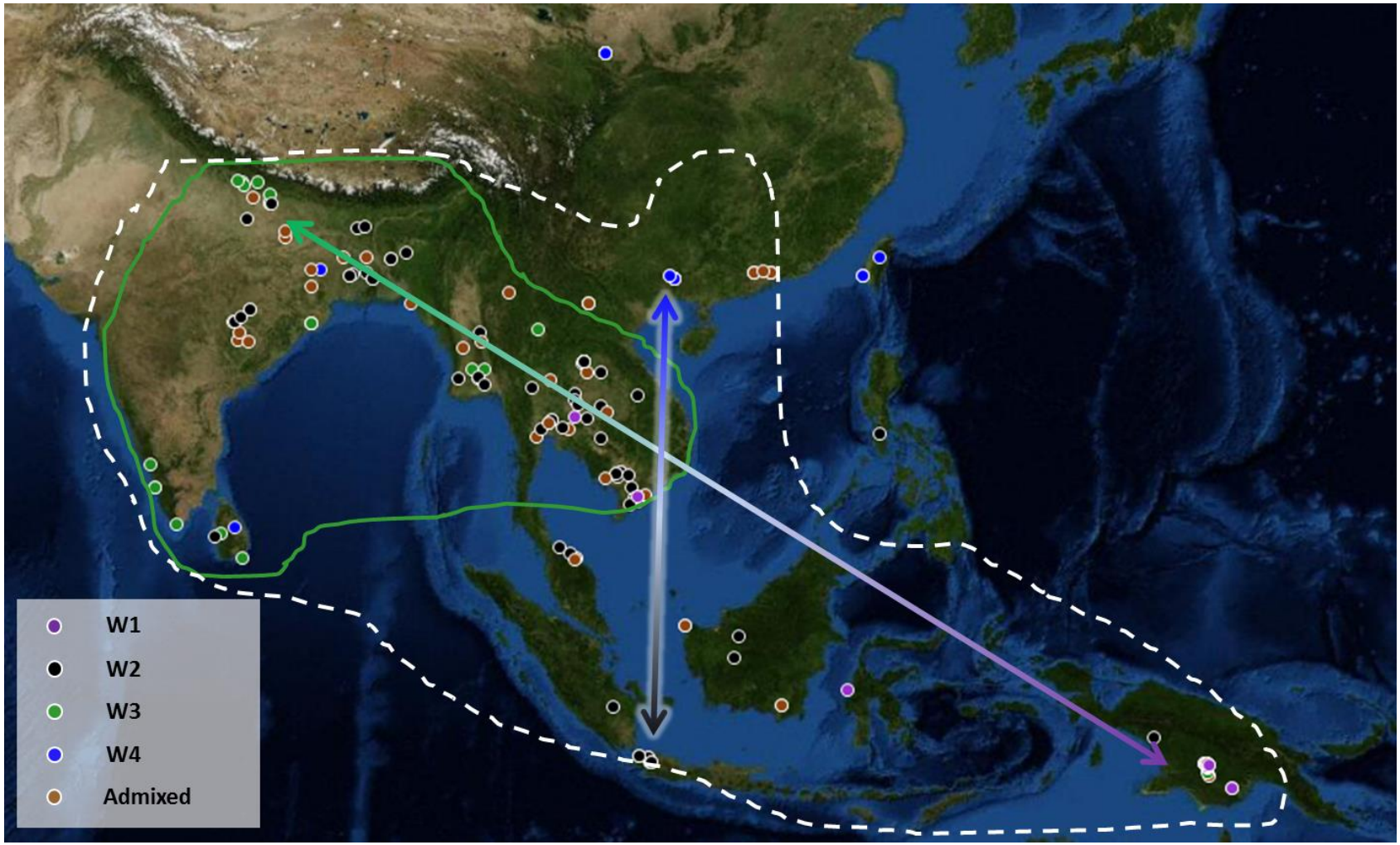


Figure 1.8 Geographic map showing the phylogeographic distribution of the 178 *ORSC* accessions genotyped in this study. Accessions are colored according to the subpopulation clusters at K=4 (See Fig. 1.7) with a 60% admixture cutoff designation.

Phylogeographic distribution of *ORSC* accessions

Evidence from the population structure analysis (Figure 1.5), and the mapping of the accessions by subpopulation on a geographic map based on collection coordinates (Figure 1.8) suggest that population substructure in the *ORSC* is partly correlated with geographic distribution and likely a result of geographic and climate change-based barriers to gene flow imposed over the past ~20,000 years. The four major subpopulations of the *ORSC* identified in this study show a northwest-southeast and a north-south clinal specificity (Figure 4.3).

Subpopulation W1, highly specific to Papua New Guinea in the extreme southeast is geographically opposed to the northwest W3 subpopulation composed mostly of accessions from north India and Nepal. The W2 subpopulation with many accessions from continental Southeast Asia, as well as Malaysia and Indonesia, separates from the W4 subpopulation whose members are largely limited to eastern China and Taiwan. This phylogeographic distribution interestingly also places W3 members almost entirely within what is thought to be the current extant range of the annual form of the species complex, *O. nivara,* and W1 members almost entirely within the current extant range of the perennial form, *O. rufipogon*, with W2 and admixed accessions largely in the shared range between the two forms. Most W3 accessions are genebank-designated as *O. nivara,* and all of the W1 accessions have been designated as *O. rufipogon*. Given that the R2, R4, and the admixed accessions include both species designations, and past literature that reports the presence of 'intermediate' form (Morishima *et al.* 1961, 1984; Vaughan *et al.* 2003), and an annual-perennial continuum (Sano *et al.* 1980), this suggests that there are genetically distinct annual (W3) and perennial (W1) subpopulations, but these do not correlate with species nomenclature or the mostly developmental and morphological distinctions by with accessions were classified as perennial/*O. rufipogon* or annual/*O. nivara*.

Hydrologic cline-based differentiation of plant morphology and reproduction

This study only measured the morphological and developmental traits in greenhouse-grown plants under a single hydrologic condition—that of water-sufficient but non-submerged plants, as the standing water level was always below the base of the culms in pots, though lodged tillers or horizontal stolons sometimes dropped below the pot rim into the standing water of the growth tanks. Based on a deeper understanding of the genetic and phenotypic diversity of the *ORSC* gained through this study, as well as the literature-based reports on intermediate plant types and an annual-perennial continuum (Morishima *et al.* 1961, 1984; Sano *et al.* 1980; Vaughan *et al.* 2003) and the semi-overlapping ecological niches preferred by the more annual, *O. nivara*-type plants or more perennial *O. rufipogon*-type plants (Tateoka 1964a; Oka and Morishima 1967; Vaughan 1989, 1994), we suggest a hydrology-centric model based on phenology, genetics, and phenotypic plasticity to account for the range in life habit and reproductive habit-related developmental and morphological variation of the species complex (Figure 1.9). Plants growing in areas with a low water table, such as on the banks of streams or lakes (Figure 1.9, far left) are seasonally flooded by monsoon rains which trigger synchronous flowering. These plants have a more upright growth habit, reproduce sexually, are largely self-pollinating, which guarantees fertilization, and produce relatively large quantities of seed. They tend to be annual and die after seed maturation and shedding. In contrast, individuals living further out in deeper water of stream or lake beds have roots and culm bases that are fully submerged year round (Figure 1.9, far right). Ethylene, abscisic acid (ABA), and giberellic acid (GA)-mediated internode elongation, node exposure, and nodal adventitious root development, promoted by submersion, as found in deepwater *O. sativa* cultivars (Hoffmann-Benning and Kende 1992; Kende *et al.* 1998; Hattori *et al.* 2008, 2009), in anddition to a prostrate shoot architecture and stolon

development are all characteristics of these aquatic individuals. Reproductively, they tend to be short-day photoperiod sensitive and therefore could be reproductive isolated from the seasonally-wet, annual individuals due to a short overlap in panicle heading period, but tend to be more outcrossing, produce few, intermittent, panicles and seed, and are perennial, reproducing vegetatively by stolon adventitious rooting and lateral meristem outgrowth to form clonal ramets. Individuals in the intermediate hydrology zones, however, may be exposed to more extreme phenological fluctuation in water levels over the course of a single season and over several seasons, and thus both show intermediate morphological, developmental, and reproductive characteristics, and retain the genetic diversity for increased phenotypically plastic response to environmental changes through outcrossing with individuals at both aquatic and seasonally-wet ecological zones (Figure 1.9, middle).

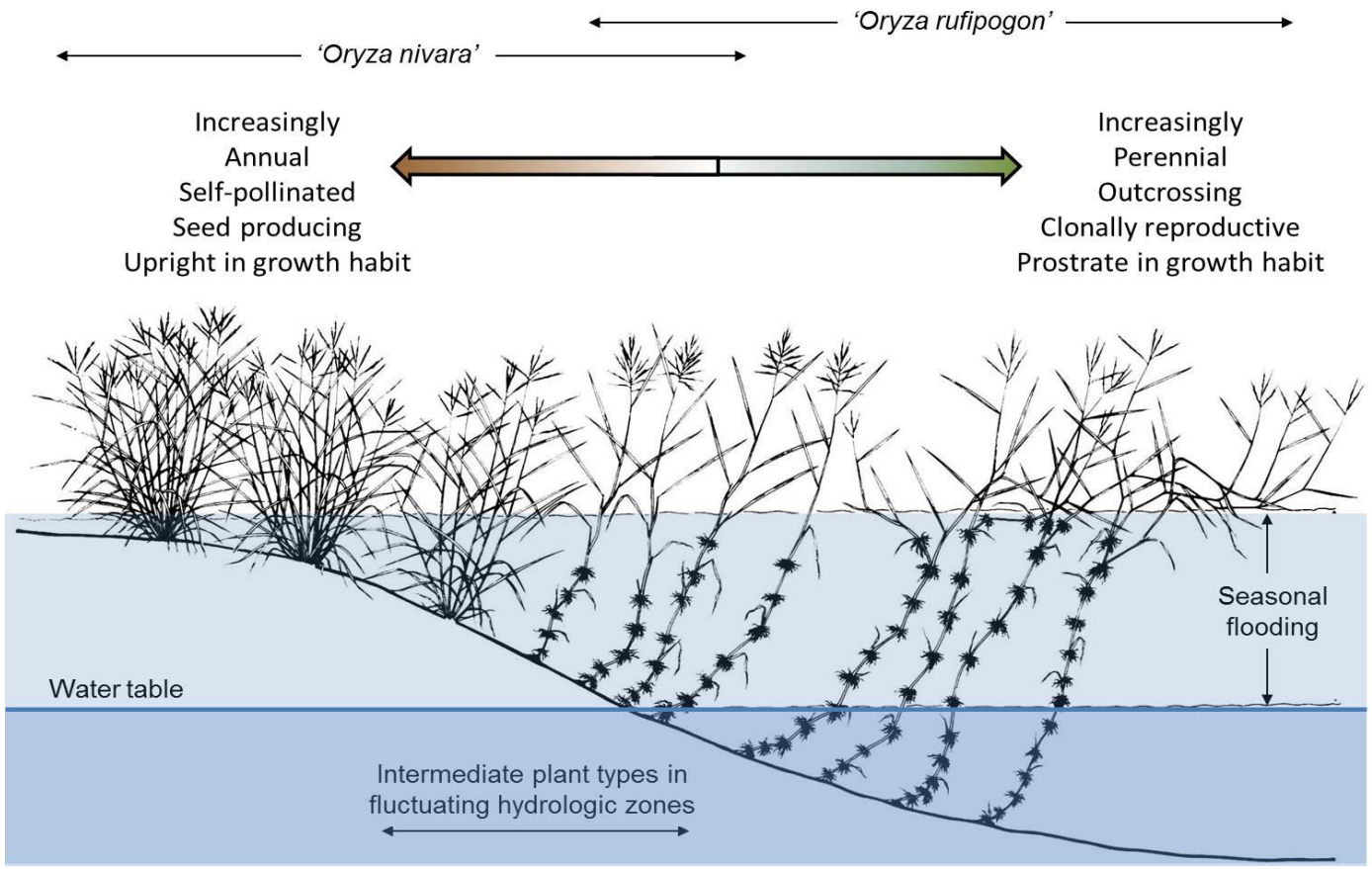


Figure 1.9 Hypothetical model for differences in plant morphology, reproductive habit, and life habit along a hydrologic cline. In their natural environment, populations of annual, perennial, and intermediate forms of the *ORSC* coexist and may have evolved though adaptation to a range of hydrologic zones. Intermediate forms in areas where the water table may fluctuate greatly between seasons and over years could retain the genetic potential and phenotypic plasticity to alternate between forms or generate more annual or perennial populations depending on environmental pressures (Modified from Vaughan, 1998).

Hypothetical models for the evolution and domestication of *O. sativa*

Two possible models for the domestication of *O. sativa* from the wild *ORSC* ancestral complex shown in Figure 1.10. The first represents a monophyletic origin, suggesting that *O. sativa* was domesticated from a single gene pool within the *ORSC* and that the subpopulations observed in *O. sativa* today differentiated following domestication. The multiphyletic origin, in which different subpopulations of *O. sativa* originate from two or more lineages of the *ORSC* that existed as predifferentiated gene pools prior to domestication. In Figure 1.10a, *indica* is pictured as being derived from the same ancestral gene pool as *japonica,* pictured here as a proto-*japonica* population, though *japonica* could alsp theoretically be the derived subpopulation. Hybridization between wild populations and cultivated subpopulations under human selection is continuous, through increasingly limited, throughout the domestication process to the present, giving rise varying degrees of admixture in extant lines of both wild and cultivated species.

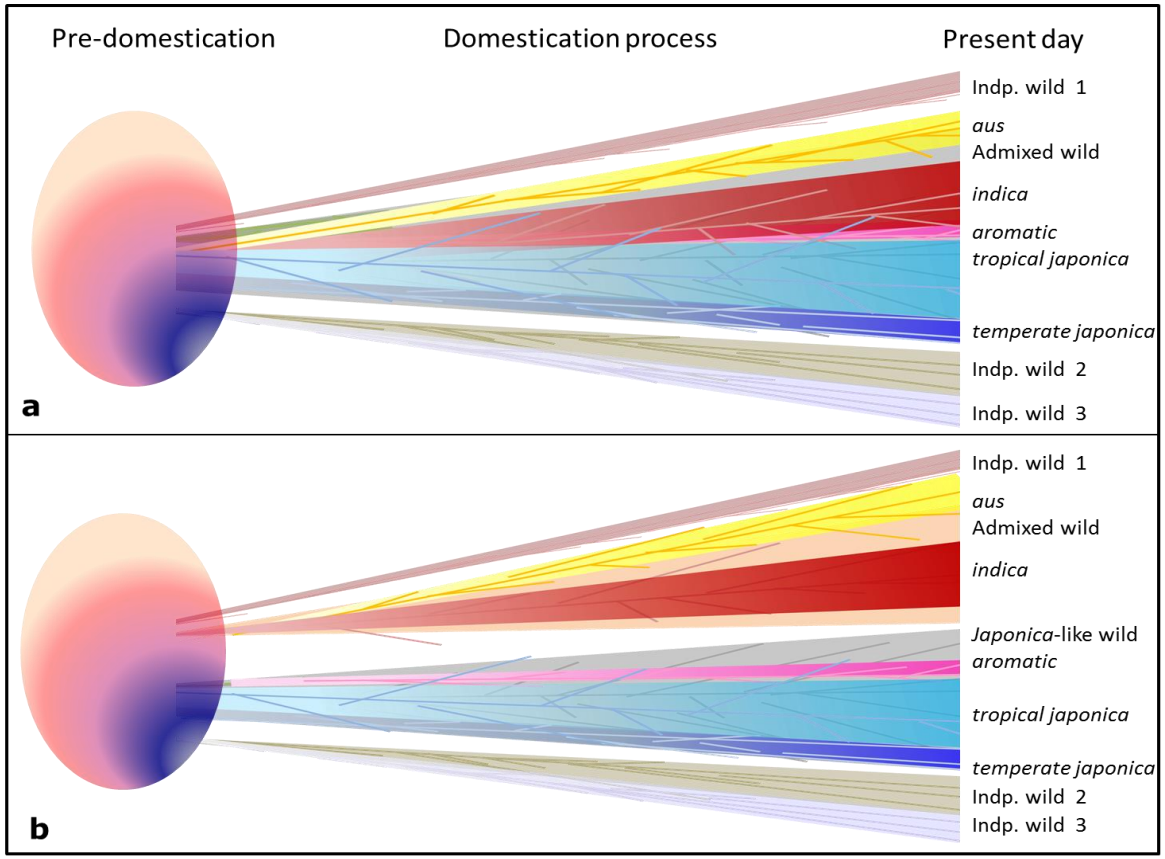


Figure 1.10 Two possible models for the domestication of O. sativa from wild *ORSC* ancestors. a. Monophyletic origin, in which a single domestication event gives rise to *O. sativa*. **b.** Multiphyletic origin, in which the five *O. sativa* subpopulations originate from two or more domestication events from a predifferentiated ancestral gene pool. In both models, *ORSC* populations are represemted as lighter transects. Hybridization between wild populations and subpopulations under human selection is continuous, through increasingly limited, throughout the domestication process to the present, giving rise to some level of admixture in extant lines of both wild and cultivated species.

Panel composition for GWAS in the *ORSC*

One of the secondary objectives of this study was to select a representative subset of around 100 geographically genetically and morphologically diverse accessions to be used for further analysis of transgressive variation and genome wide association in the *ORSC*. A total of 115 initial accessions were chosen based on the following criteria, listed in order of highest to lowest priority:

1. Enough selfed seed (≥3) for subsequent grow-out and individual plant phenotyping
2. All *japonica*-like accessions* (Supplementary Table 1.3, 2,1).
3. All three *ORSC* lines used as CSSL parents
4. All accessions with white pericarp, as representative of a possible hybrid/*O. spontanea* genetic background
5. A range in days to heading (early -- ≤55 late -- ≥118)
6. A range in selfed seed production (15≤200), representing a range
7. Accessions used in previously published research by other research groups.
8. All ‘Rufi 1’ and ‘Rufi 2’* (roughly corresponds to ‘W1’ and ‘W2’ clusters mentined in this study (Supplementary Table 1.3, 2.1).
9. All accession designated as admixed with 50-59.99% membership in a single cluster* (Supplementary Table 1.3, 2.1).
10. Country of origin- *Indica* (509), *Aus* 556, 735 (papers)

11. Plant height/tiller # - Indica 605 tall, Aus: 139, 155 tall, , 716 short, 681 high tiller#

* According to a preliminary STRUCTURE-based clustering using 49 SSR markers on a set of 283 *ORSC* and 50 *O. sativa* accessions with a 60% membership designation requirement requirement

Discussion

Genetic diversity in the *ORSC*

In this study, we set out to analyze population structure and genetic diversity in a panel of 287 diverse accessions, 2) using genotypic and phenotypic data, determine whether *ORSC* accessions in this panel were distinguishable by nomenclature, morphology and development, population structure, and geographic range, 3) select from the larger panel a subset of ~100 accessions representing maximum diversity for the aforementioned characteristics for further study of transgressive variation and GWAS.

Using genotypes from a trio of marker sets consisting of 49 SSRs, 41 MITEs, and 29 SINEs, on 178 *ORSC* accessions and 45 *O. sativa* varieties, we conducted a Bayesian-based genetic clustering analysis and determined using the DeltaK method described in Evanno et al. (2005), that K= 4 was the most significant number of clusters within the *ORSC* when analyzed without the *O. sativa* varieties. These four clusters were separated along two phylogeographic clines. The W3 and W1 subpopulations separated along a northwest to southeast axis, which appears to correspond to the proposed maximum northwestern and southeastern geographic limits of the ancestral species complex range during the Holocene, corresponding to present-day localization in Nepal and India, and Papua New Guinea, respectively. The W4 and W2 subpopulations separate along a north to south axis that approximately correspond with the northern limit of the ancestral species range during the early Holocene and the current sympatric zone shared between *O. rufipogon* and *O. nivara* type plants as suggested by Vaughan (1994).

Four out of seven morphological and developmental characteristics reported in the literature to differentiate *O. rufipogon* and *O. nivara* individuals failed to show statistically significant

variation between the accessions in our study that were genebank designated as either species, confirming the assertion held by other researchers that morphological variation for many characteristics exists in a continuous range across the species complex (Morishima *et al.* 1961, 1984; Sano *et al.* 1980) and is not a reliable or justifiable means to distinguish the complex into two "species" or even two morphotypes. Although there are almost certainly morphological and developmental traits or trait indices that separate strictly annual from strictly perennial plant types, a comprehensive phenotypic survey on a diverse range of germplasm directly categorized as annual or perennial has not yet been done to determine what these traits are. We attempted to carry out such a survey with a 64-accession subset of the wild accesions featured in this study, grown under field conditions in Nanning, China. The inherent bias of this subset of accessions to seed producers, in addition to field growout issues with experimental variation, replication, and limitations on the number and accuracy of the 13 predicted key traits surveyed produced inconclusive results.

Conflicting results in previous literature

These results contrast with some of the conclusions on the phylogeographic distribution of *ORSC* explored in earlier studies. Huang et al. (2012) used 48 genome wide sequence tagged sites (STSs) in 108 genebank-designated *O. rufipogon* and *O. nivara* accessions across the global range of the complex, and using Bayesian clustering, PCA, and AMOVA, found two genetically distinct groups, which they call 'Ruf-I' and 'Ruf-II,' in a clinal variation pattern from northeast to southwest Asia. Subgroup 'Ruf-1' highly localized to China with some continental SE Asian representation, showed genetic similarity to the *indica* subpopulation of *O. sativa,* whereas 'Ruf-

II' , localized to south Asia and the Indonesain archipelago was genetically unrelated to *O. sativa*, which the authors interpreted as supporting a single origin of domestication from the Ruf-I subpopulation. Analysis of 7 isozyme loci scored in populations of *O. rufipogon* across the global range from that populations from South China and Taiwan had a clear genetic affinity to *O. sativa* subsp. *japonica*, whereas all other *O. rufipogon* populations from South and Southeast Asia were most closely related to *O. sativa* subsp. *indica* (Second and Morishima 1980; Second 1985). These contrast with the four genetic populations—an *indica*-like group from China and S/SE Asia, a *japonica*-like group from China, an independent or ancestral group also from China, and an independent or ancestral group from S/SE Asia--found by Sun et al. using a panel of 144 geographically diverse *O. rufipogon* accessions scored with 44 RFLP markers (Sun et al., 1997; reviewed in Sun et al., 2001; Huang et al., 2012). A study of sequence haplotypes in one chloroplast and two nuclear genic regions within 129 populations of annual and perennial *O. rufipogon* across the geographic range, and 204 *O. sativa* cultivars also found an *indica*-like group from China and Southeast Asia, and a *japonica*-like group from China, as well as ancestral group spread over a wide geographic range, and an *aus*-like group from India and parts of Southeast Asia, leading the researchers to conclude that *O. sativa* was the product of at least two, and possibly three independent domestication events (Londo *et al.* 2006).

Importance of germplasm selection and genotyping methods for genetic studies

The partial overlap and conflicting conclusions from the results of these studies is likely a product of both the markers used for genetic diversity analysis and the germplasm bias in the range and number of accessions surveyed. The W1 subpopulation found in our study to be

almost entirely unique to Papua New Guinea was undetected by all of the previously mentioned phylogeographic studies because no accessions were included from that country, with the exception of five Papua New Guinea accessions in the study by Huang et al (2012). This lack of representation may be due to the fact that this W1 Papua New Guinea goup all seemed to produce very few panicles and seed, thus hampering seed-based germplasm distribution. The 178 accessions genotyped and phenotyped in this study represent the most complete range of germplasm surveyed to date, though given that we only see a single wild accession with strong genetic affinity to the *japonica* subspecies of *O. sativa,* it is likely that our study also is lacking representatives of this *ORSC* subgroup, either because it is extinct or only sparsely represented in the publicly available germplasm.

The conflicting results of these genetic studies highlight the importance of developing a a panel of diverse germplasm that fully represents the natural geographic, morphological, and genetic diversity of the *ORSC* with both range and depth of coverage, as well as neutral, whole genome marker or sequence coverage in order to draw unbiased conclusions about population structure across the complex, let alone test hypotheses on evolution and domestication.

Life habit variation along hydrologic clines: influence on selection & domestication

Several studies suggest that the differences in life habit, associated morphological traits, and main mode of reproduction between annual plant types/*O. nivara* and perennial plant types/*O. rufipogon* result from adaptation to different ecological niches along a hydrologic cline (Morishima *et al.* 1984; Barbier 1989; Morishima 2001; Sang and Ge 2007a). It is difficult to determine, however, whether the annual form was derived from an ancestral perennial state, and

divergence between the two occurred ancestrally, potentially during the Holocene expansion, or whether annuality and perenniality evolved and continues to evolve multiple times in multiple places according to environmental pressures, as the presence of annual to perennial continuum along a hydrologic cline seems to suggest.

Both annual and perennial forms and/or their genetic and phenotypic intermediates could retain the genetic potential and phenotypic plasticity to switch between annual and perennial life habits depending on changes in their growing environment. Changes between annual and perennial forms of life habit might not only be possible through natural selection over several generations, but could potentially be observed within the lifecycle of single individuals in response to specific fluctuations in water availability. This potential, and the genetic, environmental, and G x E contributions thereof would need to be further explored in a series of growth perturbation studies with both *ORSC* accessions and with *O. sativa* varieties, many of which are semi-perennial as evidenced by their ability to produce a ratoon crop, but grown as annuals. If true, results of previous studies concluding that the *indica* subspecies was domesticated from perennial *O. rufipogon* populations and the *japonica* subspecies was domesticated from annual *O. nivara* populations would need to be reconsidered.

Recommendations for effective utilization of wild germplasm

Our results, showing strong phylogeographic separation of genetic subgroups within the *ORSC* and different morphological traits/trait complexes emphasizes the importance of habitat preservation and effective, standardized germplasm conservation, characterization, and management. Habitat destruction resulting in population loss of wild germplasm has likely been

ongoing through the domestication process has natural habitats are converted for human development, but has been specifically documented as early as the late 1970s (Sano *et al.* 1980) and continues at an accelerated rate into the present (Akimoto *et al.* 1999). Results also show that there is no clear genetic basis for distinguishing genebank-characterized accessions of *O. rufipogon* from *O. nivara*, in agreement with previous studies showing a lack of clear and consistent genetic differentiation between annual and perennial, or *O. rufipogon* and *O. nivara* accessions or populations(Cheng *et al.* 2003; Yamanaka *et al.* 2003; Li *et al.* 2006a; Grillo *et al.* 2009).

Accelerating the effective utilization of *ORSC* germplasm for the discovery and introgression of agronomically useful wild alleles into *O. sativa* requires multiple prerequisites:

1. Preservation of natural habitats, especially in areas of high genetic diversity, such as southern China (Gao *et al.* 2006; Wang *et al.* 2008a), and unique subpopulation representation (ex. Papua New Guinea)

2. The informed, standardized, and well-documented collection of existing wild populations prior to further habitat destruction

3. Community-wide standardization of the nomenclature and agreement on the genetic, morphological and/or ecological characteristics used in the nomenclature conferral, particularly regarding the status of putative wild-cultivated hybrids, and annual vs. perennial types as ecotypes, species, subpopulations, or subspecies and the re-characterization of current genebank accessions accordingly

4. Standardized genetic characterization of as many publicly available genebank accessions as possible, or the informed selection of a large core set of *ORSC* germplasm for further genetic, phenotypic, and breeding evaluation

5. Database management, tracking, and standardized protocols for genebank management of germplasm as pure lines accessions through forced self pollination and single seed descent (SSD) or heterogeneous population maintenance by some form of controlled seed production. Vegetative propagation of highly clonal individuals or accessions should also be considered so as not to unintentionally drive selection toward seed producing morphotypes (Yamanaka *et al.* 2003) or increase self-sterility in obligately outcrossing accessions

6. Open access sharing of germplasm and genotypic and phenotypic data, with due consideration of intellectual property management, for further characterization and use and breeding projects

Future studies on the *ORSC*

Given the results of our study on 178 diverse *ORSC* accessions supporting the presence of four phylogeographically-distributed genetic subpopulations and also indicate modest correlations between genetic subpopulation and various developmental and reproductive phenotypic characteristics, it will be interesting to further explore the genetic and phenotypic variation in the subset of ~100 diverse accessions in additional phenotypic screens and GWAS. A more detailed series of phenotypic screens will also help identify trait complexes consistent with an annual or

perennial life habit, outcrossing versus inbreeding tendencies, and a vegetative versus clonal reproduction preference and the correlation between these factors. The effect of the growth environment, particularly with regard to drought, flooding and general hydrology in relation to life and reproductive habit, possible variations in phenotypic plasticitity, and genetic by environmental cofactors, with regard to both admixed and weedy types will be expecially interesting to explore in greater detail. The understanding gained from such studies on the evolutionary relationship between the *ORSC* and *O. sativa* could be revolutionary, and the potential to drive directed, introgression-based cultivar generation in the face of increasing competition for natural resources, changing weather patterns, and rising CO_2 temperature, and sea levels could be enormous.

Supplementary Table 1.1 Taxonomic treatment of the *O. rufipogon* species complex in the literature

Taxonomy and phylogenetics of Asian rice wild progenitors	**Specific findings about or treatment of wild ancestral species**	**Validation methods**	**Germplasm**	**Markers**	Citations*
Phylogenetics of wild ancestor	K=4 (SA, China+Taiwan, New Guinea, Aust, (+America)) Incl. Aust, America; classified as annual/perennial/interm; also found japonica (some Kwangsi +Taiwan) and indica type (some S. Asia) rufis	phenology +isozyme PCA	*O. r.* -- 28 China, 5 PnG, 10 Aust, 16 Amer; 20 O. longi, 20 brevi	24 isozyme loci	*Second, G. (1985)
	4 groups by geography (SA, SEA, China, Oceania)	Review of nuclear RFLP study in Akimoto et al, 1999			*Reviewed in Cai, H.-W. et al. (2008)
	Some differentiation by geography/isolation by distance, particularly China vs. S/SEA; more differentiation by quant. Traits (PCA1/2 >60%var. as opposed to 16% for isozymes)			17 RFLP loci and 29 isozyme loci	Cai, H.H.-W. et al. (2004)
	japonica from wilds in South China; indica from wilds around Thailand, Myanmar, India = Indochina; ID's Aus haplotypes in wilds >possible independent domestication of aus in India			allelic var in one cp and 2 nuclear genes	*Londo, J.P. et al. (2006)
	K=3; indica-type, jap-type, primitive type		122 *O. r.*, 76 O.s.	48 nuclear RFLP	Sun, C.Q. et al. (1997)

	NE->SW clinal variation and subpop structure K=2 : Ruf-I from NE Asia, is indica-like, few jap-like; Ruf-II indp, higher genetic div, SEA Indochina; *O. n.* and *O. r.* not separated genetically	STRUCTURE, Bayesian Clustering, PCA	108 *O. r.* from native range	SNPs at 42 STSs	*Huang, P. et al. (2012)
	Possible 4 clusters -- showing ind/jap clustering with sativas not correlated with nomenclature, and what looks like 2 additional indp. Clusters			416 polymorphic MITE-AFLP fragments	Park, K.C. et al. (2003)
	K=3: indica-like ann. (Thai, India, Myan)+ some interm (Cam, Bgd); peren jap-like from China; indp. perennial (Ind, Nep, Thai, China, PNG, Indo), 2 subclusters: indp, tight Inter. From Nepal and India, Malay, 1 per, 1 ann; 2 peren close to indica but indp. from India, China	NJ, UPGMA trees, Structure	68 O.s. (35 Ind, 33 Jap); 35 *O. r.* (13 ann, 16 peren, 6 intm.), 5 other O. spp	49 pSINE1 members	*Xu, J.-H. et al. (2007)
***O. nivara* (annual) derived from *O. rufipogon* (perennial)**		Review			Chang, T.T. (1976)
		Review			Khush, G.S. (1997)
	Suggested that *O. n.* (annual/seasonally wet) derived from *O. r.* (perennial, aquatic)	NJ, MP ML, AMOVA	243 indiv - in 11 *O. n.* and 15 *O. r.* pops	7cp and nuclear loci seq	*Zheng, X.-M. and Ge, S. (2010)
	O. rufipogon (*O. perennis*) ancestral to both *O. sativa* and *O. n.* + weedy forms				Oka, H. (1964)
					Oka, H. (1974)
					Oka, H.I. (1977)

	perennial originating from intermediate; annual from perennial; 2 perennial acc, grouped with the ind-like annual wild cluster	NJ, UPGMA trees, Structure	68 O.s. (35 Ind, 33 Jap); 35 *O. r.* (13 ann, 16 peren, 6 intm.), 5 other O. spp	49 pSINE1 members	*Xu, J.-H. et al. (2007)
Single species (*O. rufipogon* Griff.)	Used to refer to the "entire dataset of wild pops"	Analysis of 15 overlapping selective sweep regions and 38 co-located low-diversity genomic regions (CLDGRs)			Civáň, P. et al. (2015)
	Used to refer to "common wild rice;" study looked only at genetic div. in/btw O.s, and *O. r.*		122 *O. r.*, 75 O.s.	44 single copy RFLPs	Sun, C.Q. et al. (2001)
	But no mention of *O. n.*/annual vs. perennial wild relatives	Domestication of *O. sativa* seemed to be diphyletic - strong similarity was observed between *O. sativa* Japonica-Javanica and *O. r.* from China and between *O. sativa* Indica and *O. r.* from tropical Asia.	Twelve cultivars of *O. sativa*, one cultivar of O. glab., and 17 wild accessions (12 *O. r.*, 2 O. glum., 1 O. longi., 1 O. mer. and 1 O. barthii).	16 RAPD primers, 28 RFLP probes, 24 nuclear SSLP and 10 chloroplast SSLP	Bautista, N.S. et al. (2001)

Annual perennial continuum (*O. rufipogon* sensu lacto)					Sano, Y. and Morishima, H. (1982)
	Treated as annual-perennial continuum				Shimizu, H. et al. (2010)
	O. n. mentioned but considered as part of *O. r.*	Review			Morishima, H. (2001)
	O. n. mentioned but considered as annual ecotype	Review			*Cai, H.-W. et al. (2008)
					*Vaughan, D. a. et al. (2008)
	int/ann/per ecotypes ID'd-- MITE F1-epsilon locus dist, btw life habit in *O. r.* + (ann/intm)/- (peren), except for 4 acc. - the presence or absence of the elements at each locus was established in the ancestral plants of each species or each ecotype of *O. rufipogon* before habitat expansion or indp. selected during adaptation to a habitat via linkage to genes involved in adaptation		24 *O. r.* peren, 21 ann., 5 intm; plus other AA genome	3 MITE loci	*Kanazawa, A. and Akimoto, M. (2000)
***O. sativa* biological species complex (O.s., *O. r.* and *O. n.*, + O.s. f. spont.)**		Pointed out issues with wild accs grouping with ind/jap due to potential back-introgression	4 jap, 4 ind, 27 *O. r.*, 13 *O. n.*, 10 O.s f. spont,	30 RFLPs	*Lu, B.R. et al. (2002)
					Lu, B. et al. (2000)

	supports suggested taxonomic treatment as species complex as per Lu et al. 2000, 2002)	UPGMA	8 O.s., 5 *glab*, 9 *O. r.*, 7 *O. n.*, 3 barthii, 3 O. mer, 3 glumae, 7 longi	181 RAPD fragments from 27 primers; 101 SSR alleles from 29 SSR primer pairs	Ren, F. et al. (2003)
***ORSC* -- single large gene pool**					Zhu, Q. et al. (2007)
	All wilds found as intermediate btw ind/jap	Indels differentiating ind/jap -- wilds found as intermediate and also intermediate latitudinal growth range, but strangely the more japonica-like wilds are at lower latitudes and the more indica-like wilds at higher latitutes, opposite of cultivated	33 wilds, including *O. r.*, *O. n.* and interspecific hybrids	34 Indels diff ind/jap from Lu et al 2009	*Xiong, Z.Y. et al. (2011)
Wild ancestor treated as single large gene pool	found 2 wild groups - indica and japonica like and geographic differentiation across Himalayas			57 subsp specific intron length polymorphism markers	Zhao, X. et al. (2009)

	O. r. sensu lacto	Review			Nonomura, K.-I.K. et al. (2010)
					*Londo, J.P. et al. (2006)
	Results are also supportive - *O. n.* and *O. r.* not separated genetically, but found NE->SW clinal variation and subpop structure K=2 : Ruf-I from NE Asia, is indica-like, few jap-like; Ruf-II indp, higher genetic diversity, SEA Indochina	STRUCTURE, Bayesian Clustering, PCA	108 *O. r.* from native range	SNPs at 42 STSs	*Huang, P. et al. (2012)
***O. rufipogon* with annual and perennial individuals**			72 O.s. 42 *O. r.*	sd1 seq	Asano, K. et al. (2011)
Two ecotypes					(Oka 1988; Barbier et al. 1991; Morishima et al. 1992
					Morishima 2001; Cheng et al. 2003; Vaughan et al. 2003
	but nivara and rufi mentioned/used as annual and perennial; also O. spontanea	Review			Fuller, D.Q. et al. (2010)
	but nivara and rufi mentioned/used as annual and perennial; also O. spontanea	Review			Fuller, D.Q. (2012)

Two forms (annual/perennial) of one species (*O. rufipogon*)	Also acknolwledges annual form as *O. n.*; 3 wild subpops strongly correlated with geo. Dist.; Or-IIIa --S. China, IIIb--S.India; OrI-continental Asia; OrII--cont,+oceanic SEA	ID of selective sig, PCA, Fst, NJ, QTL analysis	446 wilds, 1083 O.s.	~5M SNPs from WGS (2X)	*Huang, X. et al. (2012)
	Uses *O. n.*=annual, *O. r.* =perennial	most *O. r.* neg for pSINE-r2, ND ORF100, J allele for CMN; most nivara opposite	23*O. r.*/peren., 23 *O. n.*/ann. All from Cmb, Thai, Viet, Laos, 1 China; 1 ind, 1 jap	distribution of p-SINE1-r2 in the waxy locus, cp ORF100, CMN-A32 primer PCR product	Yamanaka, S. et al. (2003)
Two ecotypes of one species (*O. rufipogon*)	intron seq of 4 genes -- shows ind jap-grouped rufis				Zhu, Q. and Ge, S. (2005)
	Shows ind and jap grouped rufi/niv		20 *Oryza* sp +3 outgroups; 1 ind, 1 jap. 1 niv, 3 *O. r.*	3 cp regions	Kumagai, M. et al. (2010)
	Suggested *O. n.* (annual/seasonally wet) derived from *O. r.* (perenn, aquatic)	NJ, MP ML, AMOVA	243 indiv - in 11 *O. n.* and 15 *O. r.* pops	7cp and nuclear loci seq	*Zheng, X.-M. and Ge, S. (2010)
	Supports treatment as 2 ecotypes or subsp. of a single sp. b/c of high relatedness and ambiguous sp. boundaries		16 ind, 14 jap, 18 *O. r.*, 12 *O. n.*	10 unlinked genes seq on 9 chrs.	Zhu, Q. et al. (2007)
	No sig gen diff btw ann/perenn. O.n-annual mentioned but considered as annual ecotype		3 annual, 5 perenn, 1 O. long	phytochrome and	Barbier, P. et al. (1991)

				prolamine gene seq	
Three ecotype classes of one species (*O. rufipogon*)	Ann., perenn., interm. - perennial originating from intermediate; annual from perennial.		68 O.s., 35 *O. r.*	49 pSINE1	*Xu, J.-H. et al. (2007)
	69bp deletion in ORF100 -- ann. & interm mostly del. (86.5%); perennial mostly non-del (77.2%)		137 O.s.; 82 *O. r.*, 35 from other AA spp.		Chen, W. et al. (1993)
	showing ind and jap plastotype differentiation; also divided into Asian, American, Aust. *O. r.* forms			cp RFLPs	Dally, A.M. and Second, G. (1990)
	ann/peren/interm ecotypes with 2 semi-specific markers	pSINE1 r215 (+ annual/- peren & interm.); r503 (+ annual & peren/ - interm.), but not 100% exclusive	68 O.s. (35 Ind, 33 Jap); 35 *O. r.* (13 ann, 16 peren, 6 intm.), 5 other O. spp	49 pSINE1 members	*Xu, J.-H. et al. (2007)
Two species - independent taxonomic identities					Duan, S. et al. (2007)
	Recently diverged sister species - O. rufi =perenn; O. nivi - annual				Grillo, M. a et al. (2009)
	Two morph/reprod diff. sp but sym/parapatric w/little genetic differentiation	digital gene exp, ID of directional selection	6 morphotypic *O. r.* and 6 morphotypic *O. n.*	1,717 diff. exp genes in 3 repro-rel tissues	Guo, J. et al. (2015)

	O. r. = perennial, trop, oceanic Asia; *O. n.* = annual, continental Asia		1 ind, 1, jap, 1 O.r, 1 *O. n.*, +5 other AA genome sp	53 single-copy nuclear genes representing diverse functional categories, together with 16 intergenic regions	Zhu, T. et al. (2014)
	Assumes *O. n.* and *O. r.* are separate progenitor spp of O.s. to look at genomic struct var.				Hurwitz, B.L. et al. (2010)
	Also includes 1 acc of perennis. Suggests sep. of rufi into 2 taxons based on genetics, but only surveyed limited variation	UPGMA	11 O.s., 2 *O. r.*, 1 *O. n.*, 1 O. perenn., plus other *Oryza* spp.	11 ISSR polym.	Joshi, S.P. et al. (2000)
	Confusing; *O. n.* often referred to as the annual "form" or "ecotype"	Review - looking at tiller plasticity			Mohapatra, P.K. et al. (2011)
	Together comprising progenitor group Asian wild rice	phylogenetic analyses - FRAPPE, NJ, PCA	40 O.s., 5 *O. r.*, 5 *O. n.*	6.5M SNPs from WGS	*Xu, X. et al. (2011)
	Acknowledges taxonomic ambivalence, but considered as distinct sp. accd to morphology	TESS, STRUCTURE, PCA, NJ	6 O.s., 5 O.mer, 43 *O. n.*, 47 *O. r.*, 24 interm. *O. r./O. n.*	29 SSRs	*Banaticla-Hilario, M.C.N. et al. (2013)

	No resolvable diff found in organelle genomes btw O.s., *O. r.*, *O. n.*, also some O. mer, O. glum		50 acc of 21 *Oryza* sp-incld 2 O.s., 1 *O. n.*, 4 *O. r.*	7mt, 5cp SSRs	Nishikawa, T. et al. (2005)
	Separate clustering of *O. r.* and *O. n.*, with all *O. n.* clustering with O.s. (all indica)	UPGMA, Nei's	23 O. species, incl. 3 O.s. (all ind.), 4 *O. n.*, 5 *O. r.* (all Thai or Viet)	1191 polymorphic AFLP loci	Aggarwal, R.K. et al. (1999)
	Treated as separate taxa, but showing ind/jap clustering with sativas not correlated with nomenclature, and what looks like 2 additional indp. Clusters			416 polymorphic MITE-AFLP fragments	Park, K.C. et al. (2003)
Reports of markers associated with differentiating taxa	int/ann/per ecotypes ID'd-- MITE F1-epsilon locus distinguishes between life habit in *O. r.* + (ann/intm)/- (peren), except for 4 acc. - the presence or absence of the elements at each locus was established in the ancestral plants of each species or each ecotype of *O. rufipogon* before habitat expansion or indp. selected during adaptation to a habitat via linkage to genes involved in adaptation		24 *O. r.* -- peren., 21 ann., 5 intm; plus other AA genome	3 MITE loci	*Kanazawa, A. and Akimoto, M. (2000)
	Uses *O. n.*=annual, *O. r.* =perennial	most *O. r.* neg for pSINE-r2, ND ORF100, J allele for CMN; most nivara opposite	23*O. r.*/peren., 23 *O. n.*/ann. All from Cmb, Thai, Viet, Laos, 1 China; 1 ind, 1 jap	distribution of p-SINE1-r2 in the waxy locus, cp ORF100,	Yamanaka, S. et al. (2003)

				CMN-A32 primer PCR product	
	Pox-1 isozyme locus- 2A allele associated with annuality -- >0.95 in O.s., O.glab, annual and weedy types of *O. r.*, and intermediate types to a lesser freq. (0.72) ; perennial types mostly either 2A (0.47) or 4A(0.47) or het (20%)	accessions were also phenotyped for traits related to life habit	452 O.s., 10 O.glab, *O. r.* - 24 peren, 11 interm, 16 ann., 9 weedy	Pox-1 isozyme allele	Morishima, H. (1991)
	69bp deletion in ORF100	ann. & interm mostly del. (86.5%); perennial mostly non-del (77.2%)	137 O.s.; 82 *O. r.*, 35 from other AA spp.		Chen, W. et al. (1993)
					Kanno, A. et al. (1993)
		no clear geographic diff. of ind vs. jap types in *O. r.*	70 Chinese *O. r.*, 27 Indian, 17 Thailand, 1-12 from other countries	69bp del in ORF100 only	Sun, C. et al. (1996)
	pSINE1 r215 (+ annual/-peren & interm.); r503 (+ annual & peren/ - interm.), but not 100% exclusive		68 O.s. (35 Ind, 33 Jap); 35 *O. r.* (13 ann, 16 peren, 6 intm.), 5 other O. spp	49 pSINE1 members	*Xu, J.-H. et al. (2007)
Mention of O. spontanea					
					Fuller, D.Q. et al.

					(2010)
	O. spontanea is generally considered as the progenies of the natural crosses between local cultivars and O. r or *O. n.*		316 O.s., 45 *O. r.*, 50 *O. n.*, 17 O. spont	36 SSRs	*Lu, J. et al. (2008)
***O. sativa* f. spontanea Roshev**	A synonym for *O. sativa*				http://www.theplantlist.org/tpl1.1/record/kew-465119
	A synonym for *O. n.*/Asian annual type				Sano, Y. et al. (1980)
	Weedy rice - hybrid between O.s. and *O. r.* or *O. n.*				Lu, B.-R. et al. (2002)
					Crawford, G. (2011)
					(Gealy et al., 2009; Olsen et al., 2007; Xia et al., 2011)
	Weedy rice - same species as O.s. - in upper latitudes (>35° N), similar to the locally cult. Jap vars; in reg <35° N, sim. to indica vars & common wild rice				Zhang, S. et al. (2014)
	O.s. var spontanea - A synonym for *O. rufipogon*/perennial type	review, quoting Oka, 1980; Grist, 1975			Oka, 1980; Grist, 1975 in Hill, R.D. (2010)
	O.s. subsp. Spontanea - weedy rice				Toriyama, K. et al. (2005)
Australian rufipogon different from	Australian rufi more similar to O. meridionalis	DNA sequence, undefined			Ishikawa, unpublished data in: Henry, R.J. et al.

Asian					(2009)
	Asian *O. r.* closer to O.s. than Aust. *O. r.*; Aust *O. r.* and O. mer closely related--Aust *O. r.* may be a perennial form of O. mer	cp WGS alignment (mapped to Nippon); MP, ML, Bayesian ana.	4 indiv/acc of 1 Asian *O. r.*, 1 Aust *O. r.*, 1 Aust O.mer for cp WgS + GB cp seq from 1 jap, 1 ind, 1 *O. n.* 1 O. aus	90 markers from cp WGS	*Waters, D.L.E. et al. (2012)
	also finds Australasian rufi clustering with O. mer	TESS, STRUCTURE, PCA, NJ	6 O.s., 5 O.mer, 43 *O. n.*, 47 *O. r.*, 24 interm. *O. r./O. n.*	29 SSRs	*Banaticla-Hilario, M.C.N. et al. (2013)
	Australian rufi more similar to O. meridionalis acc'd to cp WGS	MP, NJ, ML, Bayesian Inf.	2 ind, 2 jap, 3 *O. r.*, 1 *O. n.*, 6 other AA genome spp.	221 parsimony informative variations from cp WGS	Wambugu, P.W. et al. (2015)
	Mentions suspected crossing btw. Oceanic *O. r.* and O.mer	Review			*Cai, H.-W. et al. (2008)
	Meridionalis tightly clustered in RAPDs tree, but with other spp in SSR tree	UPGMA	8 O.s., 5 O. glab, 9 *O. r.*, 7 *O. n.*, 3 barthii, 3 O. mer., 3 O. glum., 7 O. longi.	181 RAPD fragments from 27 primers; 101 SSR alleles from 29 SSR	Ren, F. et al. (2003)

				primer pairs	
	Oceanian rufi (Australia + some New Guinea) different from American + Asian rufi	phenology +isozyme PCA	*O. r.* -- 28 China, 5 NG, 10 Aust, 16 Amer; 20 O. longi, 20 O. brevi	24 isozyme loci	*Second, G. (1985)
	Aust. form of annual wild morphologically distinct -> O. meridionalis	Morphology only: Annual wild indiv from Aust ("*O. n.*-like") were separated out by PC1 and 3	132 indiv of 30 accs. reID'd as: 6 *O. r.*, 4 *O. n.*, 2 *O. n.*-like fr Aust, 4 hyb, 8 weedy, 1 *O. r.*xO.s., 1 *O. n.*xO.s., 4O.s., 2 barthii		Ng, N.Q. et al. (1981)
	Aust form of annual wild (formerly a subform of *O. perennis* or *O. n.*) named O. meridionalis	Morphologically similar to nivara but more compact panicle, short 2nd rachises tightly held to main axis, and stiff, erect panicle branches			Ng, N.Q. et al. (1981)
* Denotes seminal or particularly important literature					

Abbreviations: ind - Indica; jap - japonica; O.s. - *Oryza* sativa; *O. r.* - *O. rufipogon*; *O. n.*; *O. nivara*; O. mer - O. meridionalis; O. glab - O. glaberrima; O. glum - O. glumaepatula; O. longi - O. longistaminata; O. brevi - O. brevigulata; ann. - annual; peren. - perennial; intm - intermediate; hyb - hybrid; cp - chloroplast; SEA - Southeast Asian

Supplementary Table 1.2 Domestication hypotheses and previous genetic studies on the *ORSC*.

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
Single domestication (monophyletic)	indica domesticated from *O. rufipogon*; japonica derived from indica				Ting, Y. (1957)
		Pointed out issues with wild accs grouping with ind/jap due to potential back-introgression	4 jap, 4 ind, 27 *O. r.*, 13 *O. n.*, 10 O.s f. spont,	30 RFLPs	*Lu, B.R. et al. (2002)
		O. nivara (annual) derived from *O. r.* (perennial)			Chang, T.T. (1976)
	O. rufipogon (*O. perennis*) ancestral to both *O. sativa* and *O. n.* + weedy forms				Oka, H. (1964; 1974; 1977)
	Supportive of indica domesticated from *O. rufipogon*; japonica derived from indica				Duan, S. et al. (2007)
	Single domestication of *O. sativa* from intermediate ann/perennial pops of *O. perennis*				Sano, Y. et al. (1980)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	Single domestication of *O. sativa*, subsequent split into indica and japonica				Wang, X. et al. (1984)
					Oka, H.I. and Morishima, H. (1982)
		Review - based on shattering, dormancy, rc genes			Vaughan, D. a. et al. (2008)
		Review			*Vaughan, D. a. et al. (2008)
			Chinese only	36 SSRs	Wang, M.X. et al. (2008)
		Diffusion based demographic modeling of SNP data and Bayesian phylogenetic analysis with multispecies coalescence on previous published gene sequences			Molina, J. et al. (2011)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		STRUCTURE, Bayesian Clustering, PCA	108 *O. r.* from native range	SNPs at 42 STSs	*Huang, P. et al. (2012)
	indica domesticated from Gangetic plains to continental SE Asia, S. China; [temperate race] japonica evolved from it in Yellow and Yangtze river basin; javanica evolved from indica in Malay archipelago	Review of archeological, morphology based studies		NA	Wet, J.M.J. (1981)
	"Snowballing" model: indica and japonica, derived from the hybrids between an early cultivar with a fixed set of alleles and the genetically divergent wild populations	Review			Sang, T. and Ge, S. (2007)
		Review			Sang, T. and Ge, S. (2007)
	japonica domesticated in Yangtze; proto-indica non-domesticated but eventually hybridizing with japonica varieties and receiving domestication alleles to make indica	Review; archeo-bot evidence			Fuller, D.Q. (2012)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	japonica domesticated in middle of Pearl River valley from Or-IIIa; indica developed by crossing btw ancient japonica and Or-I wild pops as rice cult. spread to S/SEA	Analysis of selective sweeps	446 wilds, 1083 O.s.	~5M SNPs from WGS (2X)	*Huang, X. et al. (2012)
Two independent domestications (di/polyphyletic)	Independent domestications of indica and japonica		234 sativa accessions	169 nuclear SSRs, 2 cp loci	Garris, A.J. et al. (2005)
					Rakshit, S. et al. (2007)
					Zhou SL. 1948. China is the place of origin of rice. J. Rice Soc. China 7: 53-54. [In Chinese.]
		intron seq of 4 genes -- div of ind and jap at 0.4mya			Zhu, Q. and Ge, S. (2005)
		LRT RT insertion - diver of ind and jap			Vitte, C. et al. (2004)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		at at least 0.2mya			
					Wang, Z.Y. et al. (1992)
		Then independent gene intro from jap -> ind. diff. expression analyses, alignment	1 jap-like perennial (W1943) and 1 ind-like perennial (W0106) *O. r.*, 1 indica, + Nippon, 93-11	transcriptome - 23-31K genes	Yang, C. et al. (2012)
			122 *O. r.*, 76 O.s.	48 nuclear RFLP	Sun, C.Q. et al. (1997)
		from partially overlapping gene pools - SD1 haplotype analysis	72 O.s. 42 *O. r.*	sd1 gene	Asano, K. et al. (2011)
	Supportive of two independent domestications of indica and japonica				Tang, T. et al. (2006)
					Nakano, M. et al. (1992)
		Ind and jap cp		cp RFLPs	Dally, A.M. and

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		differentiation found in *O. r.*			Second, G. (1990)
		Review			Kovach, M.J. et al. (2007)
		Div of ind and jap ~0.44mya	ind, Jap, glab	Indels, LRT-RT	Ma, J. and Bennetzen, J.L. (2004)
		Div of ind and jap ~0.72mya	1 ind, 1, jap, 1 O.r, 1 *O. n.*, +5 other AA genome sp	53 single-copy nuclear genes representing diverse functional categories, together with 16 intergenic regions	Zhu, T. et al. (2014)
		Review			Sweeney, M. and McCouch, S. (2007)
		nj, MP trees show ind and jap grouped rufi/niv	20 *Oryza* sp +3 outgroups; 1 ind, 1 jap. 1 niv, 3	3 cp regions	Kumagai, M. et al. (2010)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
			rufi		
		Letter contesting Molina et al, 2011 based on predomest. Div of ind, jap and arbitrary analysis rejection which could alt. stem from ancestral div. or back-introgression			Ge, S. and Sang, T. (2011)
		Domestication of *O. sativa* seemed to be diphyletic, since strong similarity was observed between *O. sativa* Japonica-Javanica and *O. r.* from China and between *O. sativa* Indica and *O. r.* from tropical Asia.	Twelve cultivars of *O. sativa*, one cultivar of O. glaberrima, and 17 wild accessions (12 *O. r.*, 2 O. glumaepatula, 1 O. longistaminata, 1 O. meridionalis and 1 O. barthii) were used	16 RAPD primers, 28 RFLP probes, 24 nuclear SSLP and 10 chloroplast SSLP	Bautista, N.S. et al. (2001)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		phenology +isozyme PCA	*O. r.* -- 28 China, 5 PnG, 10 Aust, 16 Amer; 20 O. longi, 20 brevi	24 isozyme loci	*Second, G. (1985)
		Ind and jap cp differentiation found in *O. r.*	70 Chinese *O. r.*, 27 Indian, 17 Thailand, 1-12 from other countries	69bp del in ORF100 only	Sun, C. et al. (1996)
	japonica in Yangtze river valley; indica in Ganges river basin				Second, G. (1982)
	Two independent domestications of indica and japonica; independent clustering of indica and japonica with different rufipogon and nivara accessions			416 polymorphic MITE-AFLP fragments	Park, K.C. et al. (2003)
	Indica from *O. rufipogon* (annual) and japonica from *O. rufipogon* (perennial)				Ohtsubo, H. et al. (2004)
					Ohtsubo, H. et al. (2008)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		Then independent gene intro from jap -> ind. diff. expression analyses, alignment	1 jap-like perennial (W1943) and 1 ind-like perennial (W0106) *O. r.*, 1 indica, + Nippon, 93-11	transcriptome - 23-31K genes	Yang, C. et al. (2012)
		most *O. r.* neg for pSINE-r2, ND ORF100, J allele for CMN; most nivara opposite	23*O. r.*/peren., 23 *O. n.*/ann. All from Cmb, Thai, Viet, Laos, 1 China; 1 ind, 1 jap	distribution of p-SINE1-r2 in the waxy locus, cp ORF100, CMN-A32 primer PCR product	Yamanaka, S. et al. (2003)
		but doesn't mention indp domes per se, just pre-domest diff.: 69bp deletion in ORF100 -- ann. & interm mostly del.	137 O.s.; 82 *O. r.*, 35 from other AA spp.		Chen, W. et al. (1993)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
		(86.5%); perennial mostly non-del (77.2%)			
		K=3; jap-like perennial, ind-like annual and interm, indp peren and mixed; *O. r.* - perennial originating from intermediate; annual from perennial.	68 O.s. (35 Ind, 33 Jap); 35 *O. r.* (13 ann, 16 peren, 6 intm.), 5 other O. spp	49 pSINE1	*Xu, J.-H. et al. (2007)
	Indica from annual in S. Asia; Japonica from O. rufi (perennial) in or close to Yangze River Valley	Review (japonica focus)			Bellwood, P. (2011)
	Indica from *O. n.* and japonica from *O. rufipogon*				Cheng, C. et al. (2003)
	Supportive of indica from *O. n.* (annual); japonica from O. rufi (perennial)	TESS, STRUCTURE, PCA, NJ	6 O.s., 5 O.mer, 43 *O. n.*, 47 *O. r.*, 24 interm. *O. r./O.*	29 SSRs	*Banaticla-Hilario, M.C.N. et al. (2013)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
			*n.*taxonomy assigned accd to life habit/morph		
	Indica likely from *O. n.*; japonica definitely from Chinese *O. r.*	phylogenetic analyses - FRAPPE, NJ, PCA	40 O.s., 5 *O. r.* (+10 SI), 5 *O. n.* (+5 SI)	6.5M SNPs from WGS	*Xu, X. et al. (2011)
	Maternal genome of indica from *O. n.* and japonica from *O. rufipogon* based on cp sequences	MP, NJ, ML, Bayesian Inf.	2 ind, 2 jap, 3 *O. r.*, 1 *O. n.*, 6 other AA genome spp.	221 parsimony informative variations from cp WGS	Wambugu, P.W. et al. (2015)
	new japonica domestication model- hybridization and selection btw two distant wild species			179 RFLPs and 3 FNPs	Izawa, T. (2008)
	tropical japonica domesticated in archipelagic SE Asia; indica domesticated independently	Allelic differences between qSW5, Wx, and qSH1			Shomura, A. et al. (2008)
		Allelic differences between 6 FNPs			Konishi, S. et al. (2008)
		Review			Izawa, T. et al. (2009)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	Proto-indica derived from *O. n.* (=prostrate), giving rise to indica ; japonica domesticated from *O. rufipogon*; Aus derived from hybridization between proto indica and japonica	Consideration of archeobotanical (spikelet bases and grain size) and genetic data (domestication alleles)			Fuller, D.Q. et al. (2010)
	Proto-indica derived from diverging *O. n.* (annual, extinct), with introgression of japonica alleles, giving rise to indica ; japonica domesticated from *O. rufipogon* (now extinct pops)	genetic markers and morphology - grain length		4cp, 6 nuclear markers	Castillo, C.C. *et al.* Archaeogenetic study of prehistoric rice remains from Thailand and India: evidence of early japonica in South and Southeast Asia. , *Archaeological and Anthropological Sciences*. (2015
	South and SEA CWR differentiated mainly into Indica; Chinese CWR differentiated into both indica and japonica	cp, mt, and nuclear markers			Sun, C. et al. (2002)
	Indica from trop Asia, jap from			57 subsp	Zhao, X. et al.

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	SE mainland China			specific intron length polymorphism markers	(2009)
Three or more independent domestications	*indica*, *japonica*, and *aus*, with japonica giving rise to *trop.* and *temp.*, and *aromatic* as a hybrid between japonica and aus	Analysis of 15 overlapping selective sweep regions and 38 co-located low-diversity genomic regions (CLDGRs)			Civáň, P. et al. (2015)
	japonica from wilds in South China; indica from wilds around Thailand, Myanmar, India = Indochina; possible independent domestication of aus in India			allelic var in one cp and 2 nuclear genes	*Londo, J.P. et al. (2006)
	indica and *japonica* separately; possible separate Aus and Indica domestication events from same gene pool				Caicedo, A. et al. (2007)
	Combination' model: diff wild pop in diff locations gave rise to				Sang, T. and Ge, S. (2007)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	early cultivars with different domest allele sets, then crosses betw those lead to fixation				Sang, T. and Ge, S. (2007)
Unclear/arbitrary support	*O. sativa* from *O. n.* (sensu stricto); possibly just indica from *O. n.*				Li, C. et al. (2006)
	Likely supportive of >1 domestication (India and China/broad belt btw)	Review			Khush, G.S. (1997)
	Nonindependent domestication of indica and japonica; at least partial sharing of their ancestral populations and/or recent gene flow between them				Gao, L.-Z. and Innan, H. (2008)
	More support for 2 or more domestications, but some for single, and thus still unclear	Review			Vaughan, D. et al. (2008)
	Supportive of single domestication, but mentions differences between genic and genomic domestication studies	monophyletic origin of sh4		sh4 and SH1 haplotypes	Zhang, L.L.-B. et al. (2009)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
	Polyphyletic domestication of indica and japonica from at least 4 different subgroups of *O. rufi/O.nivara* complex	Chloroplast haplotypes			Kawakami, S. et al. (2007)
	Likely supportive of >1 domestication, as wild japonica-like and indica-like genomes are recognized, but focus is on China. Also, written by an archeologist	Review			Gordon, B.C. (2010)
	Found sig diff in trop and subtrop pops of *O. r.* but not indica	InStruct clustering, D, Fst	21 indica, 50 jap, 3 *O. n.*, 13 *O. r.*, 1 O. off	4 photoperiod genes: PHYTOCHROME B (PhyB), HEADING DATE 1 (Hd1), HEADING DATE 3a (Hd3a), EARLY HEADING DATE 1	Huang, C.-L. et al. (2012)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
				(Ehd1	
	Difficult to determine domestication process b/c of introgression and gene flow	Book chapter/review			McCouch, S. et al. (2012)
	Difficult to determine -- a diffuse process in time and space	Review			Morishima, H. (2001)
	Some support of monophyletic hypo from *O. n.* grouping with O.s. and away from *O. r.*/O.peren.	UPGMA	11 O.s., 2 *O. r.*, 1 *O. n.*, 1 O. perenn., plus other *Oryza* spp.	11 ISSR polym.	Joshi, S.P. et al. (2000)
Related studies	Assumes indp. domestication; looks at introgression signatures from jap <->ind	WGS	21 jap, 22 ind, 23 *O. r.*		He, Z. et al. (2011)
	Assumes indp. domestication; looks at miRNA differences btw *O. r.* and O.s.		1 Chinese *O. r.* (Dongxiang) for denovo; 6 *O. r.*	387 miRNAs from WGS,	Wang, Y. et al. (2012)

Domestication hypothesis	Specifics - route/location	Validation methods	Germplasm	Markers	Citations*
			and 6 O.s. (3 ind/3 jap)	small RNA, degradome seq.	
* Denotes seminal or particularly important literature					
Abbreviations: indp. - independent; ind - Indica; jap - japonica; O.s. - *Oryza* sativa; *O. r.* - *O. rufipogon*; *O. n.*; *O. nivara*; O. mer - O. meridionalis; O. glab - O. glaberrima; O. glum - O. glumaepatula; O. longi - O. longistaminata; O. brevi - O. brevigulata; ann.=annual; peren. = perennial; intm =intermediate; cp = chloroplast; SEA - Southeast Asian;					

Supplementary Table 1.3 Passport information and SSR-based subpopulation identities of the *ORSC, O. sativa,* and O. meridionalis accessions analyzed in this study.

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
1	IRGC 3135	*O. sativa*	Agostano	_	_	_	_	_
5	IRGC 12440	*O. sativa*	Arc 10352	_	_	_	_	_
7	IRGC 43325	*O. sativa*	Arias	_	_	_	_	_
8	IRGC 6949	*O. sativa*	Asse Y Pung	_	_	_	_	_
12	IRGC 27805	*O. sativa*	Basmati	_	_	_	_	_
13	IRGC 27798	*O. sativa*	Basmati 1	_	_	_	_	_
14	IRGC 53637	*O. sativa*	Basmati 217	_	_	_	_	_
16	IRGC 38994	*O. sativa*	Bico Branco	_	_	_	_	_
18	IRGC 45195	*O. sativa*	Bj 1	_	_	_	_	_
28	IRGC 30238	*O. sativa*	Champa Tong 54	_	_	_	_	_
30	IRGC 10214	*O. sativa*	Chiem Chanh	_	_	_	_	_
31	PI 431222	*O. sativa*	Chinese	_	_	_	_	_
43	PI 279131	*O. sativa*	Dee Geo Woo Gen	_	_	_	_	_
44	IRGC 3686	*O. sativa*	Dhala Shaitta	_	_	_	_	_
45	PI 584607	*O. sativa*	Dom-Sofid	_	_	_	_	_
49	IRGC 8839	*O. sativa*	Dv85	_	_	_	_	_
53	IRGC 39261	*O. sativa*	Firooz	_	_	_	_	_
60	IRGC 43397	*O. sativa*	Gotak Gatik	_	_	_	_	_
61	IRGC 51300	*O. sativa*	Guan-Yin-Tsan	_	_	_	_	_
76	IRGC 11099	*O. sativa*	Jaya	_	_	_	_	_
78	IRGC 6307	*O. sativa*	Jhona 349	_	_	_	_	_
85	HO1195	*O. sativa*	Kasalath	_	_	_	_	_
88	IRGC 24224	*O. sativa*	Khao Gaew	_	_	_	_	_
93	IRGC 12793	*O. sativa*	Kitrana 508	_	_	_	_	_
94	PI 330464	*O. sativa*	Koshihikari	_	_	_	_	_
95	IRGC 2545	*O. sativa*	Kotobuki Mochi	_	_	_	_	_
101	IRGC 66756	*O. sativa*	Lemont	_	_	_	_	_
104	IRGC 8191	*O. sativa*	Mansaku	_	_	_	_	_

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
107	IRGC 25901	*O. sativa*	Miriti	_	_	_	_	_
108	IRGC 12048	*O. sativa*	Moroberekan	_	_	_	_	_
110	IRGC 6663	*O. sativa*	Mudgo	_	_	_	_	_
113	IRGC 418	*O. sativa*	Norin 20	_	_	_	_	_
131	IRGC 32399	*O. sativa*	Phudugey	_	_	_	_	_
132	IRGC 8952	*O. sativa*	Rathuwee	_	_	_	_	_
143	CIor 1642	*O. sativa*	Shinriki	_	_	_	_	_
144	PI 392539	*O. sativa*	Shoemed	_	_	_	_	_
151	PI 597021	*O. sativa*	Suweon	_	_	_	_	_
152	IRGC 6294	*O. sativa*	T 1	_	_	_	_	_
161	PI 536047	*O. sativa*	Teqing	_	_	_	_	_
163	PI 280681	*O. sativa*	Taducan	_	_	_	_	_
165	IRGC 43675	*O. sativa*	Trembese	_	_	_	_	_
173	PI 514663	*O. sativa*	Nipponbare	_	_	_	_	_
174	IRGC 328	*O. sativa*	Azucena	_	_	_	_	_
398		*O. sativa*	93-11	_	_	_	_	_
401	IRGC 80433	*O. rufipogon*	_	Admix	0.005	0.5782	0.0069	0.4099
402	IRGC 80539	*O. spontanea*	_	Admix	0.0044	0.4938	0.0244	0.4774
403	IRGC 80562	*O. rufipogon*	_	Admix	0.0046	0.362	0.2386	0.3948
404	IRGC 80582	*O. nivara*	_	Admix	0.0042	0.5007	0.0022	0.4929
405	IRGC 80586	*O. spontanea*	_	Admix	0.0025	0.4857	0.0296	0.4822
406	IRGC 80592	*O. rufipogon*	_	R2	0.0046	0.7712	0.072	0.1521
407	IRGC 80742	*O. rufipogon*	_	R2	0.0798	0.7087	0.0061	0.2053
408	IRGC 80745	*O. spontanea*	_	Admix	0.0104	0.5377	0.015	0.4369
410	IRGC 80759	*O. nivara*	_	R3	0.0041	0.1078	0.8765	0.0116
411	IRGC 81801	*O. rufipogon*	_	R2	0.0097	0.897	0.0028	0.0905
412	IRGC 81802	*O. rufipogon*	_	R2	0.1082	0.8844	0.0024	0.005
413	IRGC 81850	*O. nivara*	_	R3	0.0013	0.0021	0.9946	0.002
414	IRGC 81903	*O. spontanea*	_	Admix	0.0056	0.3956	0.1846	0.4143
415	IRGC 81909	*O. spontanea*	_	R3	0.0013	0.0027	0.99	0.006
416	IRGC 81970	*O. spontanea*	_	Admix	0.0049	0.4969	0.0034	0.4948
417	IRGC 81976	*O. rufipogon*	_	R2	0.0092	0.9451	0.0314	0.0143
418	IRGC 81977	*O. rufipogon*	_	R2	0.0238	0.9692	0.0038	0.0032
420	IRGC 81984	*O. rufipogon*	_	R2	0.0234	0.7205	0.0108	0.2453
421	IRGC 81990	*O. rufipogon*	_	R2	0.0038	0.9878	0.0059	0.0025
422	IRGC 81993	*O. rufipogon*	_	R2	0.003	0.874	0.0055	0.1175
423	IRGC 81994	*O. rufipogon*	_	R1	0.8022	0.1814	0.005	0.0114

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
424	IRGC 81996	*O. rufipogon*	_	R1	0.9011	0.0915	0.0035	0.0039
425	IRGC 82040	*O. rufipogon*	_	R2	0.1065	0.6289	0.0044	0.2602
426	IRGC 82979	*O. rufipogon*	_	R2	0.0107	0.8957	0.0341	0.0595
427	IRGC 82988	*O. rufipogon*	_	Admix	0.0401	0.1972	0.192	0.5707
428	IRGC 82989	*O. rufipogon*	_	Admix	0.0042	0.3027	0.4111	0.2819
429	IRGC 82990	*O. rufipogon*	_	R2	0.0055	0.8313	0.0089	0.1543
430	IRGC 82991	*O. rufipogon*	_	R4	0.0023	0.1361	0.1557	0.7059
431	IRGC 82992	*O. rufipogon*	_	Admix	0.0069	0.406	0.3466	0.2405
432	IRGC 83794	*O. rufipogon*	_	R2	0.0038	0.7703	0.0511	0.1748
433	IRGC 83795	*O. rufipogon*	_	Admix	0.0056	0.4227	0.0578	0.5139
434	IRGC 83823	*O. rufipogon*	_	R2	0.0868	0.8974	0.0114	0.0044
435	IRGC 86448	*O. rufipogon*	_	R2	0.003	0.8012	0.0097	0.1861
436	IRGC 86454	*O. rufipogon*	_	R2	0.0606	0.9304	0.0029	0.0061
437	IRGC 86475	*O. rufipogon*	_	Admix	0.0023	0.4499	0.3312	0.2167
438	IRGC 86476	*O. rufipogon*	_	R2	0.1721	0.6528	0.0449	0.1302
440	IRGC 88787	*O. rufipogon*	_	R2	0.0168	0.6319	0.0309	0.3204
441	IRGC 92605	*O. rufipogon*	_	R2	0.0026	0.9859	0.0037	0.0078
442	IRGC 93181	*O. nivara*	_	R3	0.002	0.0081	0.9864	0.0035
443	IRGC 93183	*O. nivara*	_	R3	0.0024	0.0124	0.8787	0.1065
444	IRGC 93188	*O. nivara*	_	R3	0.0014	0.0044	0.9795	0.0147
446	IRGC 93224	*O. spontanea*	_	R3	0.0017	0.0017	0.9949	0.0017
447	IRGC 93274	*O. rufipogon*	_	R1	0.73	0.1578	0.0175	0.0947
449	IRGC 100195	*O. nivara*	_	R3	0.0022	0.3656	0.6209	0.0113
450	IRGC 100916	*O. rufipogon*	_	R3	0.0055	0.0109	0.9815	0.0021
451	IRGC 101508	*O. nivara*	_	R3	0.0065	0.1322	0.6752	0.1861
452	IRGC 103308	*O. rufipogon*	_	Admix	0.0019	0.0996	0.3074	0.5911
453	IRGC 103404	*O. rufipogon*	_	Admix	0.0035	0.3308	0.3765	0.2892
454	IRGC 103821	*O. nivara*	_	R4	0.0062	0.2615	0.0947	0.6376
455	IRGC 103823	*O. rufipogon*	_	Admix	0.1305	0.3987	0.0788	0.392
456	IRGC 103824	*O. nivara*	_	Admix	0.002	0.389	0.1466	0.4624
457	IRGC 103838	*O. nivara*	_	Admix	0.0023	0.365	0.1144	0.5184
458	IRGC 103844	*O. rufipogon*	_	R2	0.0064	0.7224	0.0043	0.2669
459	IRGC 103847	*O. rufipogon*	_	R2	0.0025	0.915	0.0137	0.0688
460	IRGC 103848	*O. rufipogon*	_	R2	0.0226	0.6513	0.0049	0.3212
461	IRGC 104057	*O. rufipogon*	_	Admix	0.002	0.3573	0.0764	0.5644
462	IRGC 104501	*O. rufipogon*	_	R4	0.0019	0.0609	0.1799	0.7573
463	IRGC 104599	*O. rufipogon*	_	R3	0.0021	0.0192	0.9693	0.0094

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
464	IRGC 104602	*O. rufipogon*	_	R2	0.0127	0.8342	0.0164	0.1367
465	IRGC 104620	*O. spontanea*	_	R4	0.0092	0.1132	0.1718	0.7058
467	IRGC 104624	*O. rufipogon*	_	R4	0.0041	0.2128	0.1453	0.6378
468	IRGC 104626	*O. spontanea*	_	R4	0.0032	0.1817	0.1864	0.6287
469	IRGC 104628	*O. spontanea*	_	R4	0.0322	0.1309	0.2341	0.6028
470	IRGC 104632	*O. spontanea*	_	Admix	0.0517	0.27	0.1466	0.5317
471	IRGC 104634	*O. spontanea*	_	R4	0.0047	0.2983	0.0063	0.6907
472	IRGC 104636	*O. spontanea*	_	R4	0.0038	0.1604	0.1521	0.6837
473	IRGC 104644	*O. nivara*	_	Admix	0.0018	0.516	0.0038	0.4784
474	IRGC 104714	*O. rufipogon*	_	R2	0.0021	0.9561	0.0275	0.0143
475	IRGC 104823	*O. nivara*	_	Admix	0.024	0.5818	0.0314	0.3628
476	IRGC 104959	*O. spontanea*	_	Admix	0.0029	0.228	0.1899	0.5792
477	IRGC 104967	*O. spontanea*	_	Admix	0.0082	0.4761	0.0138	0.5019
478	IRGC 104971	*O. spontanea*	_	Admix	0.0022	0.2487	0.1697	0.5793
479	IRGC 105220	*O. officinalis*	_	_	_	_	_	_
480	IRGC 105250	*O. rufipogon*	_	R2	0.025	0.8922	0.0683	0.0145
481	IRGC 105343	*O. nivara*	_	R3	0.0037	0.0033	0.9903	0.0027
482	IRGC 105349	*O. rufipogon*	_	R3	0.0016	0.0017	0.9952	0.0015
483	IRGC 105375	*O. rufipogon*	_	R3	0.003	0.0021	0.9929	0.002
484	IRGC 105388	*O. rufipogon*	_	R2	0.0031	0.6526	0.0042	0.3401
485	IRGC 105400	*O. rufipogon*	_	Admix	0.0509	0.1361	0.2727	0.5403
486	IRGC 105402	*O. rufipogon*	_	Admix	0.0053	0.2294	0.1998	0.5655
487	IRGC 105428	*O. nivara*	_	R3	0.0024	0.0078	0.9727	0.0171
488	IRGC 105491	*O. rufipogon*	_	Admix	0.0017	0.4949	0.0095	0.4939
489	IRGC 105564	*O. spontanea*	_	R2	0.0072	0.648	0.0204	0.3244
490	IRGC 105567	*O. rufipogon*	_	R2	0.0024	0.8782	0.003	0.1164
493	IRGC 105706	*O. nivara*	_	R3	0.0014	0.0051	0.9799	0.0136
494	IRGC 105711	*O. rufipogon*	_	R3	0.0028	0.0044	0.9903	0.0025
495	IRGC 105717	*O. nivara*	_	Admix	0.0771	0.4963	0.098	0.3287
496	IRGC 105720	*O. rufipogon*	_	R2	0.2079	0.6005	0.0073	0.1843
497	IRGC 105726	*O. rufipogon*	_	R2	0.087	0.8971	0.0095	0.0064
498	IRGC 105735	*O. rufipogon*	_	R2	0.0029	0.6326	0.0025	0.362
499	IRGC 105767	*O. rufipogon*	_	R2	0.0093	0.7082	0.0033	0.2792
500	IRGC 105785	*O. nivara*	_	Admix	0.0015	0.4992	0.0033	0.496
501	IRGC 105821	*O. nivara*	_	Admix	0.0018	0.4389	0.1116	0.4477
503	IRGC 105843	*O. rufipogon*	_	R2	0.018	0.7065	0.0036	0.2719
504	IRGC 105847	*O. rufipogon*	_	R2	0.0025	0.8281	0.0268	0.1426

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
505	IRGC 105855	*O. rufipogon*	_	R2	0.0036	0.6437	0.065	0.2877
507	IRGC 105881	*O. rufipogon*	_	R2	0.0041	0.7178	0.0125	0.2656
508	IRGC 105890	*O. rufipogon*	_	R2	0.0131	0.8374	0.123	0.0265
509	IRGC 105897	*O. rufipogon*	_	Admix	0.0018	0.5063	0.0018	0.4901
510	IRGC 105898	*O. rufipogon*	_	R2	0.002	0.6559	0.3231	0.019
511	IRGC 105909	*O. rufipogon*	_	R2	0.0087	0.8981	0.0677	0.0254
512	IRGC 105942	*O. rufipogon*	_	R2	0.0024	0.9789	0.0082	0.0105
513	IRGC 105951	*O. rufipogon*	_	R2	0.0629	0.9275	0.0061	0.0035
514	IRGC 105956	*O. rufipogon*	_	Admix	0.149	0.5469	0.005	0.2991
515	IRGC 105958	*O. rufipogon*	_	R2	0.0045	0.9735	0.0142	0.0078
516	IRGC 106036	*O. rufipogon*	_	Admix	0.212	0.4479	0.0035	0.3366
517	IRGC 106057	*O. rufipogon*	_	Admix	0.0021	0.5489	0.0161	0.4329
518	IRGC 106078	*O. rufipogon*	_	Admix	0.0024	0.5135	0.0149	0.4692
519	IRGC 106115	*O. rufipogon*	_	R2	0.0035	0.9338	0.0136	0.0491
520	IRGC 106144	*O. rufipogon*	_	R2	0.0031	0.6071	0.0909	0.2989
521	IRGC 106145	*O. rufipogon*	_	R2	0.0051	0.9782	0.0039	0.0128
522	IRGC 106150	*O. rufipogon*	_	R2	0.0075	0.953	0.0047	0.0348
523	IRGC 106155	*O. nivara*	_	Admix	0.0029	0.4971	0.0092	0.4908
524	IRGC 106156	*O. rufipogon*	_	R2	0.0024	0.8908	0.0093	0.0975
525	IRGC 106161	*O. rufipogon*	_	R2	0.0042	0.7943	0.0026	0.1989
526	IRGC 106163	*O. rufipogon*	_	R2	0.1429	0.8392	0.0041	0.0138
528	IRGC 106168	*O. rufipogon*	_	Admix	0.0038	0.2934	0.2786	0.4242
529	IRGC 106169	*O. rufipogon*	_	R2	0.0148	0.7059	0.0033	0.276
530	IRGC 106273	*O. rufipogon*	_	R1	0.9921	0.0024	0.0026	0.0029
531	IRGC 106283	*O. rufipogon*	_	R1	0.9957	0.0015	0.0015	0.0013
533	IRGC 106327	*O. rufipogon*	_	R2	0.0542	0.6416	0.0057	0.2985
534	IRGC 106332	*O. rufipogon*	_	R2	0.0168	0.7321	0.0024	0.2487
535	IRGC 106342	*O. rufipogon*	_	R2	0.0115	0.9428	0.0205	0.0252
536	IRGC 106357	*O. rufipogon*	_	R2	0.0078	0.8411	0.0199	0.1312
538	IRGC 106410	*O. rufipogon*	_	Admix	0.002	0.5165	0.0348	0.4467
539	IRGC 106412	*O. rufipogon*	_	R2	0.0382	0.8367	0.0159	0.1091
540	IRGC 106413	*O. rufipogon*	_	R2	0.064	0.7776	0.0023	0.1561
541	IRGC 106414	*O. rufipogon*	_	R2	0.0512	0.9047	0.0248	0.0193
543	IRGC 106420	*O. rufipogon*	_	R2	0.0393	0.8864	0.0681	0.0062
545	IRGC 106453	*O. rufipogon*	_	R2	0.0025	0.9403	0.0065	0.0507
546	IRGC 106509	*O. rufipogon*	_	R2	0.0111	0.9785	0.0045	0.0059
547	IRGC 105908	*O. rufipogon*	_	R2	0.2118	0.762	0.008	0.0182

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
549	IRGC 81881	*O. rufipogon*	_	Admix	0.0073	0.4277	0.2591	0.3058
550	IRGC 81887	*O. rufipogon*	_	R2	0.0028	0.6187	0.0818	0.2967
551	IRGC 100596	*O. r. x O. n.*	_	Admix	0.0022	0.1018	0.2974	0.5986
552	IRGC 100920	*O. r. x O. n.*	_	Admix	0.003	0.5135	0.0051	0.4784
553	IRGC 100926	*O. rufipogon*	_	Admix	0.0015	0.5107	0.0026	0.4852
554	IRGC 103305	*O. rufipogon*	_	R2	0.0021	0.9592	0.0244	0.0143
555	IRGC 105349	*O. rufipogon*	_	R3	0.0023	0.0022	0.9932	0.0023
556	IRGC 105494	*O. rufipogon*	_	Admix	0.0085	0.5074	0.0058	0.4783
557	IRGC 105567	*O. rufipogon*	_	Admix	0.0358	0.4009	0.2982	0.2652
558	IRGC 105616	*O. rufipogon*	_	Admix	0.0024	0.4999	0.009	0.4887
559	IRGC 105618	*O. rufipogon*	_	R4	0.0092	0.2847	0.088	0.6181
560	IRGC 105726	*O. rufipogon*	_	R2	0.0384	0.9287	0.0125	0.0204
561	IRGC 105868	*O. rufipogon*	_	R2	0.0251	0.8424	0.0331	0.0994
562	IRGC 105890	*O. rufipogon*	_	R2	0.1357	0.6624	0.1346	0.0673
563	IRGC 105951	*O. rufipogon*	_	R2	0.0279	0.869	0.0116	0.0915
565	IRGC 106144	*O. rufipogon*	_	R2	0.0388	0.8633	0.041	0.0569
566	IRGC 106163	*O. rufipogon*	_	R2	0.157	0.8283	0.0023	0.0124
567	IRGC 106167	*O. rufipogon*	_	Admix	0.0394	0.2965	0.1123	0.5518
568	IRGC 106263	*O. rufipogon*	_	Admix	0.0356	0.5378	0.1046	0.322
569	IRGC 106264	*O. rufipogon*	_	R1	0.9937	0.0027	0.0022	0.0014
570	IRGC 106266	*O. rufipogon*	_	Admix	0.5029	0.3053	0.0019	0.1899
574	IRGC 106270	*O. rufipogon*	_	R3	0.2896	0.0062	0.6855	0.0187
575	IRGC 106272	*O. rufipogon*	_	R1	0.9943	0.0023	0.0015	0.0019
576	IRGC 106273	*O. rufipogon*	_	R1	0.9964	0.0013	0.001	0.0013
577	IRGC 106274	*O. rufipogon*	_	R1	0.9935	0.0026	0.0024	0.0015
578	IRGC 106275	*O. rufipogon*	_	R1	0.9923	0.0026	0.0033	0.0018
579	IRGC 106276	*O. rufipogon*	_	R1	0.9959	0.0013	0.0014	0.0014
581	IRGC 106278	*O. rufipogon*	_	Admix	0.4433	0.3219	0.2304	0.0044
582	IRGC 106279	*O. rufipogon*	_	R1	0.993	0.0033	0.0015	0.0022
583	IRGC 106280	*O. rufipogon*	_	R1	0.9555	0.0236	0.002	0.0189
584	IRGC 106282	*O. rufipogon*	_	R1	0.9815	0.0126	0.0015	0.0044
585	IRGC 106283	*O. rufipogon*	_	R1	0.9958	0.0015	0.0013	0.0014
587	IRGC 106285	*O. rufipogon*	_	R1	0.9903	0.0024	0.0029	0.0044
588	IRGC 106286	*O. rufipogon*	_	Admix	0.4543	0.3577	0.0498	0.1382
591	IRGC 106290	*O. rufipogon*	_	R1	0.9917	0.0033	0.0026	0.0024
592	IRGC 80671	*O. rufipogon*	_	R2	0.0093	0.6471	0.0085	0.3351
593	IRGC 105757	*O. rufipogon*	_	R1	0.9959	0.0017	0.0011	0.0013

NSF-TV ID	Accession #	Species	Accession Name	SSR-based Structure group percent identity (*ORSC* only)				
				K=4 subpop[1]	W1	W2	W3	W4
594	IRGC 106412	*O. rufipogon*	_	R1	0.7008	0.2593	0.0164	0.0235
599	IRGC 100183	*O. r. x O. s.*	_	R2	0.0028	0.6867	0.0085	0.302
600	IRGC 100187	*O. s. x O. r.*	_	Admix	0.0088	0.5137	0.0039	0.4736
602	IRGC 100900	*O. nivara*	_	R3	0.0178	0.0078	0.9663	0.0081
604	IRGC 100907	*O. r. x O. n.*	_	R4	0.0021	0.1019	0.2519	0.6441
605	IRGC 100911	*O. spontanea*	_	Admix	0.0046	0.4972	0.0024	0.4958
628	PI 593892	*O. sativa*	Jefferson	_	_	_	_	_
676	IRGC 100692	*O. rufipogon*	_	R4	0.0014	0.1014	0.2478	0.6494
700	IRGC 103423	*O. nivara*	_	R4	0.0013	0.343	0.0099	0.6458
701	IRGC 103813	*O. nivara*	_	R4	0.0024	0.205	0.1536	0.639
744	IRGC 105716	*O. nivara*	_	Admix	0.0127	0.5303	0.0422	0.4148

[1] 50% admix cutoff designation

Abbreviations: *O. r. - O. rufipogon; O. n. - O. nivara; O.s. - O. sativa*

Supplementary Table 1.4 SSR marker information

Name	Chr.	Forward primer	Reverse primer
RM1	1	GCGAAAACACAATGCAAAAA	GCGTTGGTTGGACCTGAC
RM11	7	TCTCCTCTTCCCCCGATC	ATAGCGGGCGAGGCTTAG
RM116	11	TCACGCACAGCGTGCCGTTCTC	CAAGATCAAGCCATGAAAGGAGGG
RM118	7	CCAATCGGAGCCACCGGAGAGC	CACATCCTCCAGCGACGCCGAG
RM124	4	ATCGTCTGCGTTGCGGCTGCTG	CATGGATCACCGAGCTCCCCCC
RM125	7	ATCAGCAGCCATGGCAGCGACC	AGGGGATCATGTGCCGAAGGCC
RM133	6	TTGGATTGTTTTGCTGGCTCGC	GGAACACGGGGTCGGAAGCGAC
RM136	6	GAGAGCTCAGCTGCTGCCTCTAGC	GAGGAGCGCCACGGTGTACGCC
RM142	4	CTCGCTATCGCCATCGCCATCG	TCGAGCCATCGCTGGATGGAGG
RM152	8	GAAACCACCACACCTCACCG	CCGTAGACCTTCTTGAAGTAG
RM154	2	ACCCTCTCCGCCTCGCCTCCTC	CTCCTCCTCCTGCGACCGCTCC
RM161	5	TGCAGATGAGAAGCGGCGCCTC	TGTGTCATCAGACGGCGCTCCG
RM162	6	GCCAGCAAAACCAGGGATCCGG	CAAGGTCTTGTGCGGCTTGCGG
RM169	5	TGGCTGGCTCCGTGGGTAGCTG	TCCCGTTGCCGTTCATCCCTCC
RM171	10	AACGCGAGGACACGTACTTAC	ACGAGATACGTACGCCTTTG
RM178	5	TCGCGTGAAAGATAAGCGGCGC	GATCACCGTTCCCTCCGCCTGC
RM208	2	TCTGCAAGCCTTGTCTGATG	TAAGTCGATCATTGTGTGGACC
RM214	7	CTGATGATAGAAACCTCTTCTC	AAGAACAGCTGACTTCACAA
RM215	9	CAAAATGGAGCAGCAAGAGC	TGAGCACCTCCTTCTCTGTAG
RM22	3	GGTTTGGGAGCCCATAATCT	CTGGGCTTCTTTCACTCGTC
RM237	1	CAAATCCCGACTGCTGTCC	TGGGAAGAGAGCACTACAGC
RM247	12	TAGTGCCGATCGATGTAACG	CATATGGTTTTGACAAAGCG
RM261	4	CTACTTCTCCCCTTGTGTCG	TGTACCATCGCCAAATCTCC
RM271	10	TCAGATCTACAATTCCATCC	TCGGTGAGACCTAGAGAGCC
RM277	12	CGGTCAAATCATCACCTGAC	CAAGGCTTGCAAGGGAAG
RM279	2	GCGGGAGAGGGATCTCCT	GGCTAGGAGTTAACCTCGCG
RM284	8	ATCTCTGATACTCCATCCATCC	CCTGTACGTTGATCCGAAGC
RM287	11	TTCCCTGTTAAGAGAGAAATC	GTGTATTTGGTGAAAGCAAC
RM307	4	GTACTACCGACCTACCGTTCAC	CTGCTATGCATGAACTGCTC
RM310	8	CCAAAACATTTAAAATATCATG	GCTTGTTGGTCATTACCATTC
RM316	9	CTAGTTGGGCATACGATGGC	ACGCTTATATGTTACGTCAAC
RM338	3	CACAGGAGCAGGAGAAGAGC	GGCAAACCGATCACTCAGTC
RM408	8	CAACGAGCTAACTTCCGTCC	ACTGCTACTTGGGTAGCTGACC
RM413	5	GGCGATTCTTGGATGAAGAG	TCCCCACCAATCTTGTCTTC
RM431	1	TCCTGCGAACTGAAGAGTTG	AGAGCAAAACCCTGGTTCAC
RM44	8	ACGGGCAATCCGAACAACC	TCGGGAAAACCTACCCTACC
RM447	8	CCCTTGTGCTGTCTCCTCTC	ACGGGCTTCTTCTCCTTCTC
RM452	2	CTGATCGAGAGCGTTAAGGG	GGGATCAAACCACGTTTCTG
RM454	6	CTCAAGCTTAGCTGCTGCTG	GTGATCAGTGCACCATAGCG
RM474	10	AAGATGTACGGGTGGCATTC	TATGAGCTGGTGAGCAATGG
RM475	2	CCTCACGATTTTCCTCCAAC	ACGGTGGGATTAGACTGTGC
RM484	10	TCTCCCTCCTCACCATTGTC	TGCTGCCCTCTCTCTCTCTC
RM495	1	AATCCAAGGTGCAGAGATGG	CAACGATGACGAACACAACC
RM5	1	TGCAACTTCTAGCTGCTCGA	GCATCCGATCTTGATGGG
RM507	5	CTTAAGCTCCAGCCGAAATG	CTCACCCTCATCATCGCC
RM536	11	TCTCTCCTCTTGTTTGGCTC	ACACACCAACACGACCACAC
RM55	3	CCGTCGCCGTAGTAGAGAAG	TCCCGGTTATTTTAAGGCG

Supplementary Table 1.5 MITE marker information

Marker	Forward primer	Reverse primer	Chr	Genetic location (Mb)	MSU V7. start	MSU V7. end	Length (bp)
RT4	GGGCATGTTTAAATGTTTTGGTTC	GAAACTCGAGTAAACTACGCCCAC	1	12.300977	12461155	12461699	544
RT5	CCCTTATTGACACCGATTGAGAAC	ATGGCTTAATTTGGACCCTTTTTG	1	4.664463	4666661	4667169	508
RT11	GGAATAGCTCATAGCTGGTTGTGG	AATTCTATCGAAAGCACGCCATAC	1	28.307086	29736084	29736462	378
RT38	GGTACCGAACACCATAGTACAACA	GGTGCTATCGTTTGCATGTATATT	2	19.751624	20615354	20615570	216
RT12	CGGAAAACGAGAGAGGTGAGTTAG	TGCACCCAAATATTCTGTCACAAG	3	3.016256	3015146	3015504	358
RT2	AAGGTTTGTCCCTTTCTCTGTTCC	GTTAGTGGTTGCTGTTGCTGTGAG	4	33.083767	33701369	33701889	520
RT3	AGCTCTTGCATGAGAGCTAACGTC	GCCCTGATGAGTAAAAATTCTCCC	4	29.217579	29835178	29835625	447
RT27	TTAACCTCTTTGTGATCGATCGTG	TGTACTACACCCCTCATCCTCCTC	4	30.264356	30881958	30882503	545
RT34	ATGGAGTTTTAATTGATGTATGC	AAATCCTACTGGAATTATATTTTTG	5	5.389621	5392325	5392852	527
RT35	GCCTTGAAACATGTCCACAC	AGAGGCAAGAGCTACTCCAAAC	5	20.770384			
RT40	AGAGGCAAGAGCTACTCCAAAC	GCCTTGAAACATGTCCACAC	5	3.199966	3199819	3200013	194
RT1	TTCATGCAGGTGTTTAAATGTTCG	AAAACATTTTGAAATCCGTGTTGC	6	26.389887	27267441	27267910	469
RT14	ATTGTAGCATTAATTCCGACAGGG	ATCTCCGTTTTCGTTTTGTTTCAG	6	20.143736	21021736	21022125	389
RT41	TATACCCACTTTATCCCATTGC	ACGATTTCAGTGACCTCATCA	6	4.419717	4419717	4419933	216
RT17	GCAGTCGGACACGACATGTTATAG	AATGCTTAACTAGGGCTTGCAGTG	6	2.408626	2408626	2409208	582
RT18	CTGCAAACGGTCTGAATCTAAAGG	CCTAATCCTGGTTTTGGTCTGATG	6	28.014502	28892502	28893123	621
RT23	CTATCCCTACAAAGGAACAGCACG	CTATCCCTACAAAGGAACAGCACG	7	24.095825	24757677	24757700	23
RT24	TTAAGTTGGGAAAGCCAACAAAAG	ATAAAGTCGATGCTTGTGTGATGC	7	0.10105	101751	102259	508
RT28	ATCTGCTCCAAATCTTTCACCAAG	ATTATATTACTTGGGGGCCTCTGC	7	23.493297	24155149	24155774	625
RT7	AATTTCTGCAAACAGCTGAGATCC	TTTCCTCTTTGCTCTGACGACTTC	7	27.176383	27838235	27838741	506
RT20	TCTTAGTTACATCGTCTGCCGAATC	AATCAGACTGGACCCACCTACTTG	7	18.61537	19277222	19277821	599
RT13	TGACCGGGTTCAGTAGTAGCAGTAG	ATGTGTATGACAGGTGTGACCAGG	8	22.89221	23113058	23113482	424
RT36	CACGCCTGGTGATGTATGATAA	TAAAACATGCATGGCAAAGAACTA	8	12.366413	12460760	12460973	213
RT19	GCGTAAAAACAGGGAAGTGAGTTG	CCATGGCTAGCTCTCTTTCTCATC	8	4.088953	4094300	4094873	573

RT37	GAGGAAGTTAGGGTTGGGACATTC	GCAGTAAAAGGAAGAAGAGGGAGC	8	16.206417	16427265	16427652	387
RT16	TAACCTGCTTTCTTCTTCCTTCCC	AAATGGCAGTATAGCCATGAGGAG	9	13.657	14312170	14312546	376
RT9	TGTTGTTCGCCGAAATTCTTTTAG	CAGTGATACGTCCCTTAGGATTGG	9	11.928946	12534658	12535274	616
RT10	ACCCCCTTAATAACTCTTCGCTTG	ACTCTGTTCTGGACTCTGTTGTGC	9	0.942461	890173	890744	571
RT25	CGCATGCTCTGAGTAATGTAAACG	CCTCTCCTCTCTCCTTCACACAAG	10	10.669	11257417	11257918	501
RT26	TGCAAAGGGTAGGAATGTCAAAAC	CACGTCCTTGCTACTGCACTTATG	10	2.775	2769724	2770225	501
RT6	TAGCCAGGAAGGAAGGAGAAAGAG	AGACAAAGGGCCTGTTTAGTTTCC	10	22.546665	23331630	23332043	413
RT30	AGAAACTGCAGGGTACTTGATGTG	CTCGGCCAATTACAATCTTCTCTC	10	18.449751	19236558	19237131	573
RT8	TATCATTAGTCGTCGCCCTCTCTC	CTAGGACGGGAGGTTATTTGTTCC	10	17.702071	18488673	18489250	577
RT31	GAAAAGCAATTAAGCACCAACCAC	AAGGAACAGTTATTGCTCTCACGC	10	6.878587	6794251	6794855	604
RT32	AGTCATCACAAATGACGATCGAAG	GAAATGCACTGAATGTCAAACTGC	10	18.663343	19450150	19450728	578
RT21	GCCCTTATTCAGGGACTACCAATC	CATGGCTTAAATTTGCACAAACAG	10	7.215087	7183260	7183828	568
RT33	AAAATGTTGAGCTCCCCTAAAAGC	ATTTCATGTCTCTTTCGTGGTTCG	10	8.267156	8795852	8796424	572
RT22	CGTGATATTGGAGTGTTGGACTTG	ACTCTCCTTTTGACATTTGCCATC	11	21.350731	23694588	23695215	627
RT15	GTATCTCGGTTAATGTGAGCCGTC	ACTCTCCTTTTGACATTTGCCATC	11	22.028	23695192	24372216	677024
RT39	TTACTTCAGCTGTACCCGTAGC	ATTCTCGCCTCTTGGTTCTC	12	22.037757	22274728	22274894	166
RT29	ACATATGAGATTTGGCTTTGTCACC	AGTCTCACCAAATCATAAGCAGCC	12	6.212944	6213106	6213705	599

Supplementary Table 1.6 SINE marker information

Marker	Forward primer	Reverse primer
SINE2	TTGCCCGGATACTTCTCCTC	GGAGGACGTCCAGATCGTTC
SINE25	GGATGGCTTCAGCAGGATCA	TTCTGACAGGGAATCAAATG
SINE29	CTACACTGCTAGTGGTGCTG	TTCACCAACTCTGTCAAATG
SINE30	CCACATAAGTGCTATGTAGT	GGGCTCCGTCTAGTATACCG
SINE32	AGTACAGAAGGTAATCACGT	AACTGACTCTTATTAGACTGG
SINE34	GGACCATTCTTGCACAAAGA	GGGATCACATCATCGTGCCA
SINE51	TCAGCATCTTTCAGAAGCCT	TACCACGGCTAGCTAGCAAC
SINE52	TTGATCCCGTGTGATCGTGT	GATACCTATTCGGCATGCTC
SINE54	ATGACGAAGCAAAGAGCAGA	GGTAAAGGCTTTACCAGTGT
SINE56	AACTCGAGGGCAATGACAGT	ACAGTACTCGAACACGTTAG
SINE57	AGCTAGCCATTAGTTGTGAG	AACAAGGCAGAGCCGACGAT
SINE58	ACCATAGCTGATGTACTTGT	TTGGCCTTGATCCAGAACCA
SINE59	GGACAGCACATTGATCAGCC	TCGATCGATGGGAGCAGTTC
SINE60	GACGGTCGTTGTTGTGCGGT	ATCGGTATTTCAACATCCC
SINE61	CTTGTGCCGCGTGATCCGGG	GCATCATGGCCACGTCGGTTT
SINE63	CTATGGGCGGTGAAAACCGA	GCCATGAATGAGATAAACCC
SINE69	ACGCCTTTATGTGCTAGTCA	GTTGATGTGCTGCATTATGCT
SINE102	GATCAAACAGGGTCATT	GATCTGCTGATGTGCCTC
SINE103	GATCTCTATGTACTCATGT	GATCCAACTGGCTGTTGTCT
SINE210	ATTTCCGTAGGTCTCACTAA	CGGAGATAACACCATTTCTC
SINE215	CCATCCATAAATTATTAAGG	TGGTAAGAGTTCTAACCTCT
SINE501	CCAACCATGACAGGAGAGGC	GCCCAGAACATACATCCTCG
SINE503	ACTGTACACTGCATACCTTG	ATGGCATAGATCGATGAAGT
SINE504	TTCGCAGAGCGGGAGTGGCT	CGAGCACCCAAAGCAACGAC
SINE505	TCATCTGCACCCTGCACACT	TGGCGTGTAAGGCCTTCCAG
SINE506	CAGAAACAAGTCGAGTTGTG	GCTAACTGAGCGTGAAATGC
SINE511	ATCGTCATCAAGAGCTAGCG	GAATCAAGACACAGCACGAG
SINEACH3	AGCTCCTCTACTGCATAGTC	GATTCTGAACATCTGATCGG
SINER507	GACTCCAGCCAACATGGAGA	AACCGTCCTCTACCTGATTC

Supplementary Table 1.7 Raw line means (average of 3 individual plant replicates) for 13 morphological and developmental traits for 281 *ORSC* accessions grown out from 2006-07.

NSFTV	DTHD	FLFLG	FL FLWD	PBRNB	PN LG	PNNB	CU NO	PTHT _culm	Pericarp_c olor	HUL CL	AW NPL U	CULM _ANGL E	STOLON _Maturity
401	89	_	_	5	16	5	109	93.87	1	8	9	6.33	0
402	93.5	48	0.5	6	23	7	22	115.4	1	8	8	7	1
407	_	36	1	8	19	7	76	141.13	1	2	0	3	0.33
410	85	18	0.5	5	15.5	10	63	62.7	1	8	9	5	0
413	58	17	0.8	5	1.3	9	_	77	1	8	9	9	0
415	66	37	1	5	16	11	_	90.67	1	8	9	_	_
416	98	30	1	6	19	19	6	107.85	0	2	_	5	0.5
420	97	13.7	0.5	4.7	11.3	13	95.5	114.55	1	8	9	4	0.5
427	98	30	1	5	27	7	52	103.2	1	8	9	7	0
428	76	32	1	10	19	3	121	139.7	1	8	9	3	0
431	88	22	1	4	16	8	66	119.75	1	8	9	4	0
433	_	25	1	7	18	8	47	120.3	0	2.5	7	7	0
435	94	22	1	5	16	5	55	119.9	0	8	9	6	0.5
438	95	19	1	5	13	8	66	106.2	_	8	8.33	3	0
442	61	23	1	5	12	2	48	61.05	1	8	9	4	0
443	61	_	_	_	_	_	37	94.9	1	8	9	3	0
444	82.5	_	_	_	_	_	_	51	1	8	9	9	0
445	71	_	_	_	_	_	24	110.1	1	_	_	7	0
446	56	26	1	6	19	10	22	88.1	1	8	7	3	0
449	74	18	1	5	15	22	99	70.73	1	8	6	7	0
450	_	14	1	5	16	15	62	83.1	1	8	9	7	0
451	86	17	1	5	8	2	_	64.67	1	2	9	_	_
453	76	26	1	7	20	8	85	147.5	_	8	6.33	1	0

NSFTV	DTHD	FLFLG	FL FLWD	PBRNB	PN LG	PNNB	CU NO	PTHT _culm	Pericarp_c olor	HUL CL	AW NPL U	CULM _ANGL E	STOLON _Maturity
454	64	_	_	5	16	3	37	99.85	1	8	9	5	0
457	74	_	_	7	23.5	5	39	139.8	1	8	9	3	0
461	_	32	1	7	22	11	33	130.7	1	6	3.67	2	0
465	_	22	1	6	21	13	35	138.23	1	8	9	2.33	0
467	183	31	1	5	20	8	61	117.43	1	8	9	4	0.67
472	_	35	1	7	25	6	51	103.57	1	8	9	2.33	0
477	98	36	1	7	25	17	21	113.9	1	2	7	3.67	0.33
481	83	28	1	8	18	26	74	88.15	1	8	9	4	0
482	87	25	1	5	25	1	111	78.7	1	8	7.67	3	0
483	_	28	1	5	14	4	56	151.3	1	8	9	4.33	0.67
484	90	19	1	5	21	4	150	114.8	1	8	9	7	0.33
487	_	30	1	9	17	5	64	121.37	1	8	9	7.67	0
488	95	52	1	6	19	7	27	137.5	1	8	9	1	0
490	_	12	1	5	14	8	80	107.85	1	8	9	7	0
493	_	_	_	_	_	_	_	_	1	8	9	_	_
494	108	_	_	6	11	5	111	82.55	1	8	9	9	1
495	_	25	0.5	9	21	10	55	117.1	1	8	9	_	_
496	_	27	0.8	7	20	2	22	68.1	1	8	9	_	_
498	50	31	1	6	22	4	100	130.63	1	8	9	7	0
499	48	23	1	6	21	2	80	131.77	1	8	9	5	0.67
501	97	34	1	6	21	10	58	97.37	1	8	8.33	3.33	0.67
503	111	_	_	6	20	3	57	143.6	1	8	_	5.67	0
506	98	17	0.9	8	19	1	37	143.6	1	8	7	1	1
508	91	33	1	8	22	7	90	122.67	_	5	1	2	0
509	_	44	1	10	27	10	30	145.7	1	8	7	1	0

NSFTV	DTHD	FLFLG	FL FLWD	PBRNB	PN LG	PNNB	CU NO	PTHT _culm	Pericarp_c olor	HUL CL	AW NPL U	CULM _ANGL E	STOLON _Maturity
523	_	25	0.6	5	20	8	35	_	1	8	9	_	_
549	_	34	1	5	17	3	108	139.6	1	8	7	8	0
551	_	29	1	8	20	17	51	143.3	1	8	8.33	_	_
553	_	26	1	9	22	22	40	144.65	0	8	7.67	9	1
555	89	31	1	6	17	18	59	95.85	1	8	7	3	0
568	84	32	0.6	6	18	10	71.5	156.2	1	8	9	9	1
592	_	26	1	11	23	13	52	121	1	8	5	6	0
600	_	23	0.6	6	17	3	41	110.8	1	3	1	1	0
602	_	_	_	_	_	_	49	_	1	8	9	_	_
605	_	_	_	6	16	2	19	163.33	0	2	5	_	_
665	110.5	_	_	_	_	_	57	100.95	1	2	0.5	3	0
666	75	_	_	_	_	_	53	80.87	1	8	9	4.33	0
669	124	_	_	_	_	_	52	113.63	0	8	9	5	0
673	110	_	_	_	_	_	53	82.83	1	2	1	3.67	0
676	126.5	_	_	_	_	_	31.5	126.4	0	2	2.5	1	0.5
682	95	_	_	_	_	_	22	147.5	1	8	9	5	1
683	93.5	_	_	_	_	_	92.5	105.75	1	8	9	5	0
685	96	_	_	_	_	_	52	95.63	0	2	8	3.67	0
686	54	_	_	_	_	_	55	144.17	1	_	_	5	0
687	84	_	_	_	_	_	71	82.67	1	8	5	2.33	0
691	53	_	_	_	_	_	60	90.57	1	8	9	4.33	0
701	80	_	_	_	_	_	66.5	94.2	0	2	7	4	0
704	74	_	_	_	_	_	31	97.97	0	2	9	3.67	0
707	61	_	_	_	_	_	28	132	1	8	9	3.67	0.33
708	96	_	_	_	_	_	25	103.77	1	5	5	1	0.67

NSFTV	DTHD	FLFLG	FL FLWD	PBRNB	PN LG	PNNB	CU NO	PTHT _culm	Pericarp_color	HUL CL	AW NPL U	CULM _ANGLE	STOLON _Maturity
711	37	_	_	_	_	_	18	116.75	1	8	9	1	1
715	58.5	_	_	_	_	_	71	107.15	1	8	9	5	0
716	115	_	_	_	_	_	97	81.9	1	8	9	5	0
717	88	_	_	_	_	_	75	93.47	1	8	9	5	0
719	42	_	_	_	_	_	35	97.9	1	8	9	7	0
720	78	_	_	_	_	_	114	92.5	1	8	9	5.67	0
721	97	_	_	_	_	_	110	97.9	1	8	7	3	0
722	53	_	_	_	_	_	82	83.57	_	8	9	5	0.67
723	75	_	_	_	_	_	81	84.47	1	8	9	5	0
736	94	_	_	_	_	_	54	91.05	0	_	_	8	0
738	89	_	_	_	_	_	95	78.27	1	_	9	3.67	0
743	43	_	_	_	_	_	90	80.03	1	8	9	5.67	0
746	88	_	_	_	_	_	37	104.4	0	_	_	4.33	0
751	99	_	_	_	_	_	52	105.43	1	2	3	1.67	0
757	80	_	_	_	_	_	99	91.45	_	_	_	6	0
759	79	_	_	_	_	_	50	122.03	0	8	9	5	0.33
760	96	_	_	_	_	_	78	72.57	1	8	9	4.33	0.33
762	97	_	_	_	_	_	66	76.53	1	8	9	3.67	0

REFERENCES

Aggarwal R. K., Brar D. S., Nandi S., Huang N., Khush G. S., 1999 Phylogenetic relationships among *Oryza* species revealed by AFLP markers. Theor. Appl. Genet. **98**: 1320–1328.

Akimoto M., Shimamoto Y., Morishima H., 1999 The extinction of genetic resources of Asian wild rice, *Oryza* rufipogon Griff.: a case study in Thailand. Genet. Resour. Crop **46**: 419–425.

Asano K., Yamasaki M., Takuno S., Miura K., Katagiri S., Ito T., Doi K., Wu J., Ebana K., Matsumoto T., Innan H., Kitano H., Ashikari M., Matsuoka M., 2011 Artificial selection for a green revolution gene during japonica rice domestication. Proc. **108**: 11034–9.

Banaticla-Hilario M. C. N., Berg R. G. van den, Hamilton N. R. S., McNally K. L., 2013 Local differentiation amidst extensive allele sharing in *Oryza* nivara and *O. rufipogon*. Ecol. Evol. **3**: 3047–62.

Barbier P., 1989 Genetic variation and ecotypic differentiation in the wild rice species *Oryza* rufipogon. I. Population differentiation in life-history traits and isozymic loci. Japanese J. Genet. **64**: 259–271.

Barbier P., Morishima H., Ishihama A., 1991 Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. Theor. Appl. Genet. **81**: 693–702.

Bautista N. S., Solis R., Kamijima O., Ishii T., 2001 RAPD, RFLP and SSLP analyses of phylogenetic relationships between cultivated and wild species of rice. Genes Genet. Syst. **76**: 71–9.

Bellwood P., 2011 The Checkered Prehistory of Rice Movement Southwards as a Domesticated Cereal—from the Yangzi to the Equator. Rice **4**: 93–103.

Cai H., Morishima H., 2002 QTL clusters reflect character associations in wild and cultivated rice. Theor. Appl. Genet. **104**: 1217–1228.

Cai H., Wang X., Morishima H., 2004 Comparison of population genetic structures of common wild rice (*Oryza* rufipogon Griff.), as revealed by analyses of quantitative traits, allozymes, and RFLPs. Heredity (Edinb). **92**: 409–17.

Cai H.-W., Akimoto M., Morishima H., 2008 Genetic diversity in wild relatives of rice and domestication events. In: *Rice Biology in the Genomics Era*,, pp. 261–275.

Caicedo A. L., Williamson S. H., Hernandez R. D., Boyko A., Fledel-Alon A., York T. L., Polato N. R., Olsen K. M., Nielsen R., McCouch S. R., Bustamante C. D., Purugganan M. D., 2007 Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS Genet. **3**: e163.

Cao Q., Lu B.-R., Xia H., Rong J., Sala F., Spada A., Grassi F., 2006 Genetic diversity and origin of weedy rice (*Oryza* sativa f. spontanea) populations found in North-eastern China revealed by simple

sequence repeat (SSR) markers. Ann. Bot. **98**: 1241–52.

Castillo C. C., Tanaka K., Sato Y.-I., Ishikawa R., Bellina B., Higham C., Chang N., Mohanty R., Kajale M., Fuller D. Q., 2015 Archaeogenetic study of prehistoric rice remains from Thailand and India: evidence of early japonica in South and Southeast Asia. Archaeol. Anthropol. Sci.: 1–21.

Chang T. T., 1976 The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica **25**: 425–441.

Chen W., Nakamura I., Sato Y., Nakai H., 1993 Distribution of deletion type in cpDNA of cultivated and wild rice. Japanese J. Genet. **68**: 597–603.

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. Mol. Biol. Evol. **20**: 67–75.

Civáň P., Craig H., Cox C. J., Brown T. A., 2015 Three geographically separate domestications of Asian rice. Nat. Plants **1**: 15164.

Coburn J., Temnykh S., Paul E., McCOUCH S., 2002 Design and Application of Microsatellite Marker Panels for Semiautomated Genotyping of Rice (L.). Crop Sci.

Crawford G., 2011 Early rice exploitation in the lower Yangzi valley: What are we missing? The Holocene **22**: 613–621.

Dally A. M., Second G., 1990 Chloroplast DNA diversity in wild and cultivated species of rice (Genus *Oryza*, section *Oryza*). Cladistic-mutation and genetic-distance analysis. Theor. Appl. Genet. **80**: 209–222.

Dellaporta S., Wood J., Hicks J., 1983 A plant DNA minipreparation: version II. Plant Mol. Biol. Report. **1**: 19–21.

Diarra A., Jr R. S., Talbert R., 1985 Growth and morphological characteristics of red rice (*Oryza* sativa) biotypes. Weed Sci. **33**: 310–314.

Duan S., Lu B., Li Z., Tong J., Kong J., Yao W., Li S., Zhu Y., 2007 Phylogenetic analysis of AA-genome *Oryza* species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. Biochem. Genet. **45**: 113–29.

Earl D. A., VonHoldt B. M., 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. **4**: 359–361.

Fuller D. Q., Sato Y.-I., Castillo C., Qin L., Weisskopf A. R., Kingwell-Banham E. J., Song J., Ahn S.-M., Etten J., 2010 Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol. Anthropol. Sci. **2**: 115–131.

Fuller D. Q., 2012 Pathways to Asian Civilizations: Tracing the Origins and Spread of Rice and Rice Cultures. Rice **4**: 78–92.

Gao L.-Z., Zhang C.-H., Li D.-Y., Pan D.-J., Jia J.-Z., Dong Y.-S., 2006 Genetic diversity within *Oryza* rufipogon germplasms preserved in Chinese field gene banks of wild rice as revealed by microsatellite markers. Biodivers. Conserv. **15**: 4059–4077.

Gao L.-Z., Innan H., 2008 Nonindependent domestication of the two rice subspecies, *Oryza* sativa ssp. indica and ssp. japonica, demonstrated by multilocus microsatellites. Genetics **179**: 965–76.

Garris A. J., Tai T. H., Coburn J., Kresovich S., McCouch S., 2005 Genetic structure and diversity in *Oryza* sativa L. Genetics **169**: 1631–8.

Ge S., Oliveira G., Schaal B., Gao L., Hong D., 1999 RAPD variation within and between natural populations of the wild rice *Oryza* rufipogon from china and brazil. Heredity (Edinb). **82 (Pt 6)**: 638–44.

Ge S., Li A., Lu B.-R., Zhang S.-Z., Hong D.-Y., 2002 A phylogeny of the rice tribe Oryzeae (Poaceae) based on matK sequence data. Am. J. Bot. **89**: 1967–72.

Ge S., Guo Y., Zhu Q., 2005 Molecular phylogeny and divergence of the rice tribe Oryzeae, with special reference to the origin of the genus *Oryza*. In: *Rice Is Life: Scientific perspectives for the 21st century*,, pp. 40–44.

Ge S., Sang T., 2011 Inappropriate model rejects independent domestications of indica and japonica rice. Proc. Natl. Acad. Sci. U. S. A. **108**: E755; author reply E756.

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 1. Weed Technol. **17**: 627–645.

Gordon B. C., 2010 The rise of Chinese civilization based on paddy rice agriculture. Web J. Bryan C. Gordon: 1–48.

Grillo M. a, Li C., Fowlkes A. M., Briggeman T. M., Zhou A., Schemske D. W., Sang T., 2009 Genetic architecture for the adaptive origin of annual wild rice, *Oryza* nivara. Evolution **63**: 870–83.

Guo Y.-L., Ge S., 2005 Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. Am. J. Bot. **92**: 1548–58.

Guo J., Liu R., Huang L., Zheng X.-M., Liu P.-L., Du Y.-S., Cai Z., Zhou L., Wei X.-H., Zhang F.-M., Ge S., 2016 Widespread and Adaptive Alterations in Genome-Wide Gene Expression Associated with Ecological Divergence of Two *Oryza* Species. Mol. Biol. Evol. **33**: 62–78.

Hattori Y., Nagai K., Mori H., Kitano H., Matsuoka M., Ashikari M., 2008 Mapping of three QTLs that

regulate internode elongation in deepwater rice. Breed. Sci. **58**: 39–46.

Hattori Y., Nagai K., Furukawa S., Song X.-J., Kawano R., Sakakibara H., Wu J., Matsumoto T., Yoshimura A., Kitano H., Matsuoka M., Mori H., Ashikari M., 2009 The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. Nature **460**: 1026–30.

He Z., Zhai W., Wen H., Tang T., Wang Y., Lu X., Greenberg A. J., Hudson R. R., Wu C.-I., Shi S., 2011 Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet. **7**: e1002100.

Henry R. J., Rice N., Waters D. L. E., Kasem S., Ishikawa R., Hao Y., Dillon S., Crayn D., Wing R., Vaughan D., 2009 Australian *Oryza*: Utility and Conservation. Rice **3**: 235–241.

Hill R. D., 2010 The cultivation of perennial rice, an early phase in Southeast Asian agriculture? J. Hist. Geogr. **36**: 215–223.

Hoffmann-Benning S., Kende H., 1992 On the role of abscisic acid and gibberellin in the regulation of growth in rice. Plant Physiol. **99**: 1156–1161.

Huang P., Molina J., Flowers J. M., Rubinstein S., Jackson S. A., Purugganan M. D., Schaal B. A., 2012a Phylogeography of Asian wild rice, *Oryza* rufipogon: a genome-wide view. Mol. Ecol. **21**: 4593–604.

Huang C.-L., Hung C.-Y., Chiang Y.-C., Hwang C.-C., Hsu T.-W., Huang C.-C., Hung K.-H., Tsai K.-C., Wang K.-H., Osada N., Schaal B. A., Chiang T.-Y., 2012b Footprints of natural and artificial selection for photoperiod pathway genes in *Oryza*. Plant J. **70**: 769–82.

Huang X., Kurata N., Wei X., Wang Z.-X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L., Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A., Toyoda A., Lu T., Feng Q., Qian Q., Li J., Han B., 2012c A map of rice genome variation reveals the origin of cultivated rice. Nature **490**: 497–501.

Hurwitz B. L., Kudrna D., Yu Y., Sebastian A., Zuccolo A., Jackson S. A., Ware D., Wing R. A., Stein L., 2010 Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus *Oryza*. Plant J. **63**: 990–1003.

Ishii T., Xu Y., McCouch S. R., 2001 Nuclear- and chloroplast-microsatellite variation in A-genome species of rice. Genome **44**: 658–66.

Izawa T., 2008 The Process of Rice Domestication: A New Model Based on Recent Data. Rice **1**: 127–134.

Izawa T., Konishi S., Shomura A., Yano M., 2009 DNA changes tell us about rice domestication. Curr. Opin. Plant Biol. **12**: 185–92.

Jakobsson M., Rosenberg N. a, 2007 CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics **23**: 1801–6.

Jiang Z., Xia H., Basso B., Lu B.-R., 2012 Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. Theor. Appl. Genet. **124**: 309–22.

JMP PRO 10,

Joshi S. P., Gupta V. S., Aggarwal R. K., Ranjekar P. K., Brar D. S., 2000 Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. TAG Theor. Appl. Genet. **100**: 1311–1320.

Kanazawa A., Akimoto M., 2000 Inter-and intra-specific distribution of Stowaway transposable elements in AA-genome species of wild rice. TAG Theor. Appl. Genet. **101**: 327–335.

Kanno A., Watanabe N., Nakamura I., Hirai A., 1993 Variations in chloroplast DNA from rice (*Oryza* sativa): differences between deletions mediated by short direct-repeat sequences within a single species. Theor. Appl. Genet. **86**: 579–584.

Kawakami S., Ebana K., Nishikawa T., Sato Y., Vaughan D. A., Kadowaki K., 2007 Genetic variation in the chloroplast genome suggests multiple domestication of cultivated Asian rice (*Oryza* sativa L.). Genome **50**: 180–7.

Kende H., Knaap E. van der, Cho H.-T. H., 1998 Deepwater rice: a model plant to study stem elongation. Plant Physiol. **118**: 1105–1110.

Khush G. S., 1997 Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. **35**: 25–34.

Konishi S., Ebana K., Izawa T., 2008 Inference of the japonica rice domestication process from the distribution of six functional nucleotide polymorphisms of domestication-related genes in various landraces and modern cultivars. Plant Cell Physiol. **49**: 1283–93.

Kovach M. M. J., Sweeney M. T. M., McCouch S. S. R., 2007 New insights into the history of rice domestication. Trends Genet. **23**: 578–87.

Kumagai M., Wang L., Ueda S., 2010 Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. Gene **462**: 44–51.

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

Lawton-Rauh A., Burgos N., 2010 Cultivated and weedy rice interactions and the domestication process. Mol. Ecol. **19**: 3243–5.

Li C., Zhou A., Sang T., 2006a Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza* nivara. New Phytol. **170**: 185–93.

Li C., Pan D., Mao X., Tu C., Zhou H., Fan Z., Li X., 2006b The genetic diversity of Gaozhou wild rice analyzed by SSR. Chinese Sci. Bull. **51**: 562–572.

Londo J. P. J., Chiang Y.-C. Y., 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. Proc. Natl. Acad. Sci. U. S. A. **103**: 9578–9583.

Londo J. P., Schaal B. a, 2007 Origins and population genetics of weedy red rice in the USA. Mol. Ecol. **16**: 4523–35.

Lu B., Naredo M. E. B., Juliano A. B. A., Jackson M. T., Jacobs S. W. L., Everett J., Lu B., Naredo M. E. B., Juliano A. B. A., Jackson M. T., 2000 Preliminary studies on taxonomy and biosystematics of the AA genome *Oryza* species (Poaceae). In: Jacobs S, Everett J (Eds.), *Grasses: systematics and evolution*, CSIRO Publishing, pp. 51–58.

Lu B.-R., Zheng K. L., Qian H. R., Zhuang J. Y., 2002 Genetic differentiation of wild relatives of rice as assessed by RFLP analysis. Theor. Appl. Genet. **106**: 101–6.

Lu J., Zhang X., Wang H., Yuan X., Xu Q., Wang Y., Yu H., Tang S., Wei X., 2008 SSR Analysis on Diversity of AA Genome *Oryza* Species in the Southeast and South Asia. Rice Sci. **15**: 289–294.

Ma J., Bennetzen J. L., 2004 Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. U. S. A. **101**: 12404–10.

McCouch S. R., Kovach M. J., Sweeney M., Jiang H., Semon M., 2012 The Dynamics of Rice Domestication: A Balance between Gene Flow and Genetic Isolation. In: Gepts P (Ed.), *Biodiversity in Agriculture: Domestication, Evolution, and Sustainability*, Cambridge University Press, pp. 311–329.

Mohapatra P. K., Panda B. B., Kariali E., 2011 Plasticity of Tiller Dynamics in Wild Rice *Oryza* rufipogon Griff.: A Strategy for Resilience in Suboptimal Environments. Int. J. Agron. **2011**: 1–9.

Molina J., Sikora M., Garud N., Flowers J. M., Rubinstein S., Reynolds A., Huang P., Jackson S., Schaal B. a, Bustamante C. D., Boyko A. R., Purugganan M. D., Molinaa J., Sikorab M., Garudb N., 2011 Molecular evidence for a single evolutionary origin of domesticated rice. Proc. Natl. Acad. Sci. U. S. A. **108**: 8351–8356.

Morishima H., Oka H.-I., Chang W.-T., 1961 Directions of Differentiation in Populations of Wild Rice , *Oryza* perennis and O . sativa f . spontanea. Evolution (N. Y). **15**: 326–339.

Morishima H., 1969 Phenetic similarity and phylogenetic relationships among strains of *Oryza* perennis, estimated by methods of numerical taxonomy. Evolution (N. Y). **23**: 429.

Morishima H., Sano Y., Oka H., 1984 Differentiation of Perennial and Annual Types Due to Habitat Conditions in the Wild Rice *Oryza* perennis. Plant Syst. Evol. **144**: 119–135.

Morishima H., 1991 Association between Pox-1 variation and seed productivity potential in wild rice. In: *Proceedings of the Second International Rice Genetics Symposium*, IRRI, International Rice Research Institute, Los Banos, pp. 1–9.

Morishima H., 2001 Evolution and domestication of rice. Rice Genet. IV. Proc. the. Fourth Int. Rice Genet. Symp.: 63–77.

Motohashi R., Mochizuki K., Ohtsubo H., Ohtsubo E., 1997 Structures and distribution of p-SINE1 members in rice genomes. TAG Theor. Appl. Genet. **95**: 359–368.

Ng N. Q., Chang T. T., Williams J. T., Hawkes J. G., 1981a Morphological studies of Asian rice and its related wild species and the recognition of a new Australian taxon. Biol. J. ….

Ng N. Q., Hawkes J. G., Williams J. T., Chang T. T., 1981b The recognition of a new species of rice (*Oryza*) from Australia. Bot. J. Linn. Soc. **82**: 327–330.

Nishikawa T., Vaughan D. a., Kadowaki K. I., 2005 Phylogenetic analysis of *Oryza* species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. Theor. Appl. Genet. **110**: 696–705.

Noldin J., Chandler J., McCauley G., 1999 Red rice (*Oryza* sativa) biology. I. Characterization of red rice ecotypes. Weed Technol. **13**: 12–18.

Nonomura K.-I. K., Morishima H., Miyabayashi T., Yamaki S., Eiguchi M., Kubo T., Kurata N., 2010 The wild *Oryza* collection in National BioResource Project (NBRP) of Japan: History, biodiversity and utility. Breed. Sci. **60**: 502–508.

Ohtsubo H., Cheng C., Ohsawa I., Tsuchimoto S., Ohtsubo E., 2004 Rice retroposon p-SINE1 and origin of cultivated rice. Breed. Sci. **54**: 1–11.

Ohtsubo H., Tsuchimoto S., Xu J., Cheng C., Koudo M. Y., Kurata N., Ohtsubo E., 2008 Rice Biology in the Genomics Era, Rice Retroposon, p-SINE, and Its Use for Classification and Identification of *Oryza* Species. In: Hirano H-Y, Sano Y, Hirai A, Sasaki T (Eds.), Biotechnology in Agriculture and Forestry. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 277–291.

Oka H., 1964 Pattern of interspecific relationships and evolutionary dynamics in *Oryza*. Rice Genet. Cytogenet. Proc. Symp. Rice Genet. Cytogenet.: 71–90.

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

Oka H., 1974 Experimental studies on the origin of cultivated rice. Genetics **78**: 475–486.

Oka H. I., 1977 The ancestors of cultivated rice and their evolution. In: *Meeting on African Rice Species*, IRAT-ORSTOM, Paris, pp. 57–64.

Oka H., 1988 *Origin of cultivated rice.*

Park K. C., Kim N. H., Cho Y. S., Kang K. H., Lee J. K., Kim N.-S., 2003 Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. Theor. Appl. Genet. **107**: 203–9.

Ren F., Lu B. B.-R., Li S., Huang J., Zhu Y., 2003 A comparative study of genetic relationships among the AA-genome *Oryza* species using RAPD and SSR markers. Theor. Appl. Genet. **108**: 113–20.

Rosenberg N. A., 2003 Distruct: a Program for the Graphical Display of Population Structure. Mol. Ecol. Notes **4**: 137–138.

Saitou N., Nei M., 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**: 406–25.

Sang T., Ge S., 2007a The Puzzle of Rice Domestication. J. Integr. Plant Biol. **49**: 760–768.

Sang T., Ge S., 2007b Genetics and phylogenetics of rice domestication. Curr. Opin. Genet. Dev. **17**: 533–8.

Sano Y., Morishima H., Oka H., 1980 Intermediate perennial-annual populations of *Oryza* perennis found in Thailand and their evolutionary significance. Bot. Mag. Tokyo **93**: 291–305.

Sano Y., Morishima H., 1982 Variation in Resource Allocation and Adaptive Strategy of a Wild Rice , *Oryza* perennis Moench. Bot. Gaz. **143**: 518–523.

Second G., Morishima H., 1980 *Geographical and ecological pattern of variation of Oryza perennis at 7 isozyme loci*. Mishima, Japan.

Second G., 1982 Origin of the genic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Japanese J. Genet. **57**: 25–57.

Second G., 1985 Evolutionary relationships in the Sativa group of *Oryza* based on isozyme data. Genet. Sel. Evol. **17**: 89–114.

Sharma S., 2003 Species of genus *Oryza* and their interrelationships. In: *Monograph on Genus Oryza*, Science Publishers, Enfield, New Hampshire, pp. 73–111.

Shimizu H., Maruoka M., Ichikawa N., Baruah A. R., Uwatoko N., Sano Y., Onishi K., 2010 Genetic control of phenotypic plasticity in Asian cultivated and wild rice in response to nutrient and density changes. Genome **53**: 211–23.

Shishido R., Kikuchi M., Nomura K., Ikehashi H., 2006 Evaluation of Genetic Diversity of Wild Rice

(*Oryza* rufipogon Griff.) in Myanmar using Simple Sequence Repeats (SSRs). Genet. Resour. Crop Evol. **53**: 179–186.

Shomura A., Izawa T., Ebana K., Ebitani T., Kanegae H., Konishi S., Yano M., 2008 Deletion in a gene associated with grain size increased yields during rice domestication. Nat. Genet. **40**: 1023–8.

Sun C., Yoshimura A., Li Z., Zhou H., Iwata N., 1996 Genetic differentiation of chloroplast genome in *O. rufipogon* Griff. and *O. sativa* L. In: Wang X, Sun C (Eds.), *Origin and Differentiation of Chinese Cultivated Rice*,, pp. 140–145.

Sun C. Q., Xiang-Kun W., Yoshimura A., Iwata N., 1997 RFLP analysis of nuclear DNA in common wild rice (*O. rufipogon* Griff.) and cultivated rice (*O. sativa* L.). Sci. Agric Sin. **4**: 37–44.

Sun C. Q., Wang X., Li Z., Yoshimura A., Iwata N., 2001 Comparison of the genetic diversity of common wild rice (*Oryza* rufipogon Griff.) and cultivated rice (*O. sativa* L.) using RFLP markers. TAG Theor. Appl. Genet. **102**: 157–162.

Sun C. Q., Wang X. K., Yoshimura A., Doi K., 2002 Genetic differentiation for nuclear, mitochondrial and chloroplast genomes in common wild rice (*Oryza* rufipogon Griff.) and cultivated rice (*Oryza* sativa L.). Theor. Appl. Genet. **104**: 1335–1345.

Sweeney M., McCouch S., 2007 The complex history of the domestication of rice. Ann. Bot. **100**: 951–7.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S., 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**: 2731–9.

Tang L.-H., Morishima H., 1997 Genetic characterization of weedy rices and the inference on their origins. **47**: 153–160.

Tang T., Lu J., Huang J., He J., McCouch S. R., Shen Y., Kai Z., Purugganan M. D., Shi S., Wu C.-I., 2006 Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication. PLoS Genet. **2**: e199.

Tateoka T., 1964a Taxonomic studies of the genus *Oryza*. In: *Rice genetics and cytogenetics: Proceedings of the Symposium on Rice Genetics and Cytogenetics*,, pp. 15–21.

Tateoka T., 1964b Notes on some grasses. XVI. Embryo structure of the genus *Oryza* in relation to the systematics. Am. J. Bot. **51**: 539.

Toriyama K., Heong K. L., Hardy B., 2005 Rice is life: scientific perspectives for the 21st century. In: *Proceedings of the World Rice Research Conference*, International Rice Research Institute, Tsukuba, Japan, p. .

Vaughan D. A., 1989 The genus *Oryza* L . Current status of taxonomy. IRRI Res. Pap. Ser.: 1–21.

Vaughan D. A., 1994 *The Wild Relatives of Rice: A Genetic Resources Handbook*. Int. Rice Res. Inst.

Vaughan L. K., Ottis B. B., 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 Sci. **49**: 468–476.

Vaughan D. A., Morishima H., Kadowaki K., 2003 Diversity in the *Oryza* genus. Curr. Opin. Plant Biol. **6**: 139–146.

Vaughan D. A., Kadowaki K., Kaga A., Tomooka N., 2005 On the Phylogeny and Biogeography of the Genus *Oryza*. Breed. Sci. **55**: 113–122.

Vaughan D. A., Lu B.-R., Tomooka N., 2008a The evolving story of rice evolution. Plant Sci. **174**: 394–408.

Vaughan D. A., Lu B.-R., Tomooka N., 2008b Was Asian Rice (*Oryza* sativa) Domesticated More Than Once? Rice **1**: 16–24.

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.). Mol. Genet. Genomics **272**: 504–11.

Wambugu P. W., Brozynska M., Furtado A., Waters D. L., Henry R. J., 2015 Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. Sci. Rep. **5**: 13957.

Wang Z. Y., Second G., Tanksley S. D., 1992 Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83**: 565–581.

Wang M. X., Zhang H. L., Zhang D. L., Pan D. J., Li D. Y., Fan Z. L., Qi Y. W., Sun J., Yang Q. W., Li C., Li Z. C., Cao Y. S., Qiu Z. E., Yu P., Wang X. K., Fan Z. L., Li D. Y., Pan D. J., Cao Y. S., Qiu Z. E., Yu P., Yang Q. W., Wang X. K., Li Z. C., 2008a Genetic structure of *Oryza* rufipogon Griff. in China. Chinese Sci. Bull. **53**: 527–35.

Wang M. X., Zhang H. L., Zhang D. L., Qi Y. W., Fan Z. L., Li D. Y., Pan D. J., Cao Y. S., Qiu Z. E., Yu P., Yang Q. W., Wang X. K., Li Z. C., 2008b Genetic structure of *Oryza* rufipogon Griff. in China. Heredity (Edinb). **101**: 527–35.

Wang Y., Bai X., Yan C., Gui Y., Wei X., Zhu Q.-H. H., Guo L., Fan L., 2012 Genomic dissection of small RNAs in wild rice (*Oryza* rufipogon): lessons for rice domestication. New Phytol. **39**: 914–925.

Waters D. D. L. E., Nock C. J. C., Ishikawa R., Rice N., Henry R. J., 2012 Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. Ecol. Evol. **2**: 211–7.

Wet J. M. J., 1981 Species concepts and systematics of domesticated cereals. Die Kult. **29**: 177–198.

Xiong Z. Y., Zhang S. J., Ford-Lloyd B. V., Jin X., Wu Y., Yan H. X., Liu P., Yang X., Lu B.-R., 2011 Latitudinal Distribution and Differentiation of Rice Germplasm: Its Implications in Breeding. Crop Sci. **51**: 1050.

Xu J.-H., Cheng C., Tsuchimoto S., Ohtsubo H., Ohtsubo E., 2007 Phylogenetic analysis of *Oryza* rufipogon strains and their relations to *Oryza* sativa strains by insertion polymorphism of rice SINEs. Genes Genet. Syst. **82**: 217–29.

Xu X., Walters C., Antolin M. F., Alexander M. L., Lutz S., Ge S., Wen J., 2010 Phylogeny and biogeography of the eastern Asian-North American disjunct wild-rice genus (Zizania L., Poaceae). Mol. Phylogenet. Evol. **55**: 1008–17.

Xu X., Liu X., Ge S., Jensen J. D. J. J. D. J. J. D. J., Hu F., Li X., Dong Y., Gutenkunst R. N., Fang L., Huang L., Li J., He W., Zhang G., Zheng X., Zhang F., Li Y., Yu C., Kristiansen K., Zhang X., Wang J. J., Wright M., Mccouch S., Nielsen R., Wang J. J., Wang W., 2012 Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. **30**: 105–111.

Yamanaka S., Nakamura I., Nakai H., Sato Y., 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*. Genet. Resour. Crop Evol. **50**: 529–538.

Yang C., Kawahara Y., Mizuno H., Wu J., Matsumoto T., Itoh T., 2012 Independent domestication of Asian rice followed by gene flow from japonica to indica. Mol. Biol. Evol. **29**: 1471–9.

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. Biochem. Genet. **43**: 261–270.

Zhang L. L.-B., Zhu Q., Wu Z.-Q. Z., Ross-Ibarra J., Gaut B. B. S., Ge S., Sang T., 2009 Selection on Grain Shattering Genes and Rates of Rice Domestication. New Phytol. **184**: 708–20.

Zhang S., Li J., Lee D., Xu H., Zhang L., Dongchen W., Xiong H., Zhu Q., Zhang X., Lu B., Chen L., 2014 Genetic differentiation of Asian weedy rice revealed with InDel markers. Crop Sci. **54**: 2499.

Zhao X., Yang L., Zheng Y., Xu Z., Wu W., 2009 Subspecies-specific intron length polymorphism markers reveal clear genetic differentiation in common wild rice (*Oryza* rufipogon L.) in relation to the domestication of cultivated rice (*O. sativa* L.). J. Genet. Genomics **36**: 435–42.

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

Zhu Q., Ge S., 2005 Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. New Phytol. **167**: 249–265.

Zhu Q., Zheng X., Luo J., Gaut B. S., Ge S., 2007 Multilocus analysis of nucleotide variation of *Oryza*

sativa and its wild relatives: severe bottleneck during domestication of rice. Mol. Biol. Evol. **24**: 875–88.

Zhu T., Xu P.-Z. Z., Liu J.-P. P., Peng S., Mo X.-C. C., Gao L.-Z. Z., 2014 Phylogenetic relationships and genome divergence among the AA- genome species of the genus *Oryza* as revealed by 53 nuclear genes and 16 intergenic regions. Mol. Phylogenet. Evol. **70**: 348–361.

.

CHAPTER 2 - GENOME-WIDE ASSOCIATION MAPPING IN AN *O. RUFIPOGON* SPECIES COMPLEX DIVERSITY PANEL

Note: A manuscript pertaining to much of the content of this chapter and on which I will be co-first author is currently in preparation. This manuscript will focus on phenotypic variation and GWAS in the *ORSC* diversity panel of 95 accessions using sub-population identities from the Kim, Jung et al. paper (in review; Appendix 1). This manuscript will also include re-analyzed, Bayesian-transformed line means from the phenotypic datasets generated at IRRI, Cornell and Dale Bumpers National Rice Research Center, and genotypes from the High Density Rice Array (HDRA, 557,134 SNPs) (McCouch *et al.* 2016) and Genotyping-By-Sequencing (GBS, 480,577 SNPs) datasets.

Introduction

Cultivated Asian rice, *O. sativa* L., is the staple starch for over half of the people in the world (Toriyama *et al.* 2005). As the human population continues to grow, the FAO projects that overall global food production will have to increase by approximately 70% (Alexandratos and Bruinsma 2012), and the global rice production by approximately 30-50% by 2050 in order to match the demand by U.N. projected 2050 global population of 9.3 billion (2011). With growing competition for limited land and freshwater resources, constantly evolving disease and pest pressure, rising sea levels and increasingly unpredictable weather patterns associated with climate change, gains in agricultural productivity will require a major shift in the way food is produced. New crop varieties and more environmentally and economically sustainable crop

production systems are urgently needed. This is particularly true for small-scale farmers in the developing world where populations are growing the fastest and the demand for food is greatest. Future crop productivity must be achieved using land, water, nutrients, energy and labor more efficiently. The development of new high yielding, stress tolerant and nutrient-rich crop varieties, in combination with improved agronomic and post-harvest practices and supportive economic/ policy-based systems that extend credit and expand market opportunities for poor farmers, will be essential to meet our global food production requirements.

Historically, farmers have been faced with uncertainty at every level of crop production, and they have integrated diversity into their agricultural systems as a form of risk-management. Traditionally, diversity exists at many levels, in the heterogeneity of landrace varieties within a field, in the multiplicity of crops cultivated in a single season, and in the rotation of crops over time. All are key components of sustainable agricultural systems and help ensure long-term economic as well as biological viability. Modern crop varieties tend to be genetically uniform, but planting of a succession of genetically diverse varieties can help ensure the presence of genetic variation in farmers' fields. There is a growing emphasis on utilizing novel sources of genetic variation to broaden the germplasm base and expand the potential for genetic gain in the future. Wild crop relatives represent a valuable source of novel alleles that can be productively leveraged for crop improvement as plant breeders strive to meet future production and natural resource management needs.

As discussed in Chapter 1, molecular genetic research suggests that the *O. rufipogon* species complex (*ORSC*) is the gene pool from which cultivated Asian rice was domesticated. It is an

ambiguously characterized taxon consisting of a continuum of annual to perennial and vegetative to clonally reproductive populations found throughout South and Southeast Asia. Analyses of genetic diversity and population structure of *O. rufipogon* indicate that the species is genetically divided into at least four subpopulations that are broadly distributed across its geographical range. The subpopulations are particularly difficult to differentiate phenotypically because they share various morphological and developmental characteristics related to life habit and reproductive mode. Several accessions from the *ORSC* have been successfully used to introgress novel alleles conferring higher yield (McCouch *et al.* 2006), aluminum tolerance (Nguyen *et al.* 2003), drought resistance (Tian *et al.* 2006; Zhang *et al.* 2006), and disease resistance (Zhang *et al.* 2001; Zhou *et al.* 2011) into cultivars; however, the vast genetic potential of these and other related AA genome *Oryza* species for the improvement of cultivated rice is largely unexplored and untapped.

Studies of population structure, genetic and morphological diversity, and phylogeographic distribution explored in the last chapter and in earlier research have provided a foundation for understanding the *ORSC*. Recent gains in the efficiency and accuracy of high throughput phenotyping and genotyping have also made it possible to undertake GWAS using wild as well as cultivated rice germplasm. This is important because not only does the wild germplasm harbor a much larger pool of variation than the cultivar, but LD decays much more quickly in the outcrossing *ORSC* and is estimated to decay by half within ~20kb in *O. rufipogon* (Huang *et al.* 2012), compared with a range of ~40-500kb in its cultivated, inbreeding descendant *O. sativa* The more rapid LD decay in *O. rufipogon* offers higher resolution for GWAS, and greater

probability of pinpointing the genes underlying traits of interest.

The power to reliably detect associations using the heterogeneous, heterozygous wild germplasm is dependent on multiple factors:

1. Selection of a large and genetically diverse panel of *ORSC* germplasm representing the desired range of geographic, phenotypic, and allelic diversity as the foundation for QTL/gene discovery
2. Creation of minimally heterozygous, true-breeding lines derived from *ORSC* accessions through multiple rounds of self fertilization and single seed descent (SSD) to enable estimation of phenotypic means with minimal variance.
3. A high-resolution genotyping platform providing complete genome coverage and satisfactory representation of the wild alleles contained in the panel
4. Accurate, well-replicated, phenotypic data on the panel of purified lines
5. The computational and statistical ability to identify significant associations between phenotype and genotype, with particular attention to the abundance of rare alleles in wild germplasm, and the need to control for population structure.

Phenotypic screens of *ORSC* diversity panel

As the basis for GWAS, we evaluated the *O. rufipogon* diversity panel described in Chapter 1 for a wide range of phenotypes in order to explore questions about morphology and development-related differences in life habit, reproductive habit, and phylogeographic variation discussed in the previous chapter, as well as traits related to the potential agronomic utility of this wild germplasm. The panel and various phenotypic screens are summarized in Supplementary Table

2.1. These screens included evaluations of (1) development and reproduction-related morphological characters conducted over multiple generations in the Guterman greenhouse, (2) a core set of life cycle and reproductive habit-related traits during a single-season (2009) in the field in China, (3) seedling vigor using a replicated growth chamber assay, (4) aluminum tolerance based on relative root growth of 10 day old seedlings in a hydroponic system in the growth chamber, (5) 3D root system architecture (RSA) where seedlings were grown in a gellan gum growth media and evaluated at days 3, 6, and 9 for 13 traits, and (6) micronutrient concentrations in roots and shoots of 6 week old plants grown in hydroponics using an "ionomics" approach to simultaneously analyze 24 different ions. The complementary *O. sativa* diversity panel consisting of ~400 diverse varieties (Zhao et al., 2011; Eizenga et al., 2013) was also evaluated for a similar set of agronomic and morphological characters, as summarized in Supplementary Table 2.2. The coordinated development and use of controlled trait ontology and trait measurement regimes within and across these experiments was essential for integrating data and comparing GWAS results obtained from experiments using *O. rufipogon* and *O. sativa* diversity panels.

Overview of phenotypic screens: Greenhouse and field-based evaluation

The evaluation of developmental and reproductive traits conducted on the *ORSC* panel in multiyear, multi-location, and multi-generation plantings in greenhouse and screenhouse environments was necessary because we evaluated phenotypes at the same time that we purified the accessions. There were many challenges to overcome during this 4-year process due to the radical differences in growth habit, flowering time and seed set among accessions in the *O.*

rufipogon diversity panel. To the best of our ability, we sought to standardize growth conditions to ensure accurate representation and measurement of the genetic contribution to phenotypic characteristics across the entire panel, while at the same time we sought to facilitate the bagging of panicles prior to anthesis to ensure self-fertilization and seed collection. The traits evaluated were chosen for their agronomic value, indicators of yield and quality, importance in the domestication process, and relationship to life and reproductive habit. The complete list of traits, including the ontology terms used to describe them and the methods used to measure them, with the sole exception of the ionomics screen, can be found in Supplementary Table 2.2.

Field conditions are inherently less stable than those in the greenhouse, but at the same time grant the researcher a certain degree of freedom from the spatial limitations of pot-based plantings under a climate controlled environment. Although the soil and hydrologic conditions of the irrigated rice paddy are not wholly representative of the range of variation in the natural habitats of the *ORSC*, the opportunity to observe and measure traits over the course of a season encompassing natural seasonal and diurnal fluctuations in temperature, day length, and biotic stresses afforded invaluable insight into the reproductive and life habit of the *ORSC* and the potential utility of this material as parents for introgression breeding with *O. sativa*.

Collaboration with Chinese researchers facilitated the evaluation of our *O. rufipogon* diversity panel under irrigated field conditions in China. This was an important and rare opportunity. While many evaluations of *O. rufipogon* accessions were performed in the field throughout Asia during the 1970s and 1980s (Oka and Morishima 1967; Morishima *et al.* 1984; Barbier 1989), most rice growing countries, including the US, now prohibit researchers from deliberate field

planting of *ORSC* accessions to limit the risk of undesirable trait introgression into commercial cultivars and subsequent seedbank contamination. The growing interest in the use of direct seeded (rather than transplanted) rice has greatly increased the severity of problems associated with naturalized populations of weedy rice in major rice-growing regions across the globe.

Seedling vigor evaluation

Seedling vigor is a complex trait with important ecological and agronomic implications. The evaluation of seedling vigor integrates long and short-term seed viability, degree of seed dormancy, and early seedling root system and coleoptile growth rates. Seed dormancy may be simply defined as the inhibition of seed germination under favorable conditions, but in itself, it is determined by a complex set of interactions between the genetic control of germination, environmental perception and response, and the environmental stimuli needed to trigger germination (See reviews in Koornneef et al., 2002; Finkelstein et al., 2008). Seed dormancy in *O. sativa* has also been reported to be phenotypically correlated with weedy or domestication-related characteristics such as black hull, red pericarp, shattering, and awn presence (Cai and Morishima 2000; Gu *et al.* 2004, 2005a; b), as well as heading date (Lin *et al.* 1998), due to tight genetic linkage between loci controlling these traits. Our GWAS study using an *O. rufipogon* diversity panel to assess dormancy-related traits could help fine-map known dormancy QTL, or identify additional loci involved in dormancy that may be unique to some of the *O. rufipogon* accessions or subpopulations represented in our panel.

Once dormancy is broken, seedling root and shoot growth rate are the major developmental components of seedling vigor, and can affect the speed and success of seedling establishment

with implications for later impacts on plant growth and maturity. Rice seedling vigor in general, and root and shoot growth rate in particular, are also highly dependent on the interaction between intrinsic genetic and hormonal factors that respond to external environmental conditions. Several QTL for seedling vigor have been identified in *O. sativa* under both field and controlled laboratory conditions (Redoña and Mackill 1996; Cui *et al.* 2002; Zhang *et al.* 2005; Zhou *et al.* 2007). The time–course evaluation of seedling vigor in *O. rufipogon*, as measured by early seedling radicle and coleoptile growth rate under the controlled environment conditions documented in this study, will provide the first opportunity to explore the genetics of transgresssive variation for seedling vigor in the wild ancestral complex, and GWAS will allow us to identify new loci underlying these traits in the wild, and also perhaps allow us to further explore the identity and ancestry of corresponding loci in *O. sativa* domesticates.

Root system architecture (RSA)

Root system architecture (RSA) is another trait complex composed of morphological, spatial, and temporal root growth and development characteristics, dependent on the interplay between a web of genetic and environmental factors. For a review of the genes, hormones, and major environmental components involved in RSA, see Chapter 3. The rhizosphere — the abiotic and biotic factors that together comprise the rooting environment -- strongly affects RSA and overall plant growth and development. The importance of the genotype-by-environment interaction underlies the plastic nature of root growth in response to environment, and is a fascinating subject for further exploration. It can be studied as a developmental problem within the life cycle of an individual, or as an evolutionary process in terms of the generational adaptative potential of

a population.

Taking into consideration the different rhizosphere hydrologies across which the *ORSC* is found, and the inherent fluctuation or stability therein (Fig. 1.9), it is of interest for this study to explore the inherent variation in RSA across the different accessions of *O.rufipogon* in the diversity panel in order to identify possible correlations between RSA and other phenotypic complexes, including patterns of above-ground shoot growth, life and reproductive habits, as well as genetic correlations with subpopulation membership. Whether genetically admixed individuals or individuals adapted to the greatest fluctuations in water levels show greater phenotypic plasticity in root growth response and overall fitness will be an interesting subject for future study.

The RSA screen used in this study to measure components of 3D seedling RSA in both the *O. sativa* and *ORSC* diversity panels is conducted in a tightly controlled, highly artificial, semi-sterile growth environment, as described in Clark et al., (2011), and serves to tease out the inherent genetic components of RSA from the many important and diverse external environmental influences. Now that we have a base line of information about RSA in our *O. rufipogon* diversity panel, future work will focus on the way specific types of environmental perturbations and influences impact RSA in different accessions.

Aluminum tolerance

Aluminum is the third most abundant element in the top layer of the earth's crust (Wolt 1994). Under non-flooded, acid soil conditions (pH<5.0), aluminum is solubilized from a bound component of soil clays to its phytotoxic form, Al^{3+}, which causes stunting and inhibition of root growth. In flooded soils such as those under irrigated paddy culture, or in natural environments

such as streambeds, swamps, and marshes, the water acts as a buffer to raise soil pH, preventing Al^{3+} generation and negating aluminum phytotoxicity. Cultivated rice varieties have been found to be two to six-times more aluminum tolerant than maize, wheat, and sorghum varieties (Famoso *et al.* 2010), despite the fact that *O. sativa* was domesticated from aquatic or seasonally submerged populations of the *ORSC* and in general would be less exposed to Al^{3+} and less likely to have evolved or be under selection for high levels of aluminum tolerance.

The mechanisms of Al tolerance in rice are not fully understood, but appear to differ from the major mechanism of Al tolerance in other crop grasses: Al^{3+} root uptake exclusion through chelation of Al^{3+} by Al-induced organic acid exudation from root tips (Pineros *et al.* 2002; Piñeros *et al.* 2005). For a brief review of the genetic and hormonal networks involved in aluminum tolerance, see the "Aluminum toxicity" section in Chapter 3. Screening our diversity panel of *ORSC* accessions for Al tolerance using the hydroponic assay developed by Famoso et al. (2010) will allow us to determine the variation in aluminum tolerance within the *O .rufipogon* complex. By employing GWAS to compare QTL between the wild and cultivated species complexes and look for novel alleles for aluminum tolerance, examine possible correlations between aluminum tolerance in the *ORSC* with accession geography, soil hydrology, and population structure, and to test questions on the domestication of *O. sativa* from the *ORSC* and the putative origin of aluminum tolerance in cultivated rice.

Previous GWAS in rice

This study represents only the second attempt to conduct GWAS in a wild crop relative. The first study, also in the *ORSC*, by Huang et al. (2012) looked at only two traits: leaf sheath color and

tiller angle, using a total of 256,799 SNPs segregating across a panel of 446 *O. rufipogon* accessions, with a high component of southern Chinese accessions. The authors of the study found the two strongest associations correlated with known genomic locations of two major loci for pigmentation (*OsC1*) and tiller angle (*PROG1*) in *O. rufipogon.* Their use of an *O. rufipogon* diversity panel provided an average resolution that was three times greater than when GWAS was conducted using an equivalent number of *O. sativa* accessions, validating the ability to make high-resolution QTL associations in a wild crop relative.

Our study utilizes a much smaller GWAS panel consisting of 95 geographically, genetically, and morphologically diverse *O. rufipogon* accessions, but we evaluated a much larger array of phenotypes representing agronomically important, domestication-related, and unique wild trait characteristics.

This chapter describes the development of the panel, the evaluation of diverse morphological, developmental and agronomic phenotypes, and subsequent GWAS results. The outcome of this work will be discussed in the context of six different objectives as described below:

1. To develop an immortal, wild diversity panel consisting of 95 *ORSC* accessions to serve as a core resource for GWAS in rice
2. To survey the range of variation in the *ORSC* for key morphological, developmental and agronomic phenotypes, focusing on seedling root, and shoot growth and development, life and reproductive habit, aluminum tolerance and ion content

3. To evaluate trait correlations that signify multitrait complexes and explore the relationship between those trait complexes and larger developmental schemes, population structure, phylogeography, and ecological variation
4. To identify sources of allelic variation for basic biological and agronomic traits in the *ORSC* based on GWAS
5. To compare the genetic basis of complex trait variation in *O. rufipogon* and *O. sativa*
6. To test hypotheses about the domestication of *O. sativa* from populations of the *O. rufipogon* complex based on comparisons of genotypic and phenotypic variation within and between the two species groups.

In order to at least partially address all of these study objectives, this chapter will focus on individual developmental and morphological traits evaluated on plants grown in soil in the Guterman greenhouse, aluminum tolerance traits under growth chamber hydroponic conditions and 3D-RSA traits evaluated on seedlings grown in gellan gum media.

Materials and Methods

Germplasm selection and purification

The rationale and initial accession selection for the *ORSC* GWAS panel was detailed in Chapter 1. Several accessions were dropped from the panel due to lack of germination or survivability in the first greenhouse-based growout of the panel at Cornell University in 2008 or in the two subsequent growouts, leaving the final total of 95 (Supplementary Table 2.1). This panel will henceforth be referred to as the 'Rice Diversity Panel 1 (RDP1) wild' panel.

Annual successive growouts of the RDP1 wild panel were conducted for both selfed seed production through single seed descent (SSD) and coordinated phenotypic evaluation of morphological and developmental traits. Panel growout for pureline accession generation and phenotyping was performed in two locations: the Guterman Bioclimatic Laboratory and Greenhouse Complex at Cornell University (CU), Ithaca, NY (2006/7-2010), and the greenhouse complex at the Dale Bumpers National Rice Research Center (DBNRRC), Stuttgart, AR (2007-2009). Screenhouse growout and phenotyping of a larger panel of 251 *ORSC* accessions, including those in the RDP1 wild panel was also done under screenhouse conditions at the International Rice Research Institute (IRRI) in Los Banos, the Philippines. A 2007 planting at IRRI failed due to severe brown planthopper infestation; a successful growout and phenotypic assessment was completed in 2008.

Only CU growout methods and results will be remarked further upon in this chapter. .

All CU panel growouts at the Guterman Bioclimatic Laboratory and Greenhouse Complex were performed in the same greenhouse (G160), under constant temperature and supplemented light settings: 85°F day/ 75°F night, 14 hr light/10hr lights-off.

To ensure self-pollination and prevent seed from loss by shattering, panicles were bagged prior to stigma exsertion with ventilated waxed or glassine paper seed bags, secured with twist ties. Accessions were always grown out in three single-individual replicates. Selfed seed from each individual plant representing a replicate was collected separately, bulked by individual parent plant, and given a unique identifier. Selfed seed by SSD or the related plants, tissue or growouts of accessions are referred to by the generational indicator "Sn" as follows: S0 – source seed from

genebank, S1 – seed/plants/tissue from the first generation of SSD, etc. Young leaf tissue was also collected from each plant during each growout for Biobank deposition at -20C. The seed and leaf tissue from a single individual of each accession was chosen based on phenotypic uniformity and high seed availability to represent that accession for the next generational growout and any phenotypic or genotypic screens. In total, accessions from the RDP1 wild panel were purified by three generations of SSD at Cornell University to the S3 generation. Thirty accessions with low S3 seed stock in total or after S3 seed use for phenotypic screens were planted out in 2012 for seed bulking and tissue harvest only.

Panel growout, selfed-seed generation, and limited phenotypic evaluation at Cornell University was done with the assistance of Sandra Harrington, Fumio ("Gen") Onishi, Kazi Akther, Hyunjung Kim, and David Harris. The 2006-2007 S0 CU growout (see Chapter 1) was conducted by Jennifer Kimball and Lisa Polewczak. Panel growout, selfed-seed generation, and phenotypic evaluation at the DBNRRC was done in collaboration with the lab of Georgia Eizenga by Daniel Wood and Teresa Hancock.

Morphology and development phenotypic evaluation

There were four growouts representing three generations (S0-S2) of the panel at CU: S0 in 2006-2007, S1 in 2008-2009, 40 S2 accessions in 2009-2010, and the remaining ~60 S2 accessions in 2010-2011. For all growouts, 3-4 seed of each accession were sterilized in 70% EtOH for 1 minute, followed by a solution of 20% bleach for 15 minutes, under agitation on a rotary shaker. Sterilized seeds were planted 1cm deep into moist Cornell mix in 6-inch azalea-type ceramic pots, with 1-2 seeds per pot. Pots were placed in bench-mounted tanks with 4-5 inches of

standing water, such that the stem base of plants were never submerged in standing water. For initial germination, 70 pots were evenly spaced per tank, with two tanks per bench. At 2-3 weeks past germination, seedlings were thinned to one per pot, and at 6-8 weeks past germination, pots were moved for a wider-spaced final layout with 35 pots per tank. Pots of each of the replicates blocked together and individual accession layout randomized within replicates. Greenhouse conditions were as described in the previous section on germplasm purification.

Phenotyping was conducted as noted in the ontology and phenotyping table (Supplementary Table 2.2). In the 2006-2007 S0 growout (also see Chapter 1), 11 traits were measured on 94 RDP1 accessions; in the 2008-2009 S1 growout, 16 traits were measured on 95 RDP1 accessions; in the 2009-2010 S2 growout, 46 traits were measured on 40 RDP1 accessions, and in the 2010-2011 S2 growout, 25 traits were measured on 60 RDP1 accessions--the remainder of the panel not planted in 2009-2010. At eight weeks past the heading date (HD) for each individual plant, harvest/ratoon date (RD) phenotypes were measured, bagged seed was harvested, and plants were rationed to 5 inches above the soil surface. Post-ratoon measurements on shoot regrowth and panicle production were collected two weeks past the RD (Supplementary Table 2.2).

Aluminum tolerance screen

A subset of 68 out of 95 *O. rufipogon* accessions in RDP1 wild panel the and two *O. sativa* witness lines, cv. 'Azucena' (tolerant check) and cv. 'IR64' (susceptible check) were phenotyped for their root growth in a solution containing an Al^{3+} activity of 160uM (stress) and 0uM active Al^{3+} (control) hydroponic solutions using the phenotyping platform described by Famoso (2010).

This 68 accession subset of the RDP1 wild panel was the entirety of accessions for which there were enough S3 generation SSD seed (>60 seed) to conduct this screen. For each accession, 40 seeds were surface sterilized in 20% bleach for 15 minutes, then rolled in germination paper and allowed to stand vertically in a tray of ddH2O to germinate for three days.

At three days past germination, ten seedlings from each accession with uniform coleoptile and radicle growth were transferred to foam strip floats in 30L tubs of 0uM Al^{3+} and 160uM Al^{3+} hydroponic solution. Each tub held eight strips of 10 seedlings each, with every tub having a randomized selection of six RDP1 wild panel lines plus the two witness lines. Plants were both germinated and grown under controlled growth chamber conditions with 12H, 30°C day/12H, 26°C night cycle (day neutral). Seedlings were imaged individually at 5, 10 and 13 days of growth in hydroponic solution, and images were cropped and grayscaled, and root trait measurements were extracted from the images using Root Reader 2D software developed by Clark et al. (2013).

The relative root growth indices for relative root growth based on total root length (RRG-TRL), relative root growth based on longest root length (RRG-LRL) and relative total root count (RTRC) were calculated based on Root Reader2D output values from control over stress measurements. These estimates of RRG were used as estimates of Al^{3+} tolerance. The experiment was carried out from March 20th to April 8th, 2012, at the USDA Robert Holley Center for Agriculture and Health in Ithaca, New York.

The screen was co-designed and conducted with Juan-David Arbelaez. Randy Clark assisted with image capture and analysis, and Cheryl Utter with trait measurement.

Phenotype data analysis

Raw line means for all traits were calculated and the statistical software program JMP Pro 10 (JMP PRO 10) was used to calculate means, standard deviations and oneway ANOVA analyses for traits grouped according to cluster membership at K=4 (*O. rufipogon* only analysis, 50% admixture cutoff; designated for 60/95 RDP1 wild accessions only). For all 38 CU morphological and developmental phenotypes for which were measured over more than one generational growout, trait line means by growout were also averaged across all growouts (S0-S2) to produce line means across generations.

DNA extraction and library preparation for resequencing and genotyping

Seed from the S2 generation accessions of the RDP1 panel was planted out exclusively for tissue collection and DNA extraction for resequencing and SNP discovery for construction of the high density rice array (HDRA; (McCouch *et al.* 2016)).

The S2 generation panel growout for genotying, DNA extraction, and library preparation was conducted by Chih-Wei Tung.

High density rice array (HDRA) construction

The HDRA was designed and constructed to incorporate allelic variation across all five *O. sativa* subpopulations and the *ORSC* as described in the publication by McCouch et al (McCouch *et al.* 2016)

Diversity panel genotyping on HDRA

Out of the diversity panel of 95 accessions, 91 of these were genotyped on the HDRA. Three

genotypic dataset versions were generated from the HDRA genotype calls: v0 had a non-missing allele cutoff of 0.85, with a 0.80 cutoff for v2, and 0.95 for v.2 stringent, such that any accessions which had a non-missing allele percentage below the cutoff value were dropped from the dataset. V0 of the dataset included only 68 of the 95 accessions in the RDP1, while v2 had 93 out of 95, and v2 stringent, 67 out of 95.

Genome-wide association analysis

Genome-wide association analysis using the efficient mixed model association expedited analysis (EMMA-X) which accounts for sample structure was conducted with all of the CU morphology and development phenotypes and the aluminum tolerance screen data on all three genotypic dataset versions, the GWAS results presented here were analysed with the v2 stringent HDRA genotypic dataset.

GWAS interpretation and candidate gene identification

Both forward and reverse approaches were used in order to identify possible candidate genes in regions identified as having significant trait associations with GWAS. In the forward approach, significant GWAS 'peaks' were identified, being defined as a region with a top, 'peak,' SNP having a $-\log_{10}$ P-value less than 1×10^{-4} and three or more 'supporting SNPs,'defined as SNPs with a $-\log_{10}$ P-value less than 5×10^{-3} within 500kb upstream or downstream of the peak SNP. A 100kb window (50kb upstream and 50kb downstream) for each peak SNP was delineated for for further candidate gene searching using *O. sativa* MSU rice genome annotated loci. In the reverse approach, cloned and characterized *O. sativa* genes associated with each phenotype were identified through literature searches and *in silico* by rice, maize, and arabdopsis database

searches for trait ontology (TO) and gene ontology (GO) terms relating to each phenotype. These locations of these candidate genes were plotted on the Manhattan plots of trait associations to determine if any of these genes colocalized with significant associations identified in our study.

Results

In this study, an association mapping panel of 95 diverse *ORSC* lines was developed, and successive generations of the panel grown out for phenotypic measurement and pure line accession generation using the single seed descent (SSD) method. Multiple phenotypic screens for morphological and developmental traits including: 1) a series of four panel growouts representing three generations (S0-S2) in the greenhouse at CU, phenotyped for 38 characters representing a wide range of vegetative and reproductive morphology and development-related traits, and 2) a hydroponic screen for aluminum tolerance in the panel accessions. Ranges in trait values, association with subpopulation structure, geographic origin, and pairwise correlations between line means for all traits were calculated. The panel was also genotyped using a high density rice SNP array and GWAS was performed for all traits phenotyped using line mean average values. Peak regions for significant positive phenotypic associations were identified from GWAS results and candidate genes identified within a 100kb window of peak SNPs.

Phenotypic analysis for developmental and morphological characters

Line means for all traits across all generations surveyed were calculated. Histograms of these line means across all generations are shown for a selection of traits related to plant development, vegetative and reproductive morphology, and life habit are shown in Figure 2.1. All of the traits

shown exhibit continuous ranges of trait variation. According to the Tukey outlier analysis boxplot, for nearly all of the nine shown traits, with the exception of plant height and culm angle, which was phenotyped as a categorical trait, there were outlier accessions with values <1.5x the 1st interquartile range or >1.5x the 3rd interquartile range.

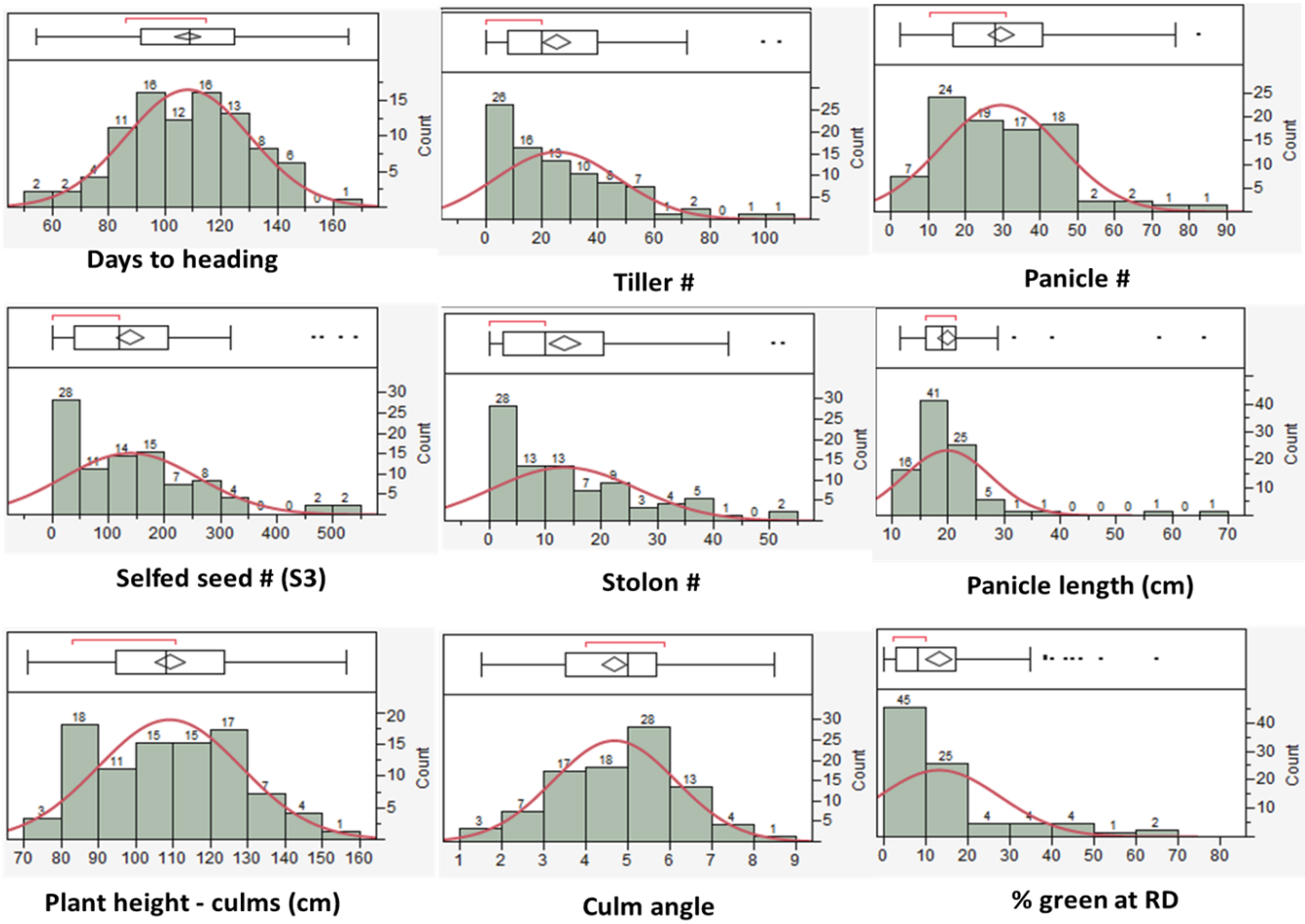


Figure 2.1 Range of variation in RDP1 wild panel for select traits pertaining to development, vegetative and reproductive morphology, and life habit. Red line indicates a normal curve fitted to the data. Top box plot indicates the 1st -3rd quartile, with the diamond representing the upper and lower 95% of the mean, and the line representing the median, and whiskers to the limits of 1.5x the interquartile range. Outliers are represented by points outside the whiskers. Abbreviations: RD – ratoon date (8 weeks post heading date).

Developmental and morphological trait correlations and trait complexes

Analysis of pairwise correlations between developmental and morphological vegetative and reproductive traits grouped by level of positive correlation revealed four major trait complexes (Figure 2.2). The trait complex I had high positive correlations between days to heading (DTHD) and various measures of plant height (related to tillers (T) –upright culms) and plant length (related to stolons (S) horizontal or near-horizontal culms). Trait complex II consisted of a suite of mostly domestication-related traits showing high positive correlations between awn presense and length (AWNPLU), hull color (HULCL_mature_seed), culm angle (CULM_ANGLE_CLASS), pericarp color (Pericarp_color), seed shattering (SDSH), stem color (StmCol_Binary), panicle number (PNNB), and culm number (CUNO). The third complex (III) grouped four different measures of tiller number, including an overlap with CUNO in the second complex, with two tiller reproductive capacity traits, panicles on tillers (P_on_T) and maximum number of panicles on a single tiller (Max_NumP_on_T), and an indicator of lateral branching on tillers (Lat_br_on_T). The fourth and largest major trait complex (IV) grouped five measures of stolon presence with culm habit (CULMHAB; a combined metric of culm angle and stolon presence), and two stolon reproductive capacity traits, panicles on stolons (P_on_S) and maximum number of panicles on a single stolon (Mar_NumP_on_S).

Of the minor trait complexes, three panicle length measurements, panicle length on the longest culm (PNLG) and panicle length on stolons (Pan_Lgth_Stolons) and tillers (Pan_Lgth_Upright) were highly correlated. Percent green leaf matter at harvest (Percent_Green_RD) and flag leaf width (FLFLWD) also had an unexpectedly high positive correlation, however this may be a factor of low sample size for FLFLWD (n=54). Most surprising was the grouping of S3 selfed and non-selfed (from panicles bagged after anthesis) seed number (S3_Max_SS and

S3_Max_nonSS) apart from any of the larger trait complexes, particularly in relation to higher seed yields that might expected to be under selection as a domestication-related trait (group II), or with the tiller panicle traits (group III) as a factor in a complex that might be correlated with a seed producing, sexually reproductive habit.

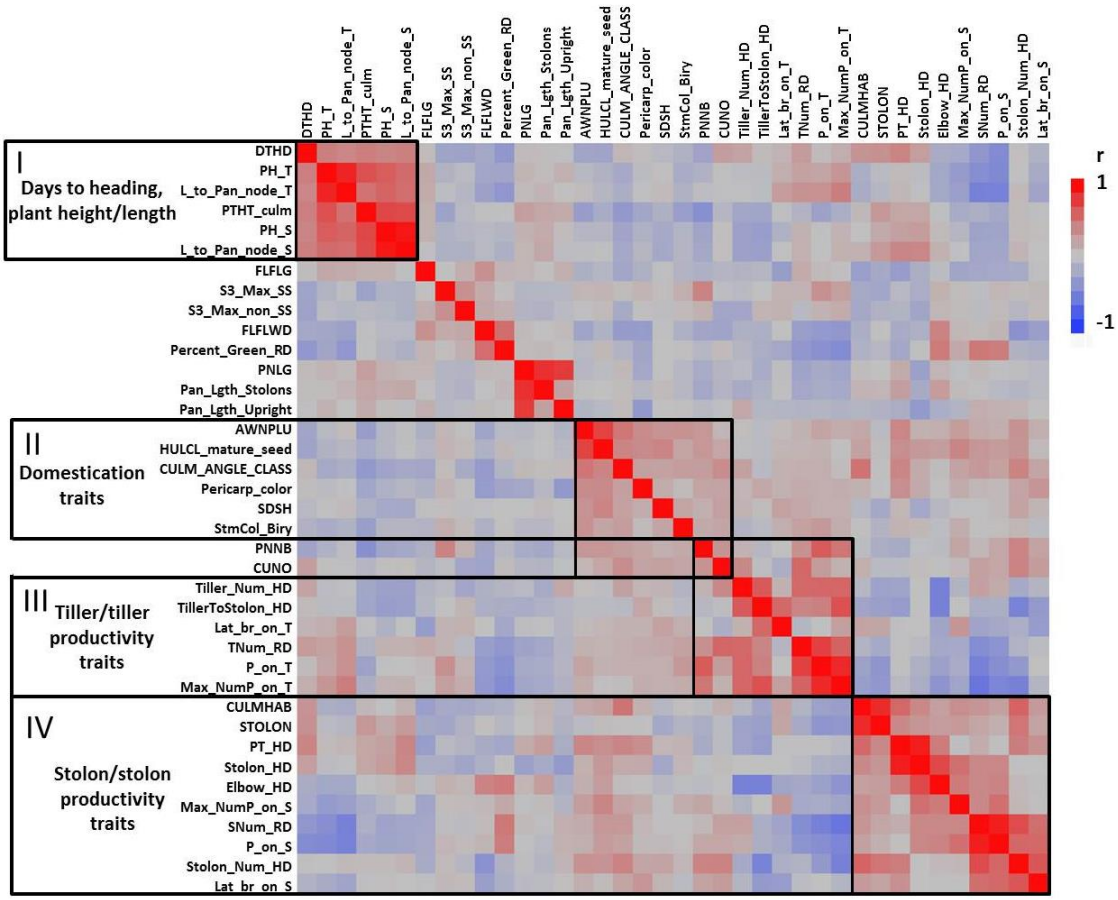


Figure 2.2 Heat map of the pairwise correlations between the generational average line means of 38 morphological and developmental traits, grouped according to strength of positive correlation. Four major trait complexes are apparent.

Developmental and morphological trait variation with subpopulation membership

Based on initial analyses showing some correlation of vegetative and reproductive traits with *ORSC* subpopulation membership (see Chapter 1), we strove to explore a greater range of morphological and developmental trait variation in our greenhouse-based growouts to determine whether the weak variation between certain traits and subpopulation membership held up in data collected from multiple years and across multiple generations. Figure 2.2 displays the 12 traits with the highest R^2 values (0.484-0.136) and greatest significance (P<.0001-0.5). These traits included developmental traits such as days to heading (DTHD), the percentage of green shoot tissue at 8 weeks post-heading date (% green), reproduction-related traits such as panicle number per plant (PNNB) and panicle length (PNLG), and also domestication related traits such as pericarp color, culm angle and stolon absence/presence (CULMHAB), and awn presence and length (AWNPLU).

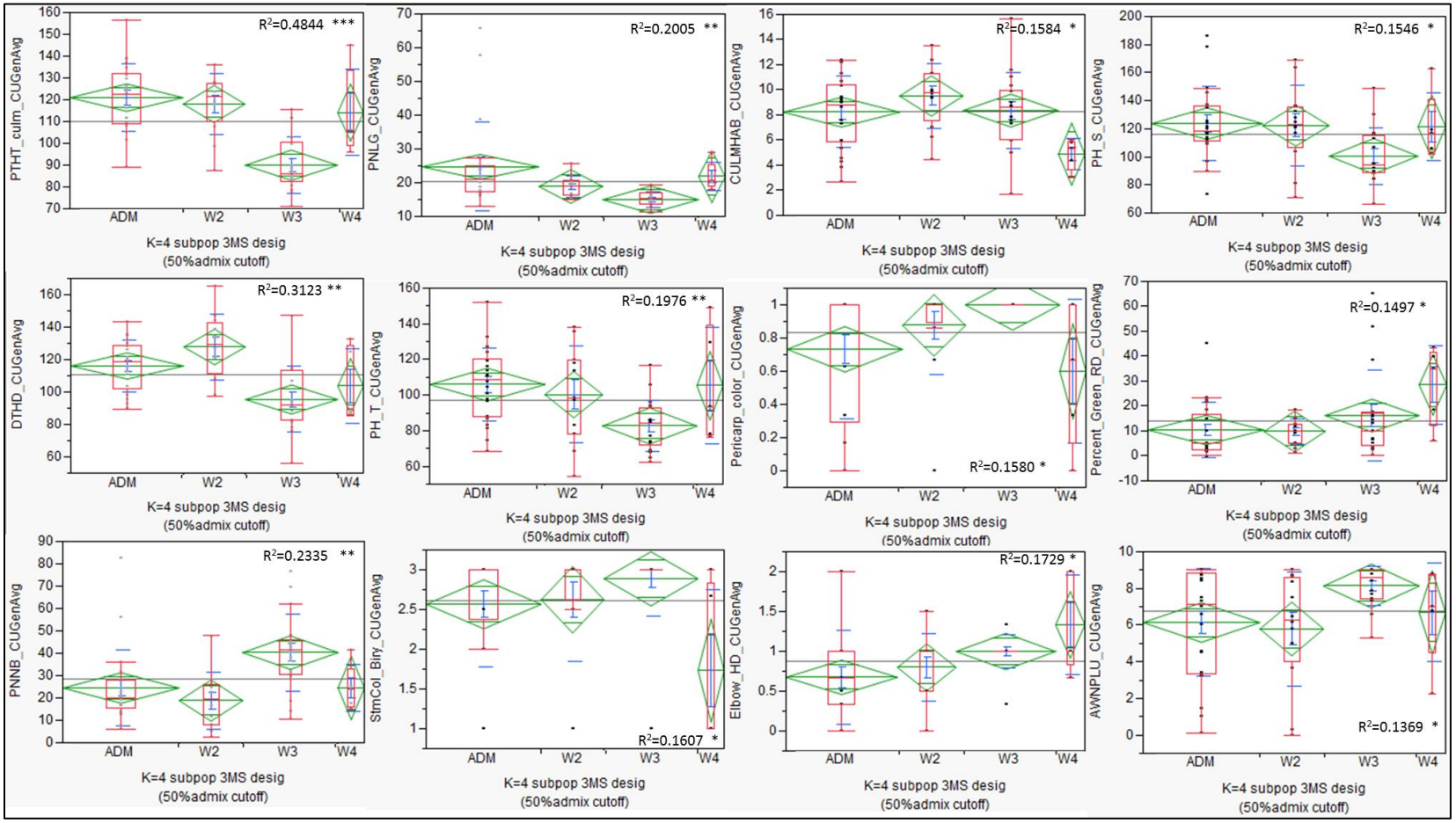


Figure 2.3 Box-plots of the 12 morphological traits showing significant variation according to subpopulation membership. The top and bottom of the green diamonds indicate the 95% confidence interval, with the width varying by the sample number. Red boxes indicate the 1st -3rd quantile, with whiskers to the limits of 1.5x the interquartile range. The blue bar indicates the mean error, and the flanking disconnected blue lines indicate the standard deviation. *P<0.05, **P<0.01, *** P<0.0001

GWAS on developmental and morphological traits

GWAS using the v.2 stringent genotype set revealed that association mapping in our RDP1 wild diversity panel was possible even with a limited number of lines in the genotypic dataset (n=67). With analyses on simple binary traits such as seed shattering, stem color, hull color and pericarp color, significant SNP associations with $-\log_{10}$ P-values as high as 1×10^{-14} were identified. As shown for hull color and pericarp color, (Figure 2.4A and B, respectively), major GWA peaks colocalized with known cloned and characterized *O. sativa* genes for those traits, cross-validating the veracity of the associations. For both hull color and pericarp color, other minor peaks above the 1×10^{-4} $-\log_{10}$ P-value threshold can also be identified; these may correspond to loci of minor effect in those phenotypes and may be unique to the *ORSC*, differing from the known genes or QTL identified in cultivated rice.

For other traits measured according to an ordinal or continuous scale, associations were less well-resolved and less powerful, with peak SNP associations ranging from a maximum $-\log_{10}$ P-value of 1×10^{-4} to 1×10^{-9}. Association mapping for culm angle, measured on a categorical scale of 1-9, shows a peak that may colocalize with the region containing the *PROG1* gene for prostrate growth, originally identified in an *O. rufipogon* accession from China (Tan *et al.* 2008), as well as two other peaks above the significance thresholdof 1×10^{-4}, indicating possible associations on chromosomes 1 and 10. Interestingly enough, there are no peaks colocalizing with the other three known genes for tiller angle in rice, *LAZY*, *REH1*, and *TAC1,* all of which were cloned and characterized from different accessions of *O. sativa* (Xu *et al.* 2005; Yoshihara and Iino 2007; Yu *et al.* 2007), which indicates that there is no apparent variation for these genes in our wild diversity panel, suggesting that the allelic variants in cultivated rice likely arose post-domestication.

In an example of association mapping for a continuously measured trait, the percentage of green leaf area on the plant at 8 weeks past the heading date (percent green), there are many sparsely supported SNPs with high association to the trait (Figure 2.6). These could be a result of spurious association from inaccurate trait phenotyping, as the trait was visually assessed in 10% intervals. However given that only 11 out of 85 phenotyped individuals had a score of higher than 30% (Figure 2.1), and that these outlier individuals came from two different subpopulations as well as admixed-designated group, it may be that the alleles for these traits are subpopulation specific and found at too low a frequency in our mapping panel to detect significant associations. There are a few significant SNPs that appear to colocalize in the regions of two of the known *O. sativa* genes for delayed leaf senescence in rice, *DOS1* (Kong *et al.* 2006) and *SGR1* (Cha *et al.* 2002); however they are not strongly supported with underlying significant SNPs within a 500kb window and $-\log_{10}$ P-values of 5×10^{-3} requirements to be considered colocalized.

.

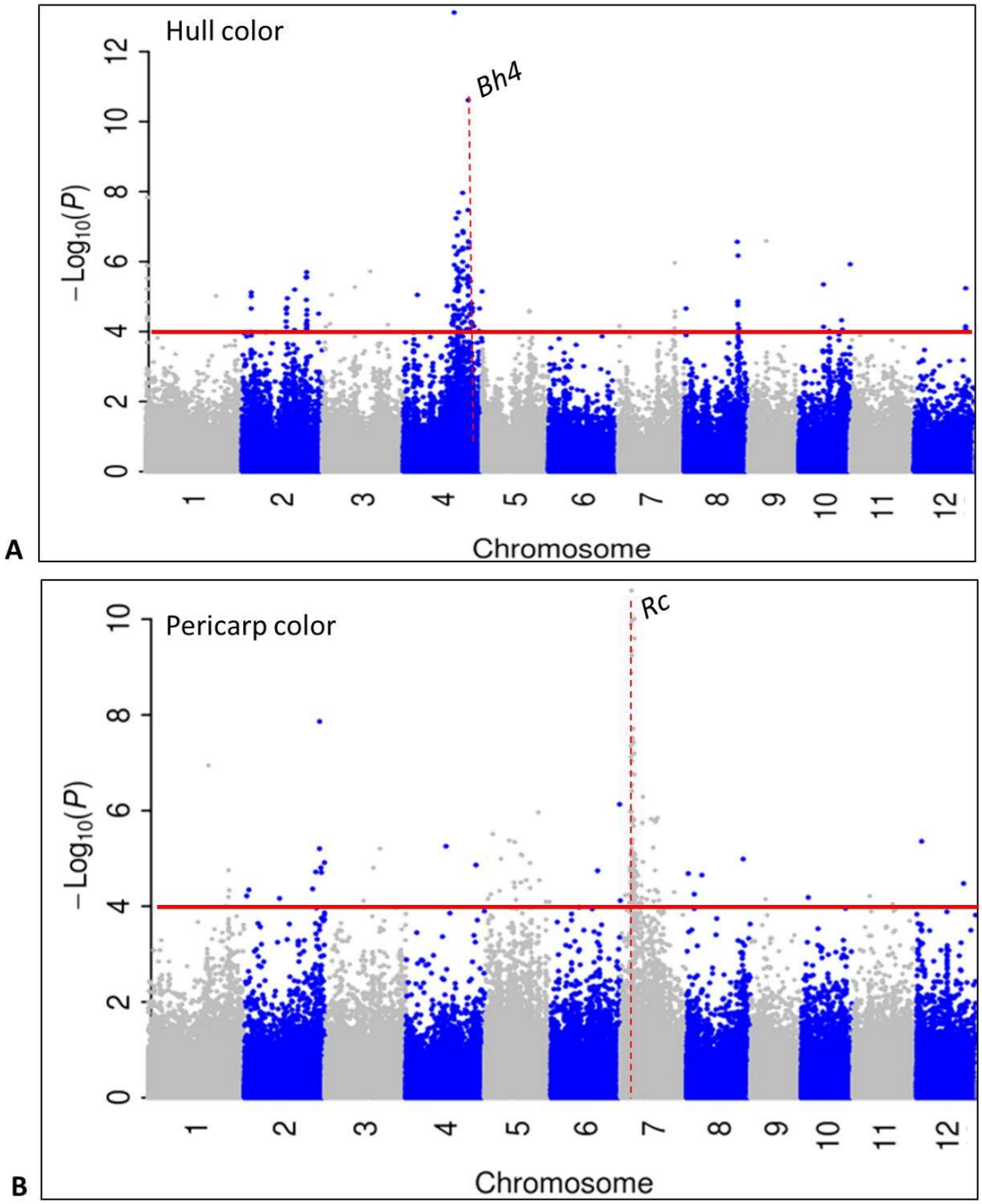


Figure 2.4 Manhattan plots of EMMA-X genetic mapping of A) hull color and B) pericarp color in the RDP1 wild panel. Each dot represents the $-\log_{10}$ P-value of the association between a SNP and the phenotype. The horizontal solid red line represents an arbitrary P-value significance threshold at 1×10^{-4}. Positions of known cloned and characterized rice genes for the respective traits are indicated by the vertical red dotted lines.

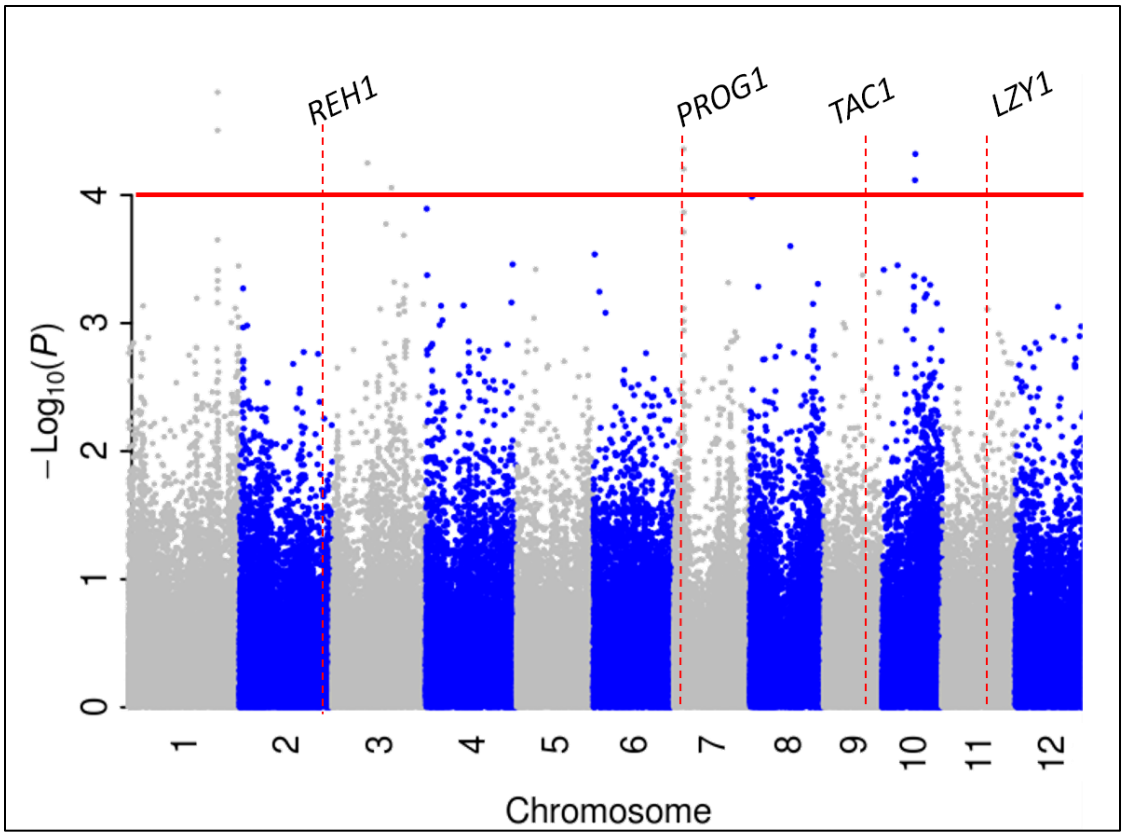


Figure 2.5 Manhattan plots of EMMA-X genetic mapping of culm angle in the RDP1 wild panel. Each dot represents the $-\log_{10}$ P-value of the association between a SNP and the phenotype. The horizontal solid red line represents an arbitrary P-value significance threshold at 1×10^{-4}. Positions of known cloned and characterized genes for the respective traits are indicated by the vertical red dotted lines.

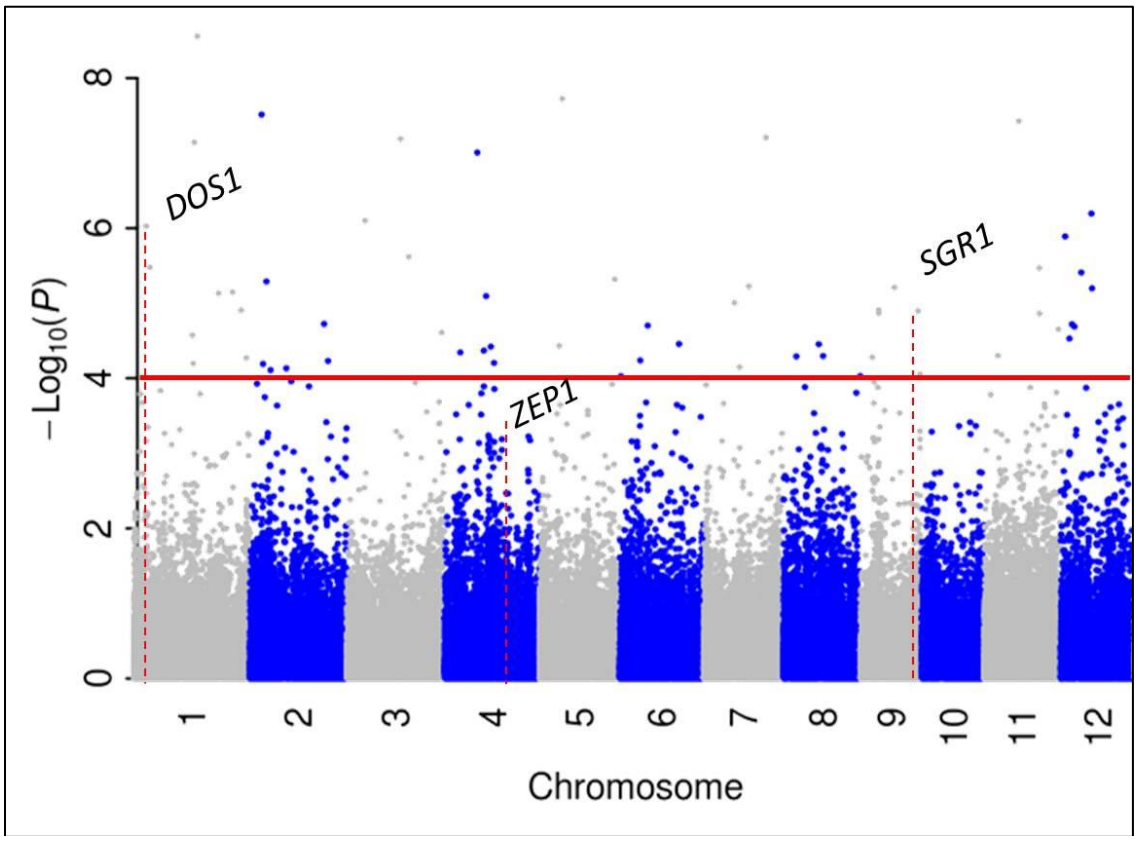


Figure 2.6 Manhattan plots of EMMA-X genetic mapping for the percentage of green vegetative material at harvest (% green) in the RDP1 wild panel. Each dot represents the $-\log_{10}$ P-value of the association between a SNP and the phenotype. The horizontal solid red line represents an arbitrary P-value significance threshold at 1×10^{-4}. Positions of known cloned and characterized genes for the respective traits are indicated by the vertical red dotted lines.

Aluminum tolerance screen phenotypic results

When compared with the susceptible ('IR64') and tolerant ('Azucena') *O. sativa* check varieties, it appears that the *ORSC* as a whole are much more susceptible to aluminum toxicity as indicated by relative root length (RRL) data (Figure 2.7). There are, however, three outlier accessions that are more aluminum tolerant than the tolerant check, two of which are from East Asia (China and Taiwan) (Figure 2.7C). Although neither of these are from the W4 subpopulation localized to East Asia (Figure 2.7B), there appears to be a positive, but non-significant correlation between the W4 subpopulation and aluminum tolerance as measured by RRL-dAvg. The lack of a statistically significant association may be due to the fact that subpopulation membership was only assigned to 40 out of 68 phenotyped *ORSC* accessions, due to the lack of complete marker set (SSR, MITE, and SINE, see Chapter 1) coverage.

When the 68 accessions screened for aluminum tolerance are color coded by RRL-dAvg values as an indicator of aluminum tolerance and mapped on a soil pH map according to their geographic origin (Figure 2.8), two interesting points are apparent. Firstly, there is a significant geographic division between higher pH non-aluminum toxic soils in South Asia and on the other side of the Himalayan mountain range, the low pH, highly aluminum toxic soils throughout continental and archipelagic Southeast Asia. Secondly, though mostly of the accessions in the screened subset of the RDP1 wild panel are highly susceptible to aluminum and are spread throughout the geographic range of the species complex, the few aluminum tolerant accession (blue, RRL-dAvg $\geq$0) are all from low pH soils with an ostensibly high selection pressure for the development of aluminum toxicity resistance mechanisms.

Given the general high susceptibility to aluminum of the *O. rufipogon* complex, it is possible that these outlier accessions may either represent the ancestral *japonica*-like wild population from which the highly aluminum tolerant *O. sativa japonica* subpopulation was domesticated. Alternately, these highly aluminum tolerant lines may carry *japonica* alleles for aluminum tolerance through back introgression with *O. sativa.* Haplotype analysis of these outlier accessions will need to be conducted to confirm or disprove these hypotheses.

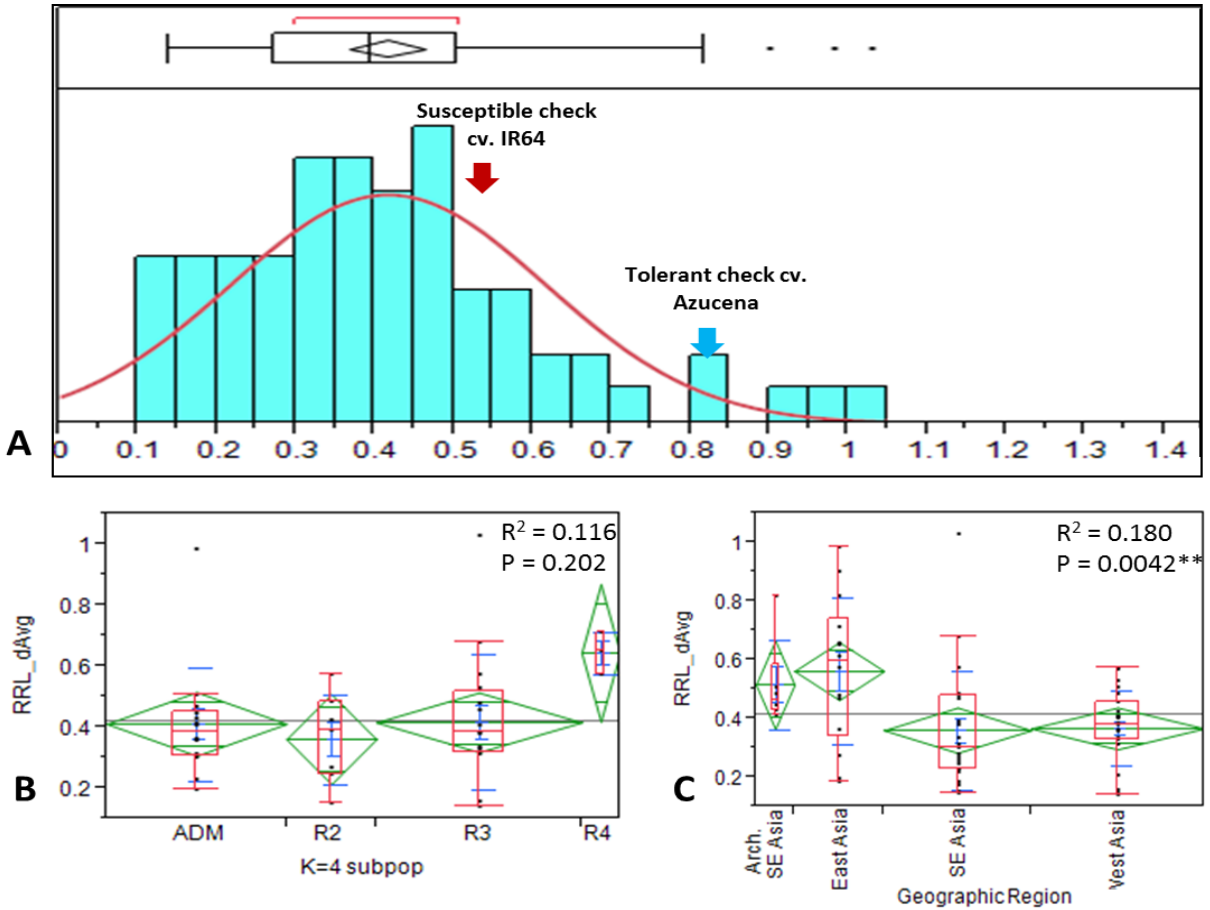


Figure 2.7 Range of variation for RRL_dAvg and variation with subpopulation membership and geographic region. A. Range of the average of day 5, 10, and 13 relative root length (RRL_dAvg) values for the 68 accession subset of the RDP1 wild panel and two *O. sativa* check varieties. Box plots of the RRL_dAvg values by subpopulation membership (B) and geographic region (C). **significant at P<0.001

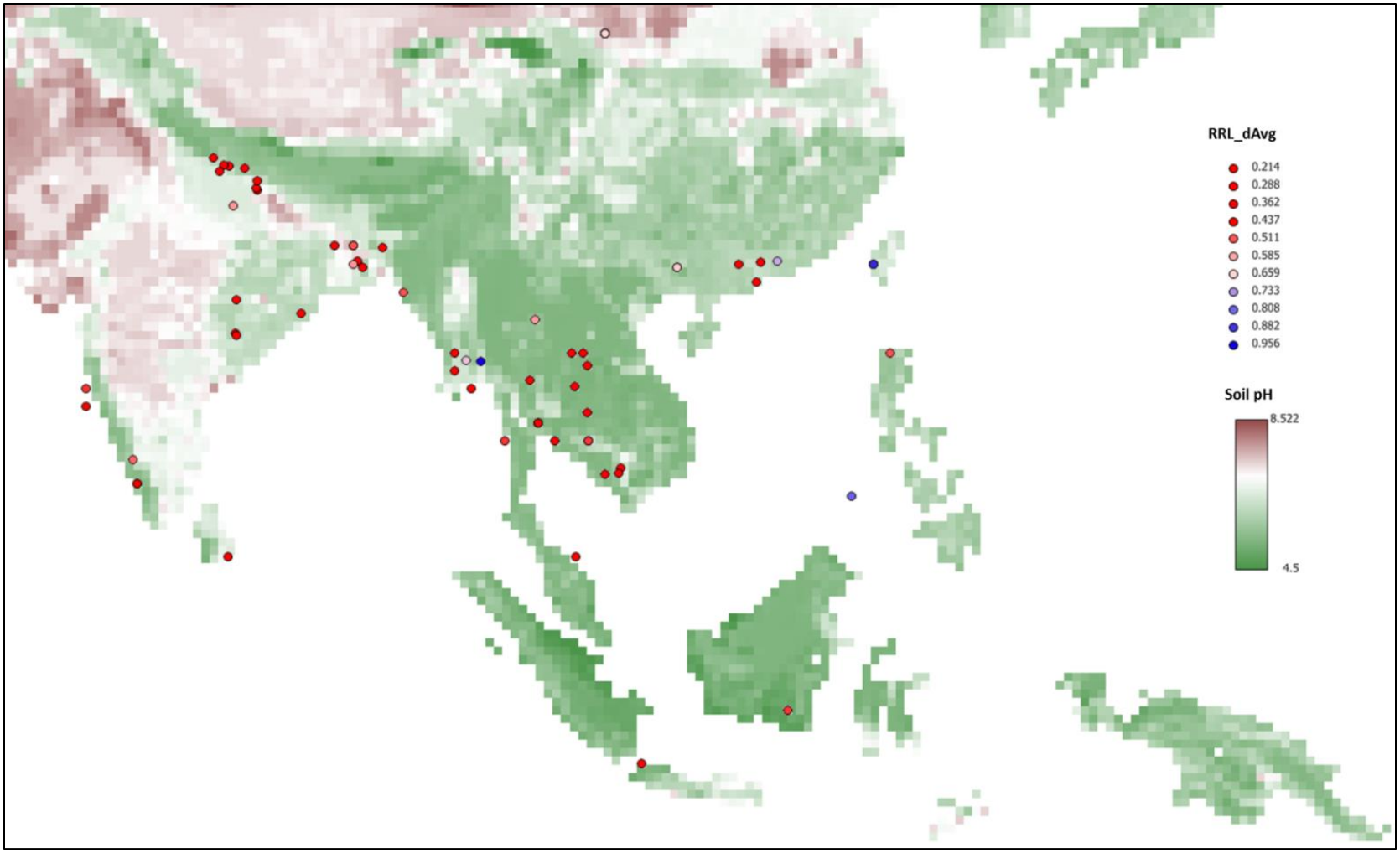


Figure 2.8 Soil pH map showing points representing the 68 RDP1accessions positioned according to their geographic origin and colored according to their RRL_dAvg values. Lower RRL_dAvg values, representing low aluminum tolerance are in redder tones and higher values representing higher aluminum tolerance (at ≥0.8)in bluer tones.

Correlations between shoot traits and aluminum tolerance

Investigating correlations between below-ground aluminum tolerance root traits and above-ground shoot development and morphological traits may hint at possible trait complexes related to aluminum tolerance development in the *ORSC*. The correlation matrix between the greenhouse screened shoot and reproductive traits sorted according to highest positive to negative correlation with RRL values for each time point and the day average reveals plant height, several indicators of tillering ability, and wild-related characteristics such as open culm angle, dark hull and pericarp color, and awnness as positively correlated with aluminum tolerance (Figure 2.9).

More telling are the traits that have the highest negative correlation with aluminum tolerance: high percent green leaf tissue--a trait correlated with lower plant senescence and high perenniality, as well as several indicators of stolon presence. Given that stolons in the *ORSC* are indicative of clonal vegetative reproduction in an aquatic ecosystem, the flooded rhizosphere of which would be pH neutral with low levels of phytotoxic Al^{3+}, it seems logical that highly stoloniferous, perennial, aquatic plants would have low levels of aluminum tolerance lacking the selection pressure of a aluminum toxic rhizosphere. Conversely, more upright, highly tillering plants in seasonally flooded soils would be more likely to be exposed to a non-flooded, acidic, aluminum-toxic soil environment and thus be under greater selection pressure to develop mechanisms of aluminum tolerance. Thus the overall regional acid soil geography of the region as shown in Figure 2.8 may be less highly correlated with aluminum tolerance in the *ORSC* then the hydrology of the microhabitat to which different populations are adapted. Given the huge range of variation for characters such as stolon development and other indicators of a perennial

vs. an annual growth habit, as well as the yet unexplored component of phenotypic plasticity in the species complex, controlled experiements would be needed to further test the veracity of these shoot trait correlations with aluminum tolerance. If certain aboveground shoot or reproductive characters are found to be consistently correlated with aluminum tolerance or susceptibility in the *ORSC*, perhaps these also could be used as an alternative means of selecting rare, highly aluminum-tolerant wild germplasm for use as parents in breeding programs.

GWAS for aluminum tolerance in the *ORSC*

Genomic mapping of RRL-dAvg as a measure of aluminum tolerance in the wild complex reveals low colocalization of significant SNP peaks with known genes involved in rice aluminum tolerance, but also indicates the presence of significant peaks which may underlie alleles unique to the wild complex (Figure 2.10). Significant peaks colocalize with the location of *ALS1* on chromosome 3, a gene than encodes an ABC transporter involved in $Al3^{+}$ vacuolar sequestration, as well as a MATE efflux gene on chromosome 10 involved in malic acid root efflux. Additionally, there are significant peaks on chromosomes 1, 2, and 11 which require further investigation to determine possible candidate gene loci underlying those regions.

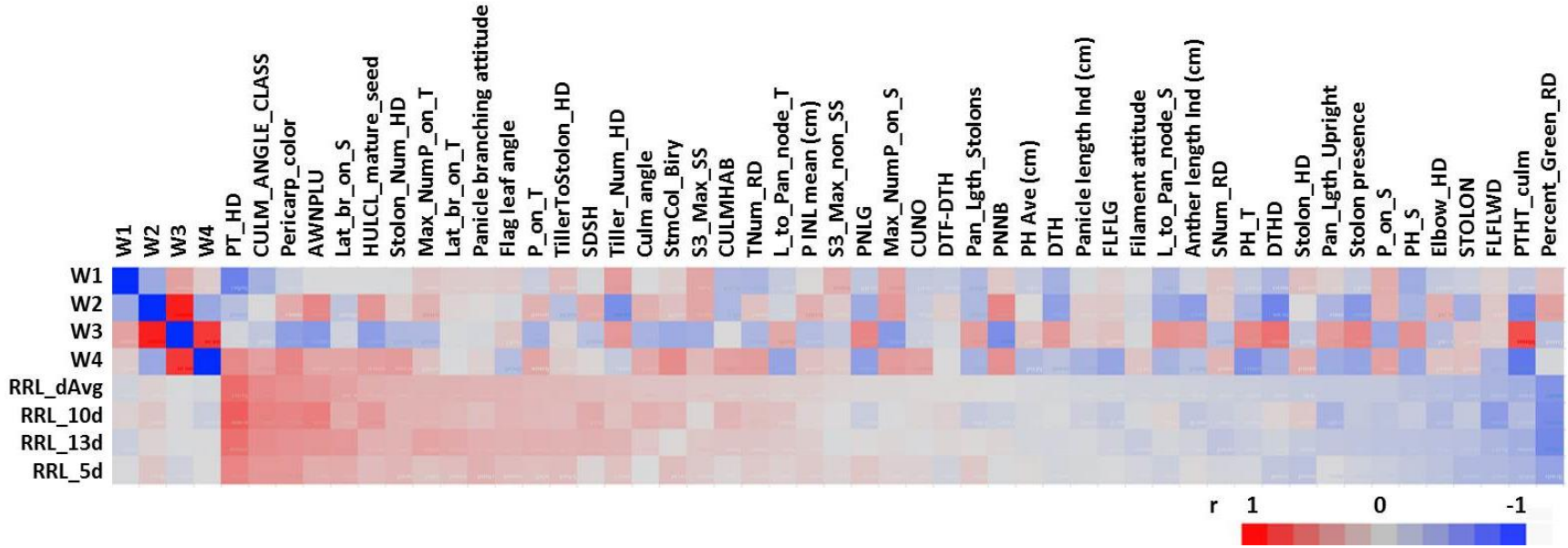


Figure 2.9 Heat map of the pairwise correlations between the generational average line means of 38 morphological and developmental traits. Morphological traits are sorted according to a higher to lower (L to R) correlation with relative root length (RRL) phenotypes for each timepoint and across all days (bottom four rows).-

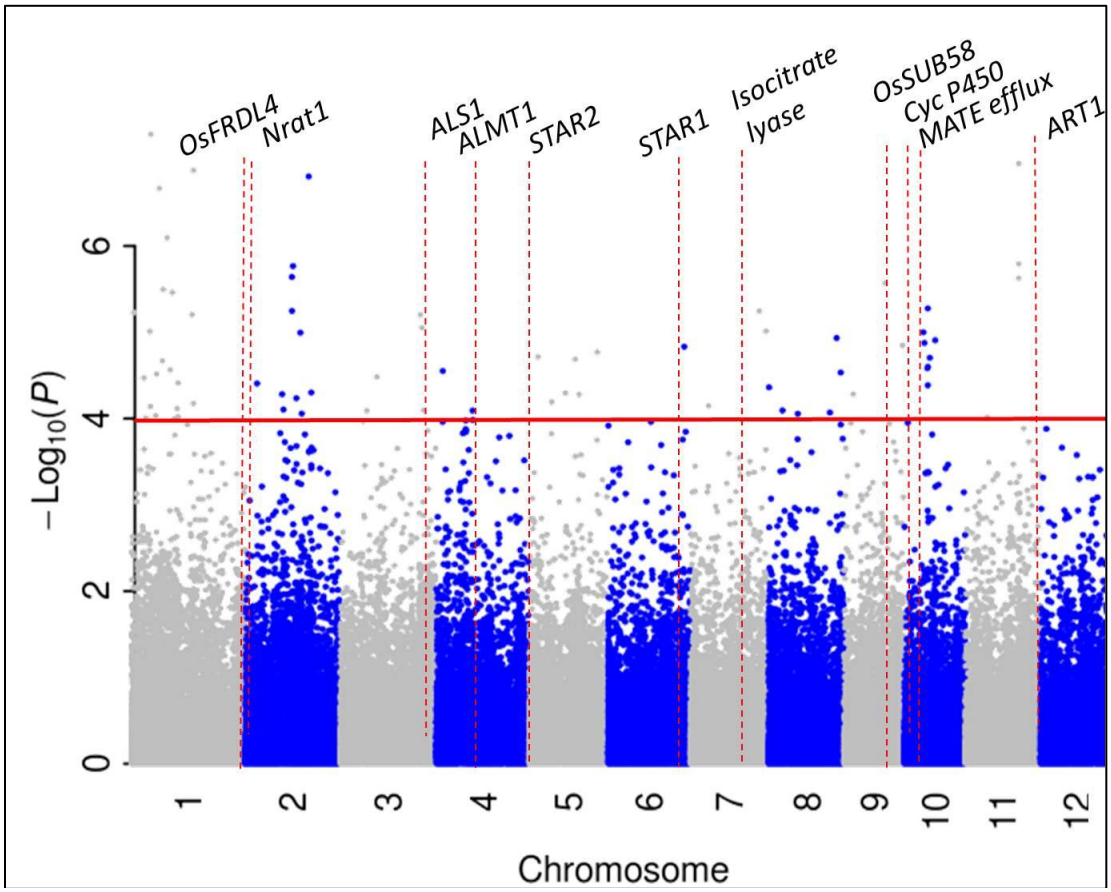


Figure 2.10 Manhattan plot of EMMA-X genetic mapping for RRL_dAvg in the RDP1 wild panel. Each dot represents the $-\log_{10}$ P-value of the association between a SNP and the phenotype. The horizontal solid red line represents an arbitrary P-value significance threshold at 1×10^{-4}. Positions of known cloned and characterized genes for the respective traits are indicated by the vertical red dotted lines.

Discussion

In this study, we set out to achieve 6 main objectives: 1) the construction of an immortal wild diversity panel for GWAS, 2) to survey the range of variation contained in this panel for various morphological and developmental traits, focusing on life habit, reproductive habit, RSA, aluminum tolerance, and seedling vigor, 3) to evaluate trait correlations signifying multitrait complexes and identify correlations with larger genetic and developmental networks, subpopulation structure and geography, 4) to identify sources of novel allelic variation for these traits using GWAS, 5) to compare the genetic basis of complex trait variation between the wild species complex and the cultivar, and 6) to use these data on trait variation, correlations, and allelic variation to test hypotheses on rice evolution and domestication.

We simultaneously conducted growouts of our panel for SSD pureline accession development and developmental and morphological trait phenotyping. Using data from a subset of the phenotypic screens, namely a multi-year evaluation of shoot and reproductive traits, and a hydroponic screen for aluminum tolerance, we present initial analyses on the range of trait variation contained within the diversity panel for these traits, conduct pairwise multivariate correlation matrices to identify trait complexes, and explore these complexes in consideration of subpopulation specific, geographic, and ecological differences within the wild complex and in comparison with *O. sativa*.. We have conducted preliminary GWAS on these traits and both validated the diverse phenotypic and genetic composition of the panel and our ability to detect significant associations in a wild crop relative by detection of significant GWAS hits colocalizing with known cloned and characterized rice genes for several traits of interest, even

with a very limited number of less than 100 accessions. Additionally we have detected significant GWAS hits for most traits that appear to be unique to the wild species complex and will need to be explored in further detail to identify possible underlying candidate gene loci.

Although we haven't yet been able to fully analyze these GWAS results to compare the genetic basis of complex trait variation in the wild vs. the cultivated species, we were able to formulate and attain foundational information to test specific evolution and domestication-related hypotheses using differential GWAS hits and morphological, developmental, and aluminum tolerance multitrait complexes we identified in this study. The results presented here demonstrate our initial success in constructing, phenotyping, genotyping, and conducting GWAS on a small, diverse, wild crop relative association mapping panel. In conducting these initial analyses, we set a standard protocol for the GWAS of the remining phenotypic screens on greenhouse and field-based life habit evaluation, 3D RSA, seedling vigor, and ionomics mentioned in the beginning of this chapter, pave the way for further candidate gene analyses of the GWAS hits already identified, and identify possible multitrait complexes and correlations with subpopulation structure, geography, and ecology to help further leverage the utilization of the valuable allelic and phenotypic variation in the *ORSC* germplasm for breeding improved varieties of cultivated rice.

Supplementary Table 2.1 Passport, genotypic, and phenotypic screen information of the 95 Rice Diversity Panel 1 (RDP1) *ORSC* accessions

NSF-TV ID	IRGC #	Species	Genotyped on HDRA*	3D RSA (seed ID)	Al tol screen (seed ID)	Phenotyped in morphological screens***	Seedling vigor screen
401	80433	*O. rufipogon*	401_C2_S2	401C2_3_S3	_	1, 2, 3, 4, 5, 7, 9	401C2_3_S3
402	80539	*O. spontanea*	402_B2_S2	402B2_3_S3	402_B2_2_S3	1, 2, 4, 5, 7, 8, 9	402B2_3_S3
407	80742	*O. rufipogon*	407_C2_1_S2	_	_	1, 2, 4, 5, 7, 8	_
410	80759	*O. nivara*	410_A2_S2	410A2_2_S3	410_A2_2_S3	1, 2, 4, 5, 7, 8, 9	410A2_2_S3
413	81850	*O. nivara*	413_A1_S2	413A1_1_S3	413_A1_1_S3	1, 2, 3, 4, 5, 7, 9	413A1_1_S3
415	81909	*O. spontanea*	415_B1_S2	415B1_3_S3	415_B1_3_S3	1, 2, 3, 4, 5, 7, 9	415B1_3_S3
416	81970	*O. spontanea*	416_A1_S2	_	_	1, 2, 4, 5, 7, 8	_
420	81984	*O. rufipogon*	420_A1_S2	_	_	1, 2, 4, 5, 7	_
427	82988	*O. rufipogon*	427_C1_S2	427C1_1_S3	_	1, 2, 4, 5, 7, 8	_
428	82989	*O. rufipogon*	428_A2_S2	428A2_2_S3	428_A2_2_S3	1, 2, 3, 4, 5, 7, 9	428A2_2_S3
431	82992	*O. rufipogon*	431_A3_S2	431A3_1_S3	431_A3_1_S3	1, 2, 4, 5, 7, 8	_
432	83794	*O. rufipogon*		_	_	1, 2, 4, 5, 7	_
433	83795	*O. rufipogon*	433_A3_S2	433A3_2_S3	433_A1_2_S3	1, 2, 4, 5, 7, 8, 9	_
435	86448	*O. rufipogon*	435_C2_S2	435C2_1_S3	_	1, 2, 4, 5, 7, 8	_
438	86476	*O. rufipogon*	438_B2_S1	438B2_1_S3	438_B2_1_S3	1, 2, 4, 5, 7, 8	_
442	93181	*O. nivara*	442_A1_S2	442A1_1_S3	442_A1_1_S3	1, 2, 4, 5, 7, 8, 9	442A1_1_S3
443	93183	*O. nivara*	443_B1_S2	443B1_2_S3	_	1, 2, 3, 4, 5, 7, 9	443B1_2_S3
444	93188	*O. nivara*	444_A2_S2	444A2_3_S3	444_A2_3_S3	1, 2, 4, 5, 7, 8, 9	444A2_3_S3
445	93189	*O. nivara*	445_B1_S2	445B1_2_S3	445_B1_2_S3	1, 2, 3, 5, 6, 7, 9	445B1_2_S3
446	93224	*O. spontanea*	446_A1_S2	446A1_1_S3	446_A1_1_S3	1, 2, 3, 4, 5, 7, 9	446A1_1_S3
449	100195	*O. nivara*	449_A1_S2	449A1_2_S3	449_A1_2_S3	1, 2, 3, 4, 5, 7, 9	449A1_2_S3
450	100916	*O. rufipogon*	450_A1_S2	450A1_1_S3	450_A1_1_S3	1, 2, 3, 4, 5, 7, 9	450A1_1_S3

NSF-TV ID	IRGC #	Species	Genotyped on HDRA*	3D RSA (seed ID)	Al tol screen (seed ID)	Phenotyped in morphological screens***	Seedling vigor screen
451	101508	*O. nivara*	451_B2_S2	451B2_2_S3	451_B2_2_S3	1, 2, 4, 5, 7, 8, 9	451B2_2_S3
453	103404	*O. rufipogon*	453_C2_S2	453C2_2_S3	453_C2_2_S3	1, 2, 4, 5, 7, 8, 9	453C2_2_S3
454	103821	*O. nivara*	454_A1_S2	454A1_3_S3	454_A1_3_S3	1, 2, 3, 4, 5, 7, 9	454A1_3_S3
457	103838	*O. nivara*	457_B3_S2	457B3_1_S3	_	1, 2, 4, 5, 7, 8	_
461	104057	*O. rufipogon*	461_A1_S2	461A1_2_S3	461_A1_2_S3	1, 2, 4, 5, 7, 8, 9	461A1_2_S3
465	104620	*O. spontanea*	465_A3_S2	465A3_3_S3	_	1, 2, 4, 5, 7, 8	_
467	104624	*O. rufipogon*	467_A2_S2	_	467_A2_1_S3	1, 2, 3, 4, 5, 7	_
472	104636	*O. spontanea*	472_B1_S2	472B1_3_S3	_	1, 2, 3, 4, 5, 7, 9	472B1_3_S3
477	104967	*O. spontanea*	477_A2_S2	477A2_1_S3	477_A2_1_S3	1, 2, 4, 5, 7, 8, 9	477A2_1_S3
481	105343	*O. nivara*	481_C3_S2	481C3_1_S3	481_C3_1_S3	1, 2, 3, 4, 5, 7, 9	481C3_1_S3
482	105349	*O. rufipogon*	482_A1_S2	482A1_3_S3	482_A1_3_S3	1, 2, 4, 5, 7, 8, 9	482A1_3_S3
483	105375	*O. rufipogon*	483_C2_S2	483C2_1_S3	483_C2_1_S3	1, 2, 3, 4, 5, 7, 9	483C2_1_S3
484	105388	*O. rufipogon*	484_B1_S2	484B1_1_S3	484_B1_1_S3	1, 2, 4, 5, 7, 8, 9	484B1_1_S3
487	105428	*O. nivara*	487_C2_S2	487C2_1_S3	487_C2_1_S3	1, 2, 4, 5, 7, 8, 9	487C2_1_S3
488	105491	*O. rufipogon*	488_B2_S2	488B2_3_S3	488_B2_3_S3	1, 2, 3, 4, 5, 7, 9	488B2_3_S3
490	105567	*O. rufipogon*	490_A_S1	_	490A2_S3	1, 2, 3, 4, 5, 7	_
492	105569	*O. rufipogon*		_	_	1, 2, 5, 6, 7	_
493	105706	*O. nivara*	493_A1_S2	493A1_1_S3	493_A1_1_S3	1, 2, 3, 4, 5, 7, 9	493A1_1_S3
494	105711	*O. rufipogon*	494_A1_S1**	494A1_1_S3	_	1, 2, 4, 5, 7, 8, 9	494A1_1_S3
495	105717	*O. nivara*	495_A1_S2	495A1_3_S3	495_A1_3_S3	1, 2, 3, 4, 5, 7, 9	495A1_3_S3
496	105720	*O. rufipogon*	496_A1_S2	_	_	1, 2, 3, 4, 5, 7	_
498	105735	*O. rufipogon*	498_A2_S2	498A2_1_S3	498_A2_1_S3	1, 2, 3, 4, 5, 7, 9	498A2_1_S3
499	105767	*O. rufipogon*	499_B3_S2	499B3_2_S3	499_B3_2_S3	1, 2, 3, 4, 5, 7, 9	499B3_2_S3
501	105821	*O. nivara*	501_B2_S2	501B1_2_S3	501_B1_2_S3	1, 2, 4, 5, 7, 8, 9	501B1_2_S3

NSF-TV ID	IRGC #	Species	Genotyped on HDRA*	3D RSA (seed ID)	Al tol screen (seed ID)	Phenotyped in morphological screens***	Seedling vigor screen
503	105843	*O. rufipogon*	503_C3_S2	_	503_C3_2_S3	1, 2, 4, 5, 7, 8	_
505	105855	*O. rufipogon*	505_A1_S2	505A1_1_S3	_	1, 2, 4, 5, 7, 9	_
506	105879	*O. nivara*	506_A2_S2	506A2_1_S3	506_A2_1_S3	1, 2, 3, 4, 5, 7, 9	506A2_1_S3
508	105890	*O. rufipogon*	508_C1_1_S2**	508C1_2_S3	508_C1_2_S3	1, 2, 4, 5, 7, 8, 9	_
509	105897	*O. rufipogon*	509_A2_S2	509A2_2_S3	509_A2_2_S3	1, 2, 3, 4, 5, 7, 9	509A2_2_S3
514	105956	*O. rufipogon*		514C1_1_S3	514_C1_1_S3	1, 2, 4, 5, 7, 8, 9	514C1_1_S3
523	106155	*O. nivara*	523_A1_S2	523A1_1_S3	523_A1_1_S3	1, 2, 4, 5, 7, 8, 9	523A1_1_S3
549	81881	*O. rufipogon*	549_A1_S1?	549A1_2_S3	_	1, 2, 3, 4, 5, 7, 9	549A1_2_S3
551	100596	*O. rufipogon/O. nivara*	551_C3_1_S2**	551C3_2_S3	_	1, 2, 4, 5, 8, 9	551C3_2_S3
553	100926	*O. rufipogon*	553_B1_S2	553B1_3_S3	_	1, 2, 3, 4, 5, 7, 9	553B1_3_S3
555	105349	*O. rufipogon*	555_B1_S2	555B1_1_S3	555_B1_1_S3	1, 2, 4, 5, 8, 9	555B1_1_S3
568	106263	*O. rufipogon*	568_A1_S2	568A1_1_S3	_	1, 2, 3, 4, 5	568A1_1_S3
592	80671	*O. rufipogon*	592_B3_1_S2**	_	_	1, 2, 4, 5, 8	_
600	100187	*O. sativa/O. rufipogon*	600_B3_S2	600B3_1_S3	600_B3_1_S3	1, 2, 3, 4, 5, 9	600B3_1_S3
602	100900	*O. nivara/O. rufipogon*	602_A3_S2	602A3_1_S3	602_A3_1_S3	1, 2, 4, 5, 8, 9	602A3_1_S3
605	100911	*O. nivara/O. rufipogon*	605_C3_S2	605C3_3_S3	_	1, 2, 4, 5, 8, 9	605C3_3_S3
665	100203	*O. rufipogon/O. sativa*	665_C1_S2	665C1_1_S3	665_C1_1_S3	1, 2, 3, 5, 6, 7, 9	665C1_1_S3
666	100211	*O. rufipogon*	666_B1_S2	666B1_1_S3	_	1, 2, 5, 6, 7, 8, 9	666B1_1_S3
669	100593	*O. nivara*	669_C2_S2	669C2_3_S3	669_C2_3_S3	1, 2, 3, 5, 6, 7, 9	669C2_3_S3
673	100647	*O. rufipogon*	673_A1_S2	673A1_1_S3	673_A1_1_S3	1, 2, 5, 6, 7, 8, 9	673A1_1_S3
676	100692	*O. rufipogon*	676_A1_S2	676A1_1_S3	676_A1_1_S3	1, 2, 3, 5, 6, 7, 9	676A1_1_S3
682	100904	*O. rufipogon*	682_C1_S2	682C1_2_S3	_	1, 2, 3, 5, 6, 7	_
683	100918	*O. nivara*	683_A1_S2	683A1_3_S3	683_A1_3_S3	1, 2, 5, 6, 7, 8, 9	683A1_3_S3
685	100923	*O. rufipogon*	685_A2_S2	685A2_2_S3	685_A2_2_S3	1, 2, 5, 6, 7, 8, 9	685A2_2_S3

NSF-TV ID	IRGC #	Species	Genotyped on HDRA*	3D RSA (seed ID)	Al tol screen (seed ID)	Phenotyped in morphological screens***	Seedling vigor screen
686	100926	*O. nivara*	686_C2_S2	686C2_3_S3	686_C2_3_S3	1, 2, 3, 5, 6, 9	686C2_3_S3
687	101193	*O. rufipogon/O. sativa*	687_A1_S2	687A1_2_S3	_	1, 2, 5, 6, 7, 8, 9	687A1_2_S3
691	101967	*O. nivara*	691_A2_S2	691A2_2_S3	691_A2_2_S3	1, 2, 5, 6, 7, 8	691A2_2_S3
701	103813	*O. nivara*	701_B2_S2**	701B2_3_S3	701_B2_3_S3	1, 2, 3, 5, 6, 7, 9	701B2_3_S3
704	103818	*O. rufipogon*	704_B1_S2	704B1_1_S3	704_B1_1_S3	1, 2, 5, 6, 7, 8, 9	704B1_1_S3
707	103835	*O. nivara*	707_B2_S2	707B2_2_S3	707_B2_2_S3	1, 2, 3, 5, 6, 7, 9	707B2_2_S3
708	103836	*O. nivara*	708_A1_S2	_	708_A1_2_S3	1, 2, 5, 6, 7, 8	_
711	103841	*O. nivara*	711_A1_S2	_	_	1, 2, 5, 6, 7, 8	_
715	104497	*O. rufipogon/O. nivara*	715_B2_S2	715B2_1_S3	715_B2_1_S3	1, 2, 3, 5, 6, 7, 9	715B2_1_S3
716	104647	*O. rufipogon*	716_B2_S2	716B2_2_S3	716_B2_2_S3	1, 2, 5, 6, 7, 8, 9	716B2_2_S3
717	104650	*O. nivara*	717_B1_S2	717B1_2_S3	717_B1_2_S3	1, 2, 5, 6, 7, 8, 9	717B1_2_S3
719	104687	*O. nivara*	719_A1_S2**	719A1_3_S3	_	1, 2, 3, 5, 6, 7, 9	_
720	104703	*O. nivara*	720_A3_S2	720A3_2_S3	720_A3_2_S3	1, 2, 3, 5, 6, 7, 9	720A3_2_S3
721	104705	*O. nivara*	721_C1_S2	721C1_2_S3	721_C1_2_S3	1, 2, 5, 6, 8, 9	721C1_2_S3
722	104962	*O. rufipogon/O. nivara*	722_A1_S2	722A1_1_S3	722_A1_1_S3	1, 2, 5, 6, 7, 8, 9	722A1_1_S3
723	104969	*O. rufipogon/O. nivara*	723_B2_S2	723B2_1_S3	723_B2_1_S3	1, 2, 5, 6, 7, 8, 9	723B2_1_S3
736	105494	*O. rufipogon*	736_B2_S2	736B2_1_S3	736_B2_1_S3	1, 2, 3, 5, 6, 9	736B2_1_S3
738	105601	*O. rufipogon/O. nivara*	738_B2_S2	738B2_2_S3	738_B2_2_S3	1, 2, 5, 6, 7, 8, 9	738B2_2_S3
743	105705	*O. nivara*	743_C1_S2	743C1_1_S3	743_C1_1_S3	1, 2, 5, 6, 7, 8, 9	743C1_1_S3
746	105740	*O. nivara*	746_C2_S2	746C2_3_S3	746_C2_3_S3	1, 2, 5, 6, 7, 8, 9	746C2_3_S3
751	105895	*O. nivara*	751_C3_S2**	751C3_2_S3	751_C3_2_S3	1, 2, 3, 5, 6, 7, 9	751C3_2_S3
757	106148	*O. nivara*	757_A1_S2	757A1_2_S3	757_A2_S2	2, 3, 5, 6, 7, 9	757A1_2_S3
759	106336	*O. rufipogon*	759_A1_S2	_	759_A1_3_S3	1, 2, 5, 6, 7, 8	_
760	106345	*O. nivara*	760_A2_S2	760A2_3_S3	_	1, 2, 5, 6, 7, 8	_

NSF-TV ID	IRGC #	Species	Genotyped on HDRA*	3D RSA (seed ID)	Al tol screen (seed ID)	Phenotyped in morphological screens***	Seedling vigor screen
762	106396	*O. nivara*	762_C1_S2	762C1_3_S3	762_C1_3_S3	1, 2, 5, 6, 7, 8, 9	762C1_3_S3

* Unless otherwise noted, all HDA-genotyped accessions are from a separate growout of materials for genotyping only by CWT, 2011

**These individuals are from growouts of materials by JJ in either 2009 or 2010

*** Accessions from various generations (Sx) were grown out, phenotyped, and bagged for self seed (DBNRRC and CU only) according to the growout location codes as follows: 1 = CU 2007-08 (CU S0); 2 = CU 2008-09 (CU S1); 3 = CU 2009-10 (CU S2); 4 = DBNRRC 2007 (CU S1); 5 = DBNRRC 2008 (DB S2); 6 = DBNRRC 2009 (DB S3)' 7 = IRRI 2008 (IRGC S0); 8 = CU 2010-11 (CU S2); 9 = China 2011 (CU S3) :

Supplementary Table 2.2 Trait ontology and phenotyping methodology for wild panel developmental and morphological traits.

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
ANTCO	030=yellow 050=brown	ANTHER COLOR	REPRODUCTIVE
APCO	IRRI:10=white; 20=straw; 52=brown; 60=green;70=red;71=red apex; 80=purple; 87=purple apex; 100=black; CU: 0=gr/white; 1=red; 2 = green	APICULUS COLOR	REPRODUCTIVE
APCO-binary	1=gr/white; 2=red/purple		REPRODUCTIVE
AUCO	0=absent; 011=whitish; 062=yellowish green; 080=purple; 081=light purple; 084=purple lines	AURICLE COLOR	VEGETATIVE
AUPIB	1=absent; 2=present	AURICLE PUBESCENCE	VEGETATIVE
AWCO	0=absent (awnless) 011=whitish 020=straw 040=gold 052=brown 061=light green 070=red 080=purple 100=black	AWN COLOR	REPRODUCTIVE
BLSCO	060=green; 080=purple; 081=light purple; 084=green with purple lines	BASAL LEAFSHEATH COLOR	VEGETATIVE
COLLCO	0=absent; 060=green; 061=light green; 080=purple; 084=purple lines	COLLAR COLOR	VEGETATIVE
CUKNAB	0 = no elbow/absence; 1 = elbow/presence; 2 = slight elbow	CULM KNEEING ABILITY	REPRODUCTIVE
CUNO		NO. OF CULMS PER PLANT	REPRODUCTIVE
FLAGATT	1=erect; 3=semi-erect (intermediate); 5=horizontal; 7=descending	FLAGLEAF ATTITUDE	REPRODUCTIVE
CUHABIT_VEG (In IRRI 2008, origi GROWTH)	1=erect; 3=semi-erect; 5=decumbent; 7=prostrate	GROWTH HABIT	VEGETATIVE
INCO_ANTHO	0=absent; 080=purple; 084=purple lines	CULM INTERNODE ANTHOCYANIN COLOR	REPRODUCTIVE
INCO_UNDER	0=absent; 041=light gold; 060=green	CULM UNDERLYING INTERNODE COLOR	REPRODUCTIVE
LEAF	1=very early 3=early 5=intermediate 7=late 9=very late	LEAF SENESCENCE	HARVEST
LIFE	1=annual 2=perennial 3=intermediate	LIFE CYCLE	HARVEST

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
LIGCO	0=absent; 062=yellowish green; 080=purple; 081=light purple; 084=purple lines	LIGULE COLOR	REPRODUCTIVE
LIGMARSH	1=entire; 2=scalloped or toothed	LIGULE MARGIN SHAPE	REPRODUCTIVE
LIGPUB	1=glabrous; 2=partially hirsute; 3=mostly or generally hirsute	LIGULE PUBESCENCE	REPRODUCTIVE
LIGSH	0=absent; 1=fringe of hairs; 2=truncate; 3=obtuse or rounded; 4=emarginate; 5=acute; 6=acuminate; 7=2-cleft	LIGULE SHAPE	REPRODUCTIVE
LMP	1=glabrous; 2=hairy or ciliated	LEAF MARGIN PUBESCENCE	VEGETATIVE
LPCO	010=white; 012=green-striped white; 042=gold and gold furrows; 052=brown (tawny); 053=brown spots on green; 054=brown furrows on green; 056=blackish brown; 060=green; 062=yellowish green; 080=purple; 082=reddish to light purple; 083=purple shade; 090=purp	LEMMA AND PALEA COLOR	REPRODUCTIVE
MARHAIR	0=absent; 1=present	LIGULE MARGIN HAIRINESS	REPRODUCTIVE
MAST	1=effectively absent (<25% sterile pollen); 2=intermediate; 3=male sterile (>95% sterile pollen)	MALE STERILITY	REPRODUCTIVE
NOCO_ANTHO	0=absent; 080=purple; 081=light purple; 084=purple lines	CULM NODE ANTHOCYANIN COLOR	REPRODUCTIVE
NOCO_UNDER	0=absent 041=light gold; 060=green	CULM UNDERLYING INTERNODE COLOR	REPRODUCTIVE
PAMA	1=upright 2=semi-upright 3=slightly drooping 4=strongly drooping	PANICLE ATTITUDE OF PRIMARY BRANCHES	REPRODUCTIVE
PAN2BR	0=absent 1=sparse (light) 2=dense (heavy) 3=clustering	PANICLE SECONDARY BRANCHING	REPRODUCTIVE
PANBR	1=whorled 2=alternate	PANICLE ARRANGEMENT OF PRIMARY BRANCHES	REPRODUCTIVE

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
PANCO	060=green 061 light green 062 yellowish green 063=dark green 080=purple 081=light purple 082=reddish purple	PANICLE BRANCHES AND AXIS COLOR	REPRODUCTIVE
PANEXS	1=enclosed 3=partly exserted 5=just exserted 7=moderately well exserted 9=well exserted	PANICLE EXSERTION	REPRODUCTIVE
PANTEXT	1=scabrous 2=smooth	TEXTURE OF PANICLE AXIS	REPRODUCTIVE
PANTYPE	1=compact 3=semi-compact 5=open 7=horizontal 9=drooping	PANICLE TYPE	REPRODUCTIVE
RHIZ	1 Vegetative crown2 Vegetative crown and stolon3 Vegetative crown and weak rhizomes4 Vegetative crown, weak stolon and weak rhizomes5 Strong rhizomes and no tubers6 Strong rhizomes with tubers	RHIZOME AND STOLON FORMATION	HARVEST
STGCO	010=white 030=yellow 061=light green 080=purple 081=light purple	STIGMA COLOR	REPRODUCTIVE
STLCO	011=whitish 020=straw 060=green 061=light green 062=yellowish green 080=purple 082=reddish purple	STERILE LEMA COLOR	REPRODUCTIVE
STLSH	0=absent 1=linear (long and slender) 2=subulate or setaceous 3=triangular	STERILE LEMMA SHAPE	REPRODUCTIVE
AUCO	0=absent; 11=whitish; 21=green; 62=yellowish green; 80=purple; 81=light purple; 84=purple lines	AURICLE COLOR	VEGETATIVE
CUKNAB	IRRI original: 0 = no elbow/absence; 1 = elbow/presence	CULM KNEEING ABILITY	REPRODUCTIVE
LIFE	1=annual; 1.5=intermediate; 2=perennial	LIFE CYCLE	HARVEST
STOL_BINARY	0=No stolons, vegetative crown only 1=vegetative crown and stolon	RHIZOME AND STOLON FORMATION	HARVEST
2LLT	Measurement (cm)	2ND LEAF LENGTH	REPRODUCTIVE
2LWD	Measurement (cm)	2ND LEAF WIDTH	REPRODUCTIVE
ANTLT	Measurement (cm)	ANTHER LENGTH	REPRODUCTIVE

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
AWNLT	Measurement (cm)	AWN LENGTH	REPRODUCTIVE
AWNPLU	DB and CUT Terminal awns only0 Absent1 short & partly awned5 short & fully awned7 long & party awned9 long & fully awned		REPRODUCTIVE
AWNPR (IRRI)	IRRI - Awn presence 0=absent; 1=partly awned; 2=fully awned	AWN PRESENCE	REPRODUCTIVE
AWNWD	Measurement (cm)	AWN WIDTH	REPRODUCTIVE
CUDI	Measurement (cm)	CULM DIAMETER	REPRODUCTIVE
CUHABIT_REPRO (IRRI 2008 -CUHABIT)	1=erect (<15 deg); 3=semi-erect (intermediate, ~20 deg); 5=open (~40 deg); 7=spreading (>60-80 deg); 9=procumbent	CULM HABIT	REPRODUCTIVE
CULT	Measurement (cm)		
DIST	Measurement (cm)	DISTANCE OF NEAREST SPIKELET TO PANICLE BASE	REPRODUCTIVE
FERT		PANICLE FERTILITY	POST HARVEST
FILLED	Count (total number of filled spikelets/panicle)		
FLFLG	Measurement (cm)		
FLFLWD	Measurement (cm)		
FLLT	Measurement (cm)	FLAG LEAG LENGTH	REPRODUCTIVE
FLWD	Measurement (cm)	FLAG LEAF WIDTH	REPRODUCTIVE
HULCL - limited	2 = white or straw or 0103 = gold4 = tawny/russet =052 IRRI5 = furrowed = 042 IRRI6 = green=053, 054, 060, 0628 = purple, brown, black, 056, 080, 082, 083, 090		

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
HULCL (IRRI)	IRRI 010=white; 012=green-striped white; 042=gold and gold furrows; 052=brown (tawny); 053=brown spots on green; 054=brown furrows on green; 056=blackish brown; 060=green; 062=yellowish green; 080=purple; 082=reddish to light purple; 083=purple shade; 090=purp CU/DB: 1 white ; 2 straw =010 IRRI; 3 gold; 4 tawny/russet =052 IRRI5 furrowed = 042 IRRI; 6 spotted, pibald; 7 purple; 8 black		
L to Pan node - S	Measurement (cm)		
L to Pan node - T	Measurement (cm)		
LBAC	0=absent; 1=present	LEAF BLADE ANTHOCYANIN COLOR	VEGETATIVE
LBDAC	080=even; 085=on margins only; 086=on tips only; 089=in blotches; (0=absent added YN22apr09)	LEAF BLADE DISTRIBUTION OF ANTHOCYANIN COLOR	VEGETATIVE
LBIGC	0=absent; 060=medium (green); 061=light; 063=dark	LEAF BLADE INTENSITY OF GREEN COLOR	VEGETATIVE
LFLPUBES (binary)	binary		
LFLPUBES (IRRI)	1=glabrous (no hairs; 2=hairy on upper surface; 3=-hairy on lower surface; 4=hairy on both sides	LEAF BLADE PUBESCENCE ON BLADE SURFACE	VEGETATIVE
LIGLT	Measurement (cm)	LIGULE LENGTH	REPRODUCTIVE
PANBASE	Count (# branches attached to base of panicle)	NO. OF BRANCHES ATTACHED TO PANICLE BASE	REPRODUCTIVE
PANLT	Measurement (cm)	PANICLE LENGTH	REPRODUCTIVE
PBRNB	Count (Panicle branch number		
PNLG	Measurement (cm)		
PNNB	Count (Panicle number per plant)	NO. OF PANICLES PER PLANT	REPRODUCTIVE

Character/Trait name	Class/scale	CHARACTER (IRRI)	STAGE TAKEN
PTHT culm	Measurement (cm) (not sensitive to tiller/stolon ID)		
SPKLT	Measurement (mm)	SPIKELET LENGTH	REPRODUCTIVE
SPKWD	Measurement (mm)	SPIKELET WIDTH	REPRODUCTIVE
STGLT	Measurement (mm)	STIGMA LENGTH	REPRODUCTIVE
STLLT	Measurement (mm)	STERILE LEMMA LENGTH	REPRODUCTIVE
STLWD	Measurement (mm)	STERILE LEMMA WIDTH	REPRODUCTIVE
STOLON +/- Maturity	Binary		
Stolon absence/presence at HD	Binary		HEADING DATE
STYLT	Measurement (mm)	STYLE LENGTH	REPRODUCTIVE
SUM_STGSTY	Sum of stigma and style lengths		REPRODUCTIVE
TOTAL	Sum of FILLED and UNFILLED counts for a given panicle	NUMBER OF SPIKELETS PER PANICLE	POST HARVEST
UNFILLED	Count (Total number of filled spikelets per panicle)	Total number of filled spikelets per panicle	POST HARVEST

REFERENCES

(UN) U. N., 2011 World Population Prospects: The 2010 Revision.

Alexandratos N., Bruinsma J., 2012 *World agriculture towards 2030/2050: the 2012 revision*. Rome.

Barbier P., 1989 Genetic variation and ecotypic diferentiation in the wild rice species *Oryza rufipogon*. II. Influence of the mating system and life-history traits on the genetic structure of populations. Japanese J. Genet. **64**: 273–285.

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

Cha K.-W., Lee Y.-J., Koh H.-J., Lee B.-M., Nam Y.-W., Paek N.-C., 2002 Isolation, characterization, and mapping of the stay green mutant in rice. TAG Theor. Appl. Genet. **104**: 526–532.

Clark R. T., MacCurdy R. B., Jung J. K., Shaff J. E., McCouch S. R., Aneshansley D. J., Kochian L. V, 2011 Three-dimensional root phenotyping with a novel imaging and software platform. Plant Physiol. **156**: 455–65.

Clark R. T., Famoso A. N., Zhao K., Shaff J. E., Craft E. J., Bustamante C. D., McCouch S. R., Aneshansley D. J., Kochian L. V, 2013 High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. Plant. Cell Environ. **36**: 454–66.

Cui H., Peng B., Xing Z., Xu G., Yu B., Zhang Q., 2002 Molecular dissection of seedling-vigor and associated physiological traits in rice. Theor. Appl. Genet. **105**: 745–753.

Famoso A. N., Clark R. T., Shaff J. E., Craft E., McCouch S. R., Kochian L. V, 2010 Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance. Plant Physiol. **153**: 1678–1691.

Finkelstein R., Reeves W., Ariizumi T., Steber C., 2008 Molecular aspects of seed dormancy. Annu. Rev. Plant Biol. **59**: 387–415.

Gu X.-Y., Kianian S. F., Foley M. E., 2004 Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). Genetics **166**: 1503–16.

Gu X.-Y., Kianian S. F., Foley M. E., 2005a Seed dormancy imposed by covering tissues interrelates to shattering and seed morphological characteristics in weedy rice. Crop Sci. **45**: 948.

Gu X.-Y., Kianian S. F., Hareland G. A., Hoffer B. L., Foley M. E., 2005b Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). Theor. Appl. Genet. **110**: 1108–18.

Huang X., Kurata N., Wei X., Wang Z.-X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L., Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A., Toyoda A., Lu T., Feng Q., Qian Q., Li J., Han B., 2012 A map of rice genome variation reveals the origin of cultivated rice. Nature **490**: 497–501.

Kong Z., Li M., Yang W., Xu W., Xue Y., 2006 A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice. Plant Physiol. **141**: 1376–88.

Koornneef M., Bentsink L., Hilhorst H., 2002 Seed dormancy and germination. Curr. Opin. Plant Biol. **5**: 33–36.

Lin S. Y. S., Sasaki T., Yano M., 1998 Mapping quantitative trait loci controlling seed dormancy and heading date in rice, *Oryza sativa* L., using backcross inbred lines. Theor. Appl. Genet. **96**: 997–1003.

McCouch S. R., Sweeney M., Li J., Jiang H., Thomson M., Septiningsih E., Edwards J., Moncada P., Xiao J., Garris A., Tai T., Martinez C., Tohme J., Sugiono M., McClung A., Yuan L. P., Ahn S.-N., 2006 Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. Euphytica **154**: 317–339.

McCouch S. R., Wright M. H., Tung C.-W., Maron L. G., McNally K. L., Fitzgerald M., Singh N., DeClerck G., Agosto-Perez F., Korniliev P., Greenberg A. J., Naredo M. E. B., Mercado S. M. Q., Harrington S. E., Shi Y., Branchini D. A., Kuser-Falcão P. R., Leung H., Ebana K., Yano M., Eizenga G., McClung A., Mezey J., 2016 Open access resources for genome-wide association mapping in rice. Nat. Commun. **7**: 10532.

Morishima H., Sano Y., Oka H., 1984 Differentiation of Perennial and Annual Types Due to Habitat Conditions in the Wild Rice *Oryza* perennis. Plant Syst. Evol. **144**: 119–135.

Nguyen B. D., Brar D. S., Bui B. C., Nguyen T. V, Pham L. N., Nguyen H. T., 2003 Identification and mapping of the QTL for aluminum tolerance introgressed from the new source, *Oryza rufipogon* Griff., into indica rice (*Oryza sativa* L.). Theor. Appl. Genet. **106**: 583–93.

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

Pineros M. A., Magalhaes J. V., Alves V. M. C., Kochian L. V., 2002 The physiology and biophysics of an aluminum tolerance mechanism based on root citrate exudation in maize. Plant Physiol. **129**: 1194–1206.

Piñeros M. A., Shaff J. E., Manslank H. S., Alves V. M. C., Kochian L. V, 2005 Aluminum

resistance in maize cannot be solely explained by root organic acid exudation: A comparative physiological study. Plant Physiol. **137**: 231–41.

Redoña E. D., Mackill D. J., 1996 Mapping quantitative trait loci for seedling vigor in rice using RFLPs. Theor. Appl. Genet. **92**: 395–402.

Tan L., Li X., Liu F., Sun X., Li C., Zhu Z., Fu Y., Cai H., Wang X., Xie D., Sun C., 2008 Control of a key transition from prostrate to erect growth in rice domestication. Nat. Genet. **40**: 1360–4.

Tian F., Li D. J., Fu Q., Zhu Z. F., Fu Y. C., Wang X. K., Sun C. Q., 2006 Construction of introgression lines carrying wild rice (*Oryza rufipogon* Griff.) segments in cultivated rice (*Oryza sativa* L.) background and characterization of introgressed segments associated with yield-related traits. Theor. Appl. Genet. **112**: 570–80.

Toriyama K., Heong K. L., Hardy B., 2005 Rice is life: scientific perspectives for the 21st century. In: *Proceedings of the World Rice Research Conference*, International Rice Research Institute, Tsukuba, Japan, p. .

Wolt J., 1994 *Soil solution chemistry : applications to environmental science and agriculture*. Wiley, New York.

Xu M., Zhu L., Shou H., Wu P., 2005 A PIN1 family gene, OsPIN1, involved in auxin-dependent adventitious root emergence and tillering in rice. Plant Cell Physiol. **46**: 1674–81.

Yoshihara T., Iino M., 2007 Identification of the gravitropism-related rice gene LAZY1 and elucidation of LAZY1-dependent and -independent gravity signaling pathways. Plant Cell Physiol. **48**: 678–88.

Yu B., Lin Z., Li H., Li X., Li J., Wang Y., Zhang X., Zhu Z., Zhai W., Wang X., Xie D., Sun C., 2007 TAC1, a major quantitative trait locus controlling tiller angle in rice. Plant J. **52**: 891–8.

Zhang Q., Wang C., Zhao K., Zhao Y., 2001 The effectiveness of advanced rice lines with new resistance gene Xa23 to rice bacterial blight. In: *Research Notes: Genetics of disease and insect resistance*,

Zhang Z.-H., Yu S. S.-B., Yu T., Huang Z., Zhu Y. Y.-G., 2005 Mapping quantitative trait loci (QTLs) for seedling-vigor using recombinant inbred lines of rice (*Oryza sativa* L.). F. Crop. Res. **91**: 161–170.

Zhang X., Zhou S., Fu Y., Su Z., Wang X., Sun C., 2006 Identification of a Drought Tolerant Introgression Line Derived from Dongxiang Common Wild Rice (*O. rufipogon* Griff.). Plant Mol. Biol. **62**: 247–259.

Zhou L., Wang J.-K., Yi Q., Wang Y.-Z., Zhu Y.-G., Zhang Z.-H., 2007 Quantitative trait loci for seedling vigor in rice under field conditions. F. Crop. Res. **100**: 294–301.

Zhou Y. Y.-L., Uzokwe V. N. E. V., Zhang C. C.-H., Cheng L.-R. L., Wang L., Chen K., Gao X.-Q., Sun Y., Chen J.-J., Zhu L.-H., Zhang Q., Ali J., Xu J.-L., Li Z.-K., 2011 Improvement of bacterial blight resistance of hybrid rice in China using the Xa23 gene derived from wild rice (*Oryza rufipogon*). Crop Prot. **30**: 637–644.

CHAPTER 3 - GENETIC AND HORMONAL CONTROL OF ROOT SYSTEM ARCHITECTURE[1]

Root system architecture (RSA) – the spatial configuration of a root system – is an important developmental and agronomic trait, with implications for overall plant architecture, growth rate and yield, abiotic stress resistance, nutrient uptake, and developmental plasticity in response to environmental changes. Root architecture is modulated by intrinsic, hormone-mediated pathways, intersecting with pathways that perceive and respond to external, environmental signals.

The recent development of several non-invasive 2D and 3D root imaging systems has enhanced our ability to accurately observe and quantify architectural traits on complex whole-root systems. Coupled with the powerful marker based genotyping and sequencing platforms currently available, these root phenotyping technologies lend themselves to large-scale genome-wide association studies, and can speed the identification and characterization of the genes and pathways involved in root system development.This capability provides the foundation for examining the contribution of root architectural traits to the performance of crop varieties in diverse environments. This review focuses on our current understanding of the genes and

[1]Jung, J. K.H., and McCouch, S.R. (2013). Getting to the roots of it: Genetic and hormonal control of root architecture. *Frontiers in Plant Science* 4:86. All supplemental tables referred to in this chapter may be found online at: http://dx.doi.org/10.3389/fpls.2013.00186

pathways involved in determining RSA in response to both intrinsic and extrinsic (environmental) response pathways, and provides a brief overview of the latest root system phenotyping technologies and their potential impact on elucidating the genetic control of root development in plants.

Introduction

The exploration of root biology lags far behind above-ground vegetative and reproductive growth and development in plants. There is a vast array of studies on root biology, but the literature is dispersed, highly fragmented, and difficult to search because there are no comprehensive phenotypic databases for plants. Many studies of root genes have been classified based on discovery technique (i.e. mutant, QTL, transgenic analyses) or response variable (hormones, microbial populations, insects, nutrients, water levels), but they have not been joined into a systemic understanding of root genetics. Furthermore, comprehensive ontology terms pertaining to root biology have yet to be established, let alone adopted, and gene functional annotation linking phenotypic characteristics into mechanistic pathways and networks is incomplete. Recently, GWAS approaches both advance and demand better integration of genetic studies, annotations, and pathways into a more complete and searchable data network.

Effective GWAS require the efficient integration of genotyping, phenotyping, and informatics capabilities. The continued development of increasingly rapid, low-cost, high-throughput genotyping and sequencing technologies, such as second and third generation sequencing and high density SNP arrays, have made it straightforward for researchers to generate massive amounts of genotypic data on individuals and populations of interest. The speed, efficiency, and

cost of high-throughput precision phenotyping of those same populations lag far behind, requiring significant investments of money, time, and labor to generate the data needed for large-scale mapping studies. The selection of traits measured may be limited due to a lack of quantitative measurement resolution and/or accuracy, leading to the frequent description of traits in qualitative classes that combine multiple biological processes, as opposed to specific quantifiable traits that each measure a distinct biological step or the result of a particular process. Furthermore, existing database resources that seek to compile and integrate phenotypic and physiological data with genotypic data, such as the Database of Genotypes and Phenotypes (dbGaP) (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap) and PhenomicDB (Groth et al, 2010), are limited by low data submission and limited curation capacity. While these databases are emerging as useful resources for human and bacterial data, plant-related datasets, particularly in relation to root system biology, are still woefully underrepresented.

Lack of comprehensive phenotypic and informatics resources is currently one of the most limiting factors for leveraging the power of GWAS. Although much about gene function, expression, and pathway or network interaction remains to be discovered, the plant genetics community has accumulated phenotypic data from both field and controlled environments during the last half-century. If properly structured and organized, these data could be interrogated to assist with candidate gene identification and interpretation of GWAS output. The problem is that there is no efficient way to access, parse, and cross-reference these data and therefore, they remain fragmented, dispersed and incompletely indexed. Because the collection, curation, and biological application of phenomic data is much more complicated and multi-dimensional than

genotypic data, it has yet to be standardized and streamlined into automated processing modules. As a result, finding, integrating and interrogating the components of complex phenotypes, particularly those associated with plant RSA, requires the intervention of expert biologists who manually search through the literature to discover relevant QTLs, pathways and candidate genes. The annotation process is a complex, multi-step, iterative adventure for the scientist interested in defining relevant genes and networks for association or linkage mapping analyses.

This review was motivated by the need to identify *a priori* candidate genes involved in rice root system architecture, morphology, growth, and development related to a the interpretation of an association mapping study based on a rice diversity panel that had been genotyped with 700,000 SNPs and screened for 19 components of seedling 3D root system architecture (RSA) (unpublished data, McCouch and Kochian labs, Cornell University and USDA-ARS). We identified known genes involved in RSA, which encompasses a range of heterogenous traits involved in many different aspects of plant growth architecture, morphology and phenology. After narrowing the search space using GWAS, we integrated information about candidate genes that mapped to candidate regions of the genome using mutant analysis, orthologous gene identification, comparative mapping, trait similarity, pathway and network extension. This was aided by the use of ontology and synteny-related informatics to find genes underlying GWAS peaks and QTLs (Lawrence and Harper 2008; Vilella *et al.* 2009; Lamesch *et al.* 2012; Chen *et al.* 2012). This article provides a comprehensive review of the genetics underlying root growth, development and response to environmental stimuli. We provide tables of genes that have been associated experimentally , and in silico by sequence homology with root development in rice,

with positional information and gene ontology (GO) evidence codes to facilitate database population and curation (Supplementary Tables 3.2 and 3.3).

Defining root system architecture (RSA)

RSA is a complex trait and refers to the spatial configuration of the root system in terms of the precise geometric arrangement of all root axes as laid down in the rooting medium. Root architecture is comprised of a whole system set of descriptors, and as such is senior to and distinct from, though naturally dependent on, the secondary fields of root anatomy, morphology, topology, and distribution; however, individual root architecture components may draw on or overlap with these fields. To clarify, root anatomy refers to the internal cellular structure and arrangement of a root; root morphology, the surface features, including diameter, root hair and cap characteristics, and contorsion; root topology, the hierarchical description of the connection of root axes to one another; and root distribution, the presence and distribution of roots in a positional gradient or grid along a horizontal and/or vertical axis.

As proposed by Fitter, there are five main components of root architecture, each of which may be comprised of several specific traits or parameters (Fitter, 1991). These components are: 1) branch magnitude – the number of interior links (internode segments between two branching points or nodes) or exterior links (internode segments between a branching point and an endpoint, i.e. root apical meristem; 2) topology, the pattern of branch distribution, which is usually herringbone (alternate lateral branching off a parent root), dichotomous (opposite, bifurcating branches), or radial (whorls of branches around a parent root (Hochholdinger 2009; Lynch and Brown 2012); 3) link/internode lengths, the distance between branch points among

different root orders of an individual root, which may be averaged across a system; 4) root angles, specifically the azimuth (radial angle) of a lateral root's emergence around the circumference of a parent root, the branching angle or departure rate of a lateral root from a parent root, and the spreading angle of the entire system; and 5) link radius, the diameter of any given root (Fitter, 1991).

Pathways and networks influencing root architecture traits

As with any phenotypic manifestation, all of these simple root architecture components: branch number, branching pattern, length, orientation, angle, and diameter are developmentally controlled by complex interacting genetic pathways, which also modulate growth and developmental responses in response to the perception of environmental cues. Malamy and Ryan refers to these familiar factors—genetics, environment, and the interaction between the two—as belonging to either 'intrinsic pathways' or extrinsic 'environmental response pathways' (Malamy and Ryan, 2001).

Hormones, their receptors, signaling components, and transcription factors make up the main chemical and molecular components of the intrinsic pathways. Extrinsic response pathways involve similar networks of receptors for environmental stimuli and their downstream signal transduction and transcription factors. Many components of the environmental perception and response networks are shared with or interregulated by intrinsic response pathways, and are also mediated by hormonal regulation in order to effect a growth response to external signals (See Table 3.1 for a review of the major hormones and their role in modulating root architectural traits; Table 3.2 for a review of the major extrinsic factors, their effects on root growth and

development, and the major genes and hormones involved, and Supplementary Table 3.1 for the key genes involved in root growth and development covered in this review). Recent studies have also identified micro-interfering RNAs (miRNAs) and small-interfering RNAs (siRNAs) which affect RSA by the post-transcriptional regulation of components involved in root growth and environmental perception and response and are themselves transcriptionally interregulated by feedback loops within the same intrinsic and extrinsic pathways (see reviews in Meng et al., 2010; Khan et al., 2011).

To date, the vast majority of research elucidating the genes and pathways involved in root architecture development has been done with the simple, dicot taproot system of *Arabidopsis thaliana* (Scheres et al., 1996; Ueda et al., 2005; Péret et al., 2009a). This has allowed for the gradual application of this knowledge in discerning conserved developmental pathways shared with monocot crown root systems, primarily studied in cereal crops such as rice (*Oryza sativa* L.) and maize (*Zea mays* L.)

Table 3.1 Hormones and their involvement in root growth and development

Hormone	Chemical compounds	Function	Hormone source	Species	Literature
Auxin	IAA	Promotes lateral root initiation by specifying lateral root founder cells	Endogenous, root tip	At	Casimiro et al, 2001; De Smet et al, 2007, Dubrovsky et al, 2008
	IAA	Promotes lateral root emergence	Endogenous, shoot	At	Bhalarao et al, 2002
	NAA	Increases lateral root primordia initiation and outgrowth	Exogenous	Os, Nt	Sreevidya et al, 2010; Campanoni et al, 2005
	2,4-D	Increases lateral root primordia initiation through cell division (but does not promote cell elongation and root outgrowth)	Exogenous	Os, Nt	Sreevidya et al, 2010; Campanoni et al, 2005
	IAA	Promotes primary root elongation by facilitating the response of root cells to GA3	Exogenous	At	Fu et al., 2003
Cytokinins	Kinetin, BAP	Inhibits lateral root primordia formation by perturbing PIN gene expression and disrupting formation of a RAM auxin gradient controlling cell division to maintain the QC and neighboring initials	Increased endogenous	At	Laplaze et al., 2007; Dello Ioio et al., 2008; reviewed in Péret et al., 2009b
	Kinetin, trans-zeatin	Stimulates lateral root elongation	Exogenous	Os	Rani Debi, et al., 2005; Laplaze et al., 2007; Dello Ioio et al., 2008; reviewed in Bishopp, et al., 2009
	Kinetin, trans-zeatin	Stimulates crown root primordia formation	Exogenous	Os	Rani Debi, et al., 2005; Hirose et al., 2007; Zhao et al., 2009

	Zeatins, other endogenous cytokinins	Inhibits primary root elongation by reducing cell division in RAM, thus regulating RAM size	Increased endogenous	At	Ruzicka et al., 2009; Kuderova et al., 2008
Gibberellins	GA3	Interacts with ethylene to promote crown root primordia outgrowth and elongation	Exogenous	Os	Steffens et al., 2006
	GA3	Promotes primary root elongation in the presence of auxin by repressing growth-repressing DELLA proteins	Decreased endogenous and increased exogenous	At	Fu et al., 2003
	GA3	Inhibits lateral root primordia initiation	Exogenous	Pt	Gou et al., 2010
Ethylene	Ethylene	Promotes crown root formation at submerged nodes	Internode	Os	Lorbiecke et al., 1999
	Ethylene	Promotes crown root emergence at submerged nodes through induction of epidermal cell death over sites of lateral root primordia formation	Internode	Os	Mergemann and Sauter, 2000
	Ethylene				
Jasmonates	MeJA	Promotes lateral root formation through interaction with auxin pathway	Increased endogenous	At, Gm	Xue et al., 2007; Sun et al., 2009
	MeJA	Inhibits primary root growth	Increased endogenous	Gm	Xue et al., 2007

Abscisic Acid	ABA	Induces lateral root primordia formation under non-stress conditions by modulating the auxin response	Endogenous	At	Brady et al, 2003
	ABA	Maintains primary root elongation under drought stress	Endogenous	Zm	Saab et al., 1990
	ABA	Inhibits lateral root outgrowth prior to lateral root meristem formation under non-stress conditions	Exogenous	At	De Smet et al., 2006
Brassinosteroids	BL	May induce lateral root initiation in the presence of auxin, through modulating auxin signalling	Exogenous	At	Bao et al., 2004
	BL	Induces primary root elongation in the presence of exogenous auxin (IAA) by affecting ethylene biosynthesis and the gravitropic response	Exogenous	Zm, At	Yun et al., 2009; Kim et al., 2007; Chang et al., 2004
	HBR	Induces primary and crown root elongation possibly through modulating auxin signalling	Exogenous	Hv	Kartal et al., 2009
Strigolactone	GR24 (synthetic strigolactone analog)	May either inhibit primary root elongation in low concentrations, or stimulate primary root growth in high concentrations, in the presence of auxin, by putative regulation of auxin efflux carriers	Exogenous	At	Koltai et al., 2010; Kapulnik et al, 2011; Ruyter-Spira, 2011
	GR24 (synthetic strigolactone analog)	Induces primary root curving in high concentrations, in the presence of no-low auxin by inducing asymmetric cell elongation	Exogenous	At	Koltai et al., 2010
	SLs	Promote crown root elongation by inducing meristematic cell division, possibly through the modulation of local auxin concentrations that regulate meristem cell number	Endogenous	Os	Arite et al., 2012

SLs	Putatively modulates auxin sensitivity by downregulating auxin efflux carrier expression to inhibit lateral root formation under low auxin levels by reducing auxin accumulation in roots, or inducing lateral root formation under high auxin concentrations by allowing optimal auxin levels to be met	Endogenous	At	Kapulnik et al, 2011; Ruyter-Spira, 2011

Abbreviations: 2,4-D, 2,4-Dichlorophenoxyacetic acid; ABA, Abscisic acid; At, Arabidopsis thaliana; BAP, 6-benzylaminopurine; BL, Brassinolide; GA3, Gibberellic acid; Gm, Glycine max; HBL, Homobrassinolide; Hv, Hordeum vulgare; IAA, Indole-3-acetic acid; MeJA, Methyl jasmonate; NAA, 1-Naphthaleneacetic acid; Nt, Nicotiana tabacum; Os, *Oryza* sativa; Pt, Populus tremulus; QC, quiescent center; RAM, root apical meristem; SLs, endogenous strigolactones; Zm, Zea mays

Table 3.2 Effects of extrinsic factors in modulating root system architecture					
Factor	**Condition**	**General effect on root growth**	**Genes with known involvement**	**Interactions with hormone pathways**	**References**
Environmental stimuli					
Gravity	Normal	Growth toward the gravity vector	*ARG1* and *2*, *PIN3* (At)	Auxin	Boonsirichai et al., 2003; Harrison and Masson, 2008a; Harrison and Masson, 2008b
Light (direct root exposure)	Presence	Negative growth to blue light; positive to red/far red light	*PHOT1*, *NPH1*, *PhyA*, *PhyB* (At)	Auxin, JA	Galen et al., 2007; Huala et al., 1997; Christie et al., 1998; Kiss et al., 2003; Kurata and Yamamoto, 1997; Costigan et al., 2011; Correll et al., 2003
Water/oxygen	Root system submergence/hypoxia	CR primordia development and outgrowth (deepwater rice)	*SUB1* (Os)	GA, ethylene	Xu et al., 2006; Fukao et al., 2011
	Drought	Mixed. General decreased LR and PR growth and LR emergence, but ABA has been shown to stimulate PR elongation and LR emergence in response to drought	*LACS2* (At); *SUB1* (Os)	ABA	Sharp et al., 1988; Deak and Malamy, 2005; De Smet et al., 2006; Macgregor et al., 2008; Wiegers et al., 2009
Soil nutrients					
Nitrogen	High nitrate availability	Inhibition of LR outgrowth, development and elongation	*AtNRT1.1*, *ANR1*, *AtOCT1*(At)	Auxin, ABA	Zhang and Forde, 1998; Liu et al., 1999; Zhang et al., 1999; Guo et al., 2002; Munos et al., 2004; Lelandais-Briere et al., 2007; Krouk et al., 2010
	Low nitrate availability	Localized stimilation of LR growth, branching in high inorganic N soil patches	*AtNRT1.1*, *AtNRT2.1*, *ANR1*, *AtOCT1*(At)	Auxin	Zhang and Forde, 1998; Zhang et al., 1999; Malamy and Ryan, 2001; Lelandais-Briere et al., 2007; Krouk et al., 2010

Phosphorus	High phosphate availability	PR growth promoted, LR growth prohibited	*PDR2*, *LPR1*, *WRKY75* (At)	Auxin	Linkohr et a;., 2002; Ticconi et al., 2004; Shane and Lambers, 2005; Schulze et al., 2006; Reymond et al., 2006; Devaiah and Raghothama, 2007; Devaiah et al., 2007; Perez-Torres et al., 2008; Ticconi et al., 2009
Phosphorus (con't)	Low phosphate availability	Root foraging: increased LR initiation, outgrowth, forming a shallow, highly branched system	*PDR2*, *LPR1*, *PHR1*, *AtSIZ1*, *PHO2* (At), *OsPTF1*(Os), *PHI2* (Nt)	Auxin, CK, ethylene, GA, SLs	Schmidt and Schikora, 2001; Williamson et al., 2001; Franco-Zorrilla et al., 2002; Lopez-Bucio et al., 2002; Sano and Nagata, 2002; Ma et al., 2003; Shimizu et al., 2004; Ticconi et al., 2004; Miura et al., 2005; Shane and Lambers, 2005; Yi et al., 2005; Bari et al., 2006; Schulze et al., 2006; Reymond et al., 2006; Devaiah and Raghothama, 2007; Devaiah et al., 2007; Jiang et al., 2007; Perez-Torres et al., 2008; Ticconi et al., 2009
Sulfur	High sulfate availability/ sufficiency	Not highly studied	*SULTR1:2*, *SLIM1* (At)	Auxin, JA, CK	Ohkama et al., 2002; Hirai et al., 2003; Maruyama-Nakashita et al., 2003; Nikiforova et al., 2003; Buchner et al., 2004; Hoefgen and Nikiforova, 2008; Lewandowska and Sirko, 2008; Takahashi, 2010
	Low sulfate availability	Mixed. Short-term sulfur limitation proposed to stimulate LR growth with longer-term deficiency causing overall decreased growth	*SULTR1:1*, *SULTR1:2*, *SLIM*, *OAS*, *NIT3*, *BIG*, *IAAs* (At)	Auxin, JA, CK	Leustek et al., 2000; Saito, 2000; Takahashi et al., 2000; Kutz et al., 2002; Ohkama et al., 2002; Yoshimoto et al., 2002; Maruyama-Nakashita et al., 2003; Nikiforova et al., 2003; Buchner et al., 2004; Maruyama-Nakashita et al., 2004; Hoefgen and Nikiforova, 2008; Lewandowska and Sirko, 2008; Bouranis et al., 2008
Phytotoxins					

Aluminum	High Al^{3+}	Inhibition of LR intiation and outgrowth, swollen, malformed root tips	*ETR1*, *EIN2*, *AtACSs*, *AtACOs*, *AtPIN*, *AUX1*, *PME*, *AtCHIA*, *CALS* (At); *EXPA10*, *STAR1* and *2*, *ART1* (Os)	Auxin, ethylene	Foy, 1984; Delhaize et al., 1993; Kochian, 1995; Alonso et al., 1999; Matsumoto, 2000; Lee and Kende, 2002; Tsuchisaka and Theologis, 2004; Yokoyama and Nishitani, 2004; Eticha et al., 2005; O'Malley et al., 2005; Jones et al., 2006; Sivaguru et al., 2006; Sun et al., 2010
Sodium chloride	High salinity	Mixed. General decrease in root growth due to slower epidermal cell division and elongation	*HKTs*, *GLRs*, *NSCCs*, *CNGCs*, *SOS1-3*, *NX1* (At)	Auxin, ABA, CK, ethylene, GA	Kuiper et al., 1990; Zidan et al., 1990; Liu and Zhu, 1998; Apse et al., 1999; Gaxiola et al., 1999; Leng et al., 2002; Quintero et al., 2002; Tester and Davenport, 2003; He et al., 2005; Mahajan and Tuteja, 2005; Khadri et al., 2006; Cao et al., 2008; Bano, 2010
Symbioses					
Root nodulation	Pre-symbiosis Nod factor-induced	None known.			
	High colonization	Nodule formation, putative suppresion of LR emergence	*Nodulins*, *LHK1* (Lj), *MtCRE*, *ARR*, *NSP1 and 2*, *NIN*, *ENOD11*, *ERFs* (Mt)	Auxin, ABA, BRs, CK, ethylene, GA, SA	Nutman, 1948; Nap and Bisseling, 1990; Verma et al., 1992; Catoira, 2000; Journet et al., 2001; Borisov et al., 2003; Charron et al., 2004; Lohar et al., 2004; Kalo et al., 2005; Smit et al., 2005; Gonzalez-Rizzo et al., 2006; Marsh et al., 2007; Middleton et al., 2007; Frugier et al., 2008; Vernie et al., 2008; Ding and Oldroyd, 2009; Markmann and Parniske, 2009; Ferguson et al., 2010
Arbuscular mycorrhizal	Pre-symbiosis Myc-factor-induced	LR elongation	*DMI1* and *2*, *MtENOD11*(Mt), *OsPOLLUX2*, *OsCCAMK2*, *Os CYCLOPS1*	Auxin, ABA, CK,SLs	Endre et al., 2002; Stracke et al., 2002; Kosuta et al., 2003; Olah et al., 2005; Hogg et al., 2006; Gutjahr et al., 2009

			(Os)		
	High AM colonization	Variable increases in root mass, thickness, length, and LR number dependant on host species	*LRT1*(Zm)	Auxin, ABA, CK, ethylene	Hetrick et al., 1988; Berta et al., 1990; Dixon, 1990; Hetrick, 1991; Berta et al., 1995; Barker and Tagu, 2000; Berta et al., 2002; Paszkowski and Boller, 2002; Vierheilig et al., 2002; Fitze et al., 2005; Olah et al., 2005; Ludwig-Muller and Guther, 2007; Parniske, 2008; Gutjahr et al., 2009
Abbreviations: At, Arabidopsis thaliana; BRs, brassinosteroids; CR/CRP, Crown root/Crown root primordia; Lj, Lotus joponicus; LR/LRP, Lateral root/Lateral root primordia; Nt, Nicotiana tabacum; Os, *Oryza* sativa; PR, primary root; Ps, Pisum sativum; QC, quiescent center; RAM, root apical meristem; SA, salicylic acid; SLs, strigolactones; Zm, Zea mays					

The importance of root architecture

The 3D configuration of a root system is important mechanically, providing physical anchorage of the plant in soil, and physiologically, in nutrient and water sensing and uptake, and in response to soil biota. The rate of root system growth and its vertical and horizontal spread can affect seedling vigor, neighbor competition, and exploitation of different limiting resources, such as phosphorus, nitrogen, and water, through root growth or support of symbioses, and can be highly specific to environmental conditions--a root architecture which may favor the growth of a plant under low water conditions, may impede its growth in flooded soil. The specific growth and development characteristics of a plant's root system also confers some degree of developmental plasticity to the organism in dealing with nutrient and water availability, seasonal and climate changes, beneficial or disease causing organisms or toxic compounds in soil. Together, these qualities of anchorage, soil nutrient exploitation, and developmental plasticity as determined by root architecture can have far-reaching effects on maximal yield, especially under stress, and yield stability, and a greater understanding of the genes and pathways involved in root architectural development may be translated into the breeding of improved crop varieties.

Intrinsic Pathways – Genetic and hormonal regulation of root architecture

Primary root initiation, development, and elongation

The primary root, derived from the radicle and laid down during embryogenesis, grows to form the foundation of the dicotyledonous taproot system, and is the first root of the fibrous, crown root-based root system of monocots. Establishment of the root apical meristem (RAM) of the primary root involves cell identity differentiation and the formation and maintenance of a

quiescent center (QC) and stem cell population. In Arabidopsis, auxin signaling and its antagonistic feedback by cytokinins (CKs) have been implicated in the development of a root stem cell niche (Muller and Sheen, 2008; Kartal et al., 2009; Moubayidin et al., 2009; Pernisova et al., 2009; Ruzicka et al., 2009). The secondary regulation of auxin signaling by gibberellins, and brassinosteroids has also been implied (Sabatini et al., 1999; Frigerio et al., 2006). Polar auxin transport by the AUXIN1/LIKE AUXIN (AUX1/LAX) family of auxin influx transporters and the PIN-FORMED 3 (PIN3) and PIN7 auxin efflux transporters lead to the creation and maintenance of an auxin concentration gradient with a root tip maximum (Bennett et al., 1996; Parry et al., 2001; Kramer, 2004; Blilou et al., 2005; Carraro et al., 2006; Swarup et al., 2008; Liu et al., 2009; Wang et al., 2009; see reviews in Petrásek and Friml, 2009; Overvoorde et al., 2010) (Figure.3.1). Several multidrug resistant/P-glycoprotein (MDR-PGP) subfamily members of the ATP-binding cassette subfamily B (ABCB) are also key auxin influx and efflux membrane transporters (Noh et al., 2001; Luschnig et al, 2002; Noh et al., 2003).

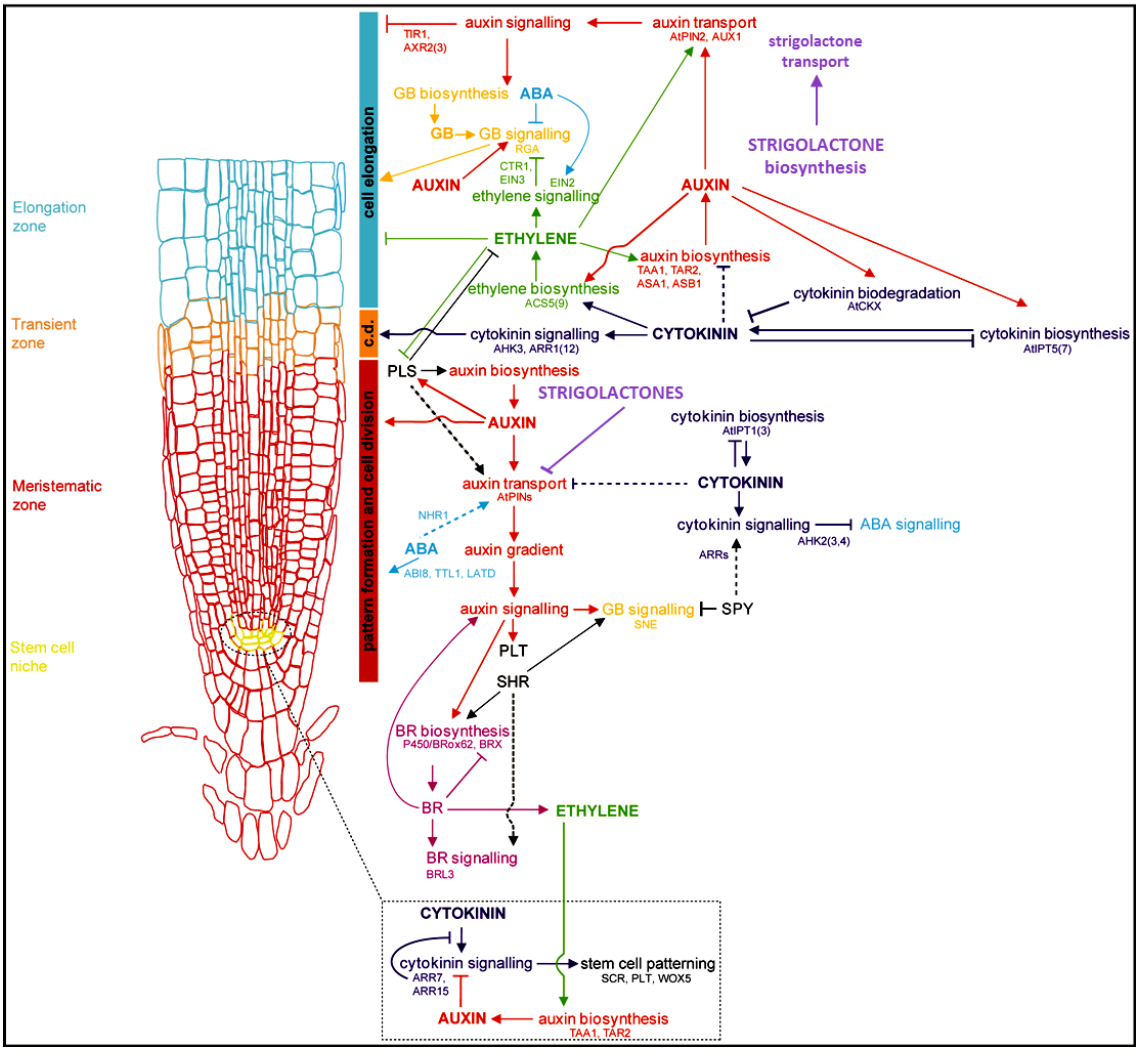


Figure 3.1. Genetic and hormonal control of primary root development in Arabidopsis. Model of the current understanding of hormone interaction and genetic regulation of primary root and general root apical meristem growth and development in Arabidopsis. Important genes involved in integrating signals from different hormone pathways are shown in black; hormone networks are color-coded; dashed lines represent unresolved or indirect relations. The fundamental role of auxin-mediated signaling in controlling all major aspects of root growth, from cell division, differentiation, and elongation, can be visualized, as well as the antagonistic regulation of auxin by cytokinins, and secondary regulation by other hormones, including ABA, ethylene, GA, BRs, and strigolactones. c.d. is "cell differentiation," in reference to the transition zone where cell differentiation is initiated. (Modified from Benkova et al., 2009).

Strigolactones (SLs), a new class of plant hormones and rhizosphere signaling molecules have also been implicated in primary root development based on crosstalk with auxin signaling. In the presence of auxin, exogenous application of the synthetic SL analogue GR24 has been found to either inhibit Arabidopsis primary root elongation in low concentrations, or stimulate primary root growth in high concentrations by putative regulation of the auxin efflux carriers PIN1, PIN3, and PIN7 (Kapulnik et al, 2011; Ruyter-Spira et al, 2011). GR24 has also been found to induce primary root curving in high concentrations, in the presence of no or low auxin levels by inducing asymmetric cell elongation (Koltai et al, 2010). It should be noted, however, that due to the increased stability of GR24 in aqueous solution, as compared with natural SLs, the effects of this synthetic strigolactone on root growth may be misrepresented (Akiyama et al,. 2010).

The presence or absence of auxin transcriptionally regulates many genes involved in general root growth and development through the action of auxin/indole-acetic acid (Aux/IAA) and Auxin Response Factor (ARF) modules (De Smet *et al.* 2010; Goh *et al.* 2012). When not bound to Aux/IAA proteins, ARFs are free to recognize and bind to auxin-responsive elements (AREs) in the promoters of target genes, activating or repressing their transcription. In the absence of auxin or under low auxin concentrations, AUX/IAA proteins, negative regulators of auxin response genes (Abel 1994) bind with their ARFs, inactivating ARF activity. Under high auxin concentrations, AUX/IAA proteins are targeted for degradation by the SCF^{TIR} E3 ubiquitin ligase complex (Gray et al., 2001; Reed, 2001; Dharmasiri et al., 2005; Kepinski and Leyser, 2005; Badescu and Napier, 2006; Maraschin et al., 2009) (Figure 3.1).

Other layers of ARF regulation involve miRNAs. The miR160 family has been found to play a role in Arabidopsis primary and lateral root development through its regulation of the ARF TFs,

ARF10 and ARF16, which are functionally redundant but both required for root cap cell formation and development (Wang et al, 2005). Transgenic overexpression of miR160 in rice also induced severe root cap defects, suggesting the presence of a similar regulatory pathway in monocots, although the target(s) of miR160 in rice have not yet been determined (unpublished data as cited in Meng et al., 2010). Normal root cap formation in all roots is necessary for normal root system development and impinges on multiple downstream RSA components, specifically, root elongation, lateral root production, and root growth angle as dictated by the gravitopic response through root tip sensing (Wang et al., 2005; Band et al., 2012).

In Arabidopsis, a second set of TFs: SHORTROOT (SHR) and its target, SCARECROW (SCR), both GRAS TFs, are involved in the specification and localization of stem cells and the QC, as well as root radial patterning. They affect not only primary root initiation, but also root diameter, and the regulation of cell division and differentiation necessary for downstream lateral root development (Di Laurenzio et al., 1996; Helariutta et al., 2000; Sabatini et al., 2003; Paquette and Benfey, 2005; Lucas et al., 2011). SCR is also suggested to have a possible role in mediating a cross-response between gibberellic acid, brassinosteroid, and auxin signaling involved in stem cell maintenance (Muller and Sheen, 2008; Ruzicka et al., 2009; reviewed in Benkova and Hejatko, 2009). The maize SCR homolog, *ZmSCR*, was shown to be essential for the development of the maize radicle during the formation of the coleorhizae, the unique grass structure that sheathes and protects the primary root meristem during embryogenesis and germination (Tillich, 1977; reviewed in Hochholdinger and Zimmermann, 2008).

A third set of TFs, related to the second set, are the DELLA proteins, including the Arabidopsis GIBBERELLIN INSENSITIVE (GAI), REPRESSOR OF GA1 (RGA) and RGA-LIKE 1, RGA-

LIKE 2, and RGA-LIKE 3 (RGL1, RGL2, and RGL3), rice SLENDER RICE (SLR), and its barley homologue, SLENDER1 (SLN1), are negative regulators of GA-mediated root growth, and appear to be negatively regulated by auxin. The ubiquitination and destruction of these DELLA TFs in the presence of auxin and GA thus allow for root cell division and elongation (Dill and Sun, 2001; Ikeda et al., 2001; Chandler et al., 2002; Ikeda et al., 2002; Fleet and Sun, 2005; Perez-Perez, 2007) (Figure 3.1).

Lateral root growth – From primordia initiation to elongation

First order (or primary) lateral roots (LR) are roots that branch off of the taproot or adventitious roots in dicots, and the primary seminal root or crown roots in monocots. These first order laterals may be short and determinate, or they may develop higher orders of ramification (second, third, fourth-order, etc. laterals). Lateral roots account for the majority of the root mass in most plant root systems, and perform key functions in soil exploration, nutrient and water uptake, and symbiosis development. While lateral root production is generally developmental, it may also be adaptive, in response to environmental influences within the rhizosphere. Lateral roots are similar in anatomy, but usually smaller in diameter than their parent root, due to a reduced number of cortical cell layers and xylem and phloem poles (Coudert et al., 2010).

Lateral root growth may be organized into four stages with different implications for root system architecture: 1) LR initiation, 2) LR primordia formation, 3) LR meristem outgrowth and emergence from the parent root, and 4) LR elongation (Malamy and Benfey, 1997). The first three stages all affect the potential number and radial orientation of lateral roots. Development may be halted at any stage during this process which, prior to emergence would reduce the

number, position, and pattern of mature LRs; LR elongation affects LR branching angle, branch length, development rate, and whole system topology.

Lateral root initiation

The first stage in lateral root development takes place in the parent root pericycle in Arabidopsis, and the pericycle and endodermis layers in crop cereals like maize and rice (Casimiro et al., 2001; Fahn, 1990). This process is characterized by founder cell identity priming and fate fixation by auxin, cell cycle activation of the founder cells, and asymmetric cell division (Malamy and Benfey, 1997; De Smet et al., 2007; reviewed in Fukaki and Tasaka, 2009). The IAA28-ARFs module, the first of three known AUX/IAA-ARF modules regulating LR development is active in this LR initiation stage for LR founder cell specification (De Rybel et al., 2010). Cell cycle reactivation and control is fundamental to LR initiation and is partially induced by the accumulation of high auxin levels in quiescent xylem pole pericycle or endodermal cells (Casimiro et al., 2001; Beeckman et al., 2001; Malamy, 2005), and the priming of specific xylem pole or endodermal cells to become LR founder cells by 15-hour oscillations in the auxin level (De Smet et al., 2007). In Arabidopsis, this root-tip synthesized auxin gradient was found to promote asymmetric cell division of xylem pole pericycle founder cells (Casimiro et al., 2001; De Smet et al., 2007) by the auxin-induced up-regulation of cell cycle genes, including cyclins and cyclin dependent kinases (*CDK*s) (Soni 1995; Meijer and Murray 2000; Boniotti and Gutierrez 2001), and the synchronous down-regulation of CDK repressors, such as *KRP1* and *KRP2*, which inhibit the G1 to S transition phase in lateral root primordial (Himanen et al., 2002; reviewed in Fukaki et al., 2007) (Figure 3.2).

Further research has suggested that cyclic changes in auxin concentration are insufficient as the sole trigger of lateral root initiation, and that molecular clock-coordinated oscillating gene expression within the so-called 'oscillation zone,' a region encompassing the primary root basal meristem and elongation zone, is also necessary for the spatial and temporal definition of lateral root pre-branching sites. These pre-branching sites develop lateral root primordia, but may not always grow out into fully emerged lateral roots (Moreno-Risueno, et al, 2010). In Arabidopsis, two sets of 2084 and 1409 genes were found to oscillate either in phase or in antiphase, respectively, with specific waves of each phase being associated with increased expression of particular genes, mostly notably members of the ARF, NAC, MYB, and SOMBRERO TF families. T-DNA insertions in several of these genes also showed defects in LR pre-branching site initiation and reduced lateral root number (Moreno-Risueno, et al, 2010).

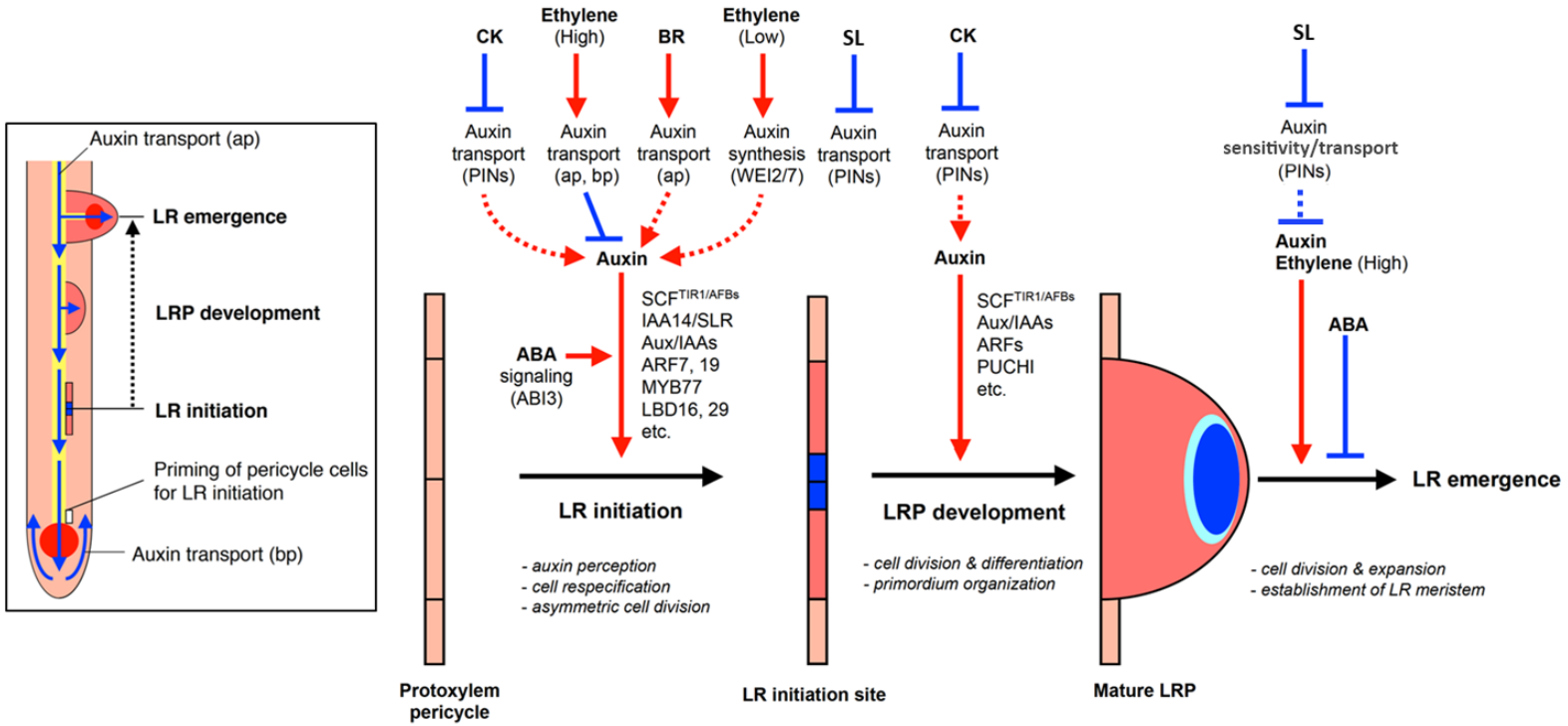


Figure 3.2. Hormonal and genetic control of lateral root formation in Arabidopsis. LR formation is a three-stage process consisting of LR initiation, LRP development, and LR emergence. LR initiation is positively regulated by auxin but negatively regulated by CK and high concentrations of ethylene (high concentrations of exogenous ACC). The polar auxin transport with a balance of influx and efflux in both acropetal and basipetal directions is necessary for LR initiation and setting up auxin gradient to organize LR primordium (LRP) (blue color in LR initiation site and primordium). CK inhibits auxin maxima by altering the expression of PINs, thereby inhibiting auxin gradient for LR initiation. High concentrations of ethylene or exogenous ACC, an ethylene precursor, inhibited LR initiation by enhancing acropetal (ap) and basipetal (bp) auxin transport. BR promotes LR initiation by increasing acropetal (ap) auxin transport. Low concentrations of ethylene (low concentrations of exogenous ACC) promote LR initiation by increasing Trp-dependent auxin synthesis mediated by WEI2 and WEI7. Normal ABA signaling mediated by ABI3 is necessary for proper auxin responsiveness for LR initiation. Auxin also promotes LR primordium development but CK inhibits LR primordium development and affects auxin maxima by altering the expression of PINs. ABA inhibits LR emergence whereas auxin and ethylene (via high concentrations of exogenous ACC) promotes LR emergence (Modified from Fukaki et al, 2009).

Lateral root primordia formation

The formation of lateral root primordia (LRP) is characterized by several rounds of anticlinal and periclinal cell division (Malamy and Benfey, 1997).As modeled in Arabidopsis, this process generates a patterned LRP similar to the primary root tip (DiDonato et al., 2004). Mutant and transgenic studies in Arabidopsis suggest that the formation of both the lateral root meristems (LRM) and the primary RM are driven by equivalent, if not the same, hormonal and genetic factors (Malamy and Benfey, 1997). Auxin is the primary signaling hormone regulating LRP development through the formation of an auxin gradient (Péret et al., 2009a). This gradient is modulated upstream by low levels of antagonistic CKs which would otherwise repress LRP formation via the disruption of auxin efflux PIN protein localization, which itself is partly responsible for creating the auxin gradient (Laplaze et al., 2007). Cytokinin specifically affects the rate of cell differentiation between the cell division and elongation/differentiation zones but does not affect the rate of cell division in the meristem (Dello Ioio et al., 2008). Strigolactones may also regulate LRP formation, possibly by altering auxin sensitivity through downregulating the expression of auxin efflux carriers such as PIN1, thus inhibiting lateral root formation under low auxin levels by reducing auxin accumulation in roots, or inducing lateral root formation under high auxin concentrations by allowing optimal auxin levels to be met (Ruyter-Spira et al., 2011).

The Arabidopsis GTP:GDP ANTIPORTER/ PROTEIN HOMODIMERIZATION (GNOM) protein also appears to play an essential role in regulating PIN protein trafficking for auxin gradient formation (Steinmann et al., 1999; Geldner et al., 2003; Laplaze et al., 2007). The accumulation of auxin in the central cells and later in the tip of the LRP signals the targeted degradation of AUX/IAA proteins, repressors of auxin-induced transcription. Furthermore, the

auxin gradient enables ARF7/NPH4 and ARF19 module-upregulated transcription of target genes for cell ID and pattern formation, including other downstream TFs, such as LATERAL ORGAN BOUNDARIES DOMAIN 16/ ASYMMETRIC LEAVES2-LIKE 18 (LBD16/ASL18) and LBD29/ASL16 (Okushima et al., 2005; Lee et al., 2009; Goh et al., 2012a) (Figure 3.2).

Lateral root outgrowth

Lateral root primordia emergence through the overlying tissues of the parent root involves both further growth, in terms of cell elongation and division, and further differentiation, particularly the development and activation of the LRM, the definitive feature of a newly formed lateral root (Malamy and Benfey, 1997). Primordia emergence requires the coordinated separation of the overlying cells in the parent root in order to avoid excessive damage and infection risk (Swarup et al., 2008; Laskowski et al., 2006). In Arabidopsis, only three single-cell tissue layers have to be penetrated; in rice as many as 15 cell layers must be penetrated for LRP emergence (Osmont et al., 2007; Péret et al., 2009b).

This process of root cell separation for root primordial emergence is regulated by basipetal, shoot-derived auxin (Bhalerao et al., 2002) and LRP-derived auxin (Swarup et al., 2008), promoting cell separation and upregulating the expression of cell-wall-remodeling genes in the endodermal, cortical, and epidermal cells layers overlaying the LRP (Swarup *et al.* 2008). LAX3, a high-affinity auxin influx transporter, upregulated in response to LRP-derived auxin, and specifically expressed in the epidermal and cortical cells overlaying LRP, facilitates auxin influx in these cells, spatially regulating the subsequent expression of auxin-induced genes involved in cell wall remodeling (Swarup et al., 2008). These cell-wall modification genes encode a suite of enzymes, including pectate lyases such as PLA2, pectin methylesterases

(MPEs), polygalacturonidase (PG), an expansin (EXP17), and at least one known glycosyl hydrolase, GLH17, all of which are implicated in facilitating cell wall loosening and separation for LRP outgrowth to occur (Henrissat and Davies, 1997; Cosgrove, 2000; Marin-Rodriguez et al., 2002; Laskowski et al., 2006; Swarup et al., 2008; Ogawa et al., 2009) (Figure 3.2)

The activation of the LRM is also thought to occur during LRP emergence from the parent root (Laskowski et al., 1995). While the genes and pathways involved in this process have yet to be elucidated, a shift in auxin signaling or source of synthesis from the parent root to the new LRM is implicated, as the arrested post emergence growth of the Arabidopsis *aberrant lateral root formation3* (*alf3)* mutant can be rescued with the application of exogenous auxin, suggesting that the ability of the new LR to synthesize its own auxin may coincide or cause LM activation (Celenza, 1995; Péret et al., 2009b) (Figure 3.2). Multiple Aux/IAA-ARF modules, including the SHY2/IAA3–ARF module (Goh et al., 2012b), may play a role in the complex networks regulating LRP development and LR emergence. These networks may also be mediated post-transcriptionally by the down-regulation of LR emergence through the auxin-induced expression of *miRNA164a* and *miR164b* which target for degradation the mRNAs of NAM/ATAF/CUC 1 (NAC1) (Guo et al, 2005), a TF involved in transmitting auxin signals for LR emergence (Xie, 2000). Preliminary research shows this miR164-NAC1 regulatory module may also be conserved in tomato (Zeng et al, 2010) and rice (Meng et al, 2009).

Lateral root elongation

The genetic control of post-emergence lateral root elongation affects the rate and angle of LR growth, LRM determinancy and branching potential, all of which are important considerations in root system architecture. Not much is known about the genetic control of these traits; however,

these are areas under active research. The Arabidopsis *PLETHORA 1* and *2* (*PLT 1* and *2*)and *CLAVATA 3* (CLV3) genes are implicated in both primary and LRM maintenance of the root stem cell niche and QC, as mutants of these genes fail to maintain the QC and root stem cells, and thus stop root elongation (Aida et al., 2004; Fiers et al., 2004). In vitro application of the artificially synthesized, mature CLV3 peptide, a 12-amino acid ligand, processed from the conserved 14-amino acid CLE domain of a larger peptide (Fiers *et al.* 2006), and peptide synthesis or overexpression of other members of its greater CLV3/ESR (CLE) family of related proteins sharing the conserved and essential CLE motif, all caused the termination of root development (Strabala and O'Donnell 2006; Kinoshita *et al.* 2007), suggesting other CLE genes could be involved in regulating RAM identity (reviewed in Miwa et al., 2009). Cell division and elongation, particularly elongation or expansion is one of the primary drivers of root growth rate, and while the genes involved have not yet been cloned, the maize mutants *short lateral root1* and *2* (*slr1* and *slr2)* display short, slow-growing lateral roots on their primary and embryonic crown roots, which microscopy studies haves attributed to a decrease in cell elongation (Hochholdinger et al., 2001). Hormonal interactions also play a role in lateral root growth: auxins, ethylene, and ABA have been shown to inhibit lateral root elongation, while CKs promote elongation (Rani Debi et al., 2005; Iwama et al., 2007) (Figure 3.2). Amongst the many auxin transporters potentially involved in LR elongation, ABCB19/MDR1, an important shoot basipetal auxin transporter, has also been shown to be important for root acropetal auxin transport and necessary for maintenance of a high enough auxin concentration to support post-emergence LR elongation at a normal rate (Wu et al., 2007).

The angle of LR growth is thought to be at least partially under genetic control due to tropic responses, as different Arabidopsis and rice accessions display variations in lateral root angle

(Mullen and Hangarter, 2003; Iyer-Pascuzzi et al., 2010), which may be attributable to differences in intrinsically programmed LR gravitropic setpoint angle (GSA), the angle of growth relative to the gravity vector (Digby and Firn, 2002). Mutant analyses of Arabidopsis lines with a normal primary root gravitropic response, but variations in LR GSA suggest that the genetic control of GSA may be independent between lateral and primary roots, and that GSA may be mediated by auxin signaling and a root phototropic response (Mullen and Hangarter, 2003).

Crown roots – From initiation to elongation

Crown roots, also called nodal or shoot-borne roots, are adventitious roots unique to monocots and part of normal monocot root system development. Along with their associated lateral roots, crown roots make up the bulk of the fibrous monocot root system. Crown roots may be developmentally separated into two different types: the embryonic crown roots--seminal roots which form around the coleoptilar node along with the primary root (radicle) during embryogenesis, and the post-embryonic crown roots that arise during germination and throughout the life of the plant (Hochholdinger and Tuberosa, 2009). Along with dicot root and the monocot seminal primary root, all crown roots, both embryonic and post-embryonic, can be considered primary order roots, as like the radicle they arise from the main stem of the plant and not from another root as do lateral roots.

Crown root primordia initiation and development

Most root development research has focused on primary and lateral roots, thus much if the current knowledge about the genetic control of crown root development is deduced from studies of maize and rice mutants or based on comparative analysis with Arabidopsis primary, lateral, and adventitious root studies. The overarching hormonal regulation and the gene families

regulating primary, lateral, crown (in monocots), and adventitious (in dicots) root growth appear to be largely conserved (Coudert et al., 2010; Hochholdinger et al., 2004). The functions of individual genes in the genetic pathways regulating the development may, however, be slightly different.

Crown root primordia (CRP) initials are produced from periclinal divisions of parenchyma cells which give rise to the pattern arrangement of differentiated epidermis/endodermis initials, central cylinder cells, and root cap initial cells (Itoh et al., 2005). This is followed by the establishment of epidermis and endodermis by periclinal divisions of the endodermis-endodermis initials, and then the formation of the cortical cells and central metaxylem (Itoh et al., 2005).

Similar to early processes in primary and lateral root development, the initiation and development of crown roots is also controlled by auxin mediated signaling (reviewed in Rebouillat et al., 2009). OsGNOM1, an ortholog of Arabidopsis GNOM1, was found to be involved in regulating proper PIN1 auxin efflux protein trafficking, and thus the polar auxin transport necessary for auxin gradient formation to signal the proper asymmetrical division of parenchyma cells for CRP development (Geldner et al., 2003; Liu et al., 2009; Péret et al., 2009b; Richter et al., 2010). Maize and rice homologs of the Arabidopsis *SHR* and *SCR* genes, GRAS TFs, also have been shown to be essential for the radial patterning necessary for CRP development. With a similar endogenous expression pattern to the Arabidopsis genes and *in vitro* evidence of the capacity for interaction between each species pair, it is likely that in monocots the two TFs share a similar role in crown root, as opposed to lateral root primordia development and interact with each other to restrict the formation of the endodermis to a single cell layer (Cui et al., 2007).

There is also evidence to suggest that the monocot radicle/primary seminal root, the embryonic crown roots, and the postembryonic crown roots may be under different genetic control. The monogenic maize mutant *rootless concerning crown and seminal roots* (*rtcs*) does not form any crown roots, just the primary root and its associated laterals (Hetz et al., 1996). Other monogenic maize mutants display less severe root developmental phenotypes: *lateral rootless 1* (*lrt1*) does not develop crown roots at the coleoptilar node or any lateral roots on the primary root or remaining embryonic crown roots (Hochholdinger and Feix, 1998), whereas the *rum1* mutant has no embryonic crown roots, and few, late-developing lateral roots and postembryonic crown roots (Woll et al., 2005). Rice mutants *crown rootless1* (*crl1*) and *adventitious rootless 1* (*arl1*), found to be allelic, have no crown roots or crown root primordia, fewer lateral roots off the primary root, and an abnormal gravitropic response (Inukai et al., 2001). Rice *ARL1/CRL1* and *RTCS* have been shown to encode LBD (Lateral organ Boundary Domain) proteins similar to those encoded by the Arabidopsis *LBD16* and *LBD29* genes (Inukai et al., 2005; Liu et al., 2005; Taramino et al., 2007). All genes are members of the same family and are probably auxin responsive, having auxin response elements (AREs); however, they each have different functions. *LBD16* and *29* are involved in lateral root formation in Arabidopsis, the maize *RTCS* gene is involved only in crown root development, and the rice *ARL1/CRL1* gene in both lateral and crown root development (Figure 3.3) (Inukai et al., 2005; Liu et al., 2005; Taramino et al., 2007).

Similar to lateral root formation in Arabidopsis, CKs also plays a secondary role in mediating crown root development in monocots through antagonism of auxin-related signaling pathways. The rice *WUSCHEL-RELATED HOMEOBOX 11* (*WOX11*) gene encodes an auxin and CK-induced TF expressed in early crown root primordia and the actively dividing regions of

the shoot apical meristem (Zhao *et al.* 2009a) and found to repress the CK-upregulated type-A response regulator gene, *RR2* (Jain *et al.* 2006), which may function as a negative regulator of CK signaling and may repress cell proliferation in the CR meristem, thus repressing CR emergence (Zhao *et al.* 2009a). Knockout mutants of *WOX11* exhibited inhibited crown root growth, while overexpression of the gene increased rates of crown root cell division, leading to precocious crown root growth. Additionally both mutant and overexpressor lines also showed altered transcription of auxin and CK-responsive genes, suggesting that *WOX11* may play a pivotal role in integrating auxin and CK signaling to control cell division rates in the crown root primordia (Figure 3.3) (Zhao *et al.* 2009a).

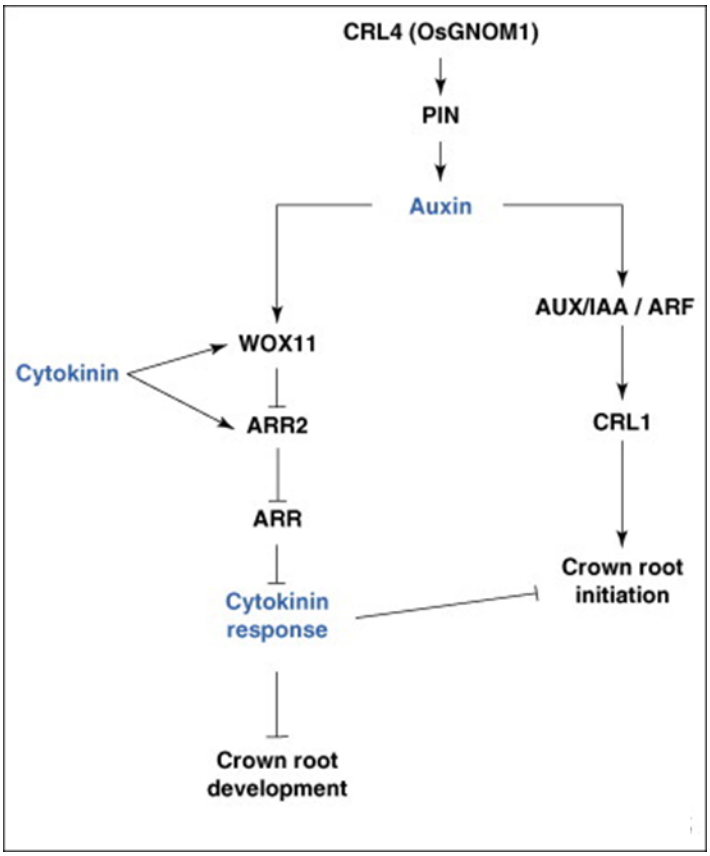


Figure 3.3. Hormonal and genetic control of crown root formation in rice. Crown root initiation in rice is promoted by auxin, and regulated by the inhibitory influence of cytokinin. Arrows represent the positive regulatory action of one element of the network on another one. A line ending with a bar represents the negative regulatory action of one element of the network on another one. Abbreviations: ARF, Auxin Response Factor; ARL, ARR, type-A RESPONSE REGULATOR; AUX/IAA, AUXIN/INDOLE-3-ACETIC ACID; CRL4, CROWN ROOTLESS4; GNOM1, GTP:GDP ANTIPORTER/ PROTEIN HOMODIMERIZATION1; PIN1, PINFORMED1; WOX11, WUSHEL-Related Homeobox 1 (Coudert et al., 2009).

Crown root outgrowth and elongation

While the formation of CRP is under genetic and physiological control, the emergence of developing crown roots from stem nodes is at least partially influenced by the environment. Mergemann and Sauter found that in accessions of deep-water rice, the buildup of ethylene caused by submergence induces the death of epidermal cells above CRP, thus promoting emergence of crown roots through the epidermis of the submerged nodal branches (Mergemann and Sauter, 2000).

Recent studies on this phenomenon have shown that GA is also involved as a non-essential but synergistic upregulator of CRP emergence and elongation rate in the presence of ethylene, and ABA as a likely inhibitor of both ethylene and GA signaling pathways (Steffens and Sauter, 2005; Steffens et al., 2006). While the specific hormone biosynthesis, signaling, and target genes implicated in this H_2O_2 programmed cell death pathway have not yet been identified, it has been shown that the epidermal cells overlying CRP may be predestined to die, exhibiting a lower transcription level of *METALLOTHIONEIN 2b* (*MT2b*), which encodes a reactive oxygen scavenger that, in higher levels, would prevent cell damage by H_2O_2 (Steffens and Sauter, 2009). It is possible that CRP emergence may also be auxin-regulated, as rice RNAi-knockdown lines of the *OsPIN1* gene, which encodes an auxin efflux carrier, show arrested CRP emergence (Xu et al., 2005); however, the physiological mechanism by which auxin signaling influences CRP emergence is yet unknown.

Strigolactones may play a role in positively regulating CR elongation through promoting root meristematic cell division (Arite et al, 2012), potentially through modulating auxin flux. Rice *dwarf* mutants for genes involved in SL biosynthesis (SL-deficient rice mutants *max3/rms5/d17*,

max4/rms1/d10, and *d27*) or SL signaling (SL-insensitive rice mutants *max2/rms4/d3* and *d14*) were found to have a short CR phenotype due to an apparent decrease in cell division, leading to a narrower meristematic zone (Arite et al, 2012). This decreased cell division may be due to SL-modulation of local auxin levels, affecting meristem cell number as seen in primary roots of homologous Arabidopsis SL-deficient and SL-insensitive mutants (Kapulnik et al, 2011; Ruyter-Spira, 2011); however, the specific mechanism of SL effect on root growth has yet to be fully elucidated.

Extrinsic Pathways - Root system architecture changes in responses to environmental stimuli

The intrinsic genetic pathways detailed previously control the normal development of plant root systems by directing the primordia initiation, outgrowth, and elongation of various root types. Modulation of these pathways in response to the environment allow plants the phenotypic plasticity to modify specific components of their root system architecture to exploit limiting nutrient resources and respond to a constantly fluctuating complex of biotic and abiotic stresses. Even different ecotypes or varieties from the same species that are adapted for growth in dissimilar rhizosphere environments can vary widely in intrinsic root system development schemes and plasticity responses, resulting in heritably different RSAs (Malamy 2005; Suralta *et al.* 2008; Gowda *et al.* 2011; Clark *et al.* 2011; Pacheco-Villalobos and Hardtke 2012) (Figure 3.4).

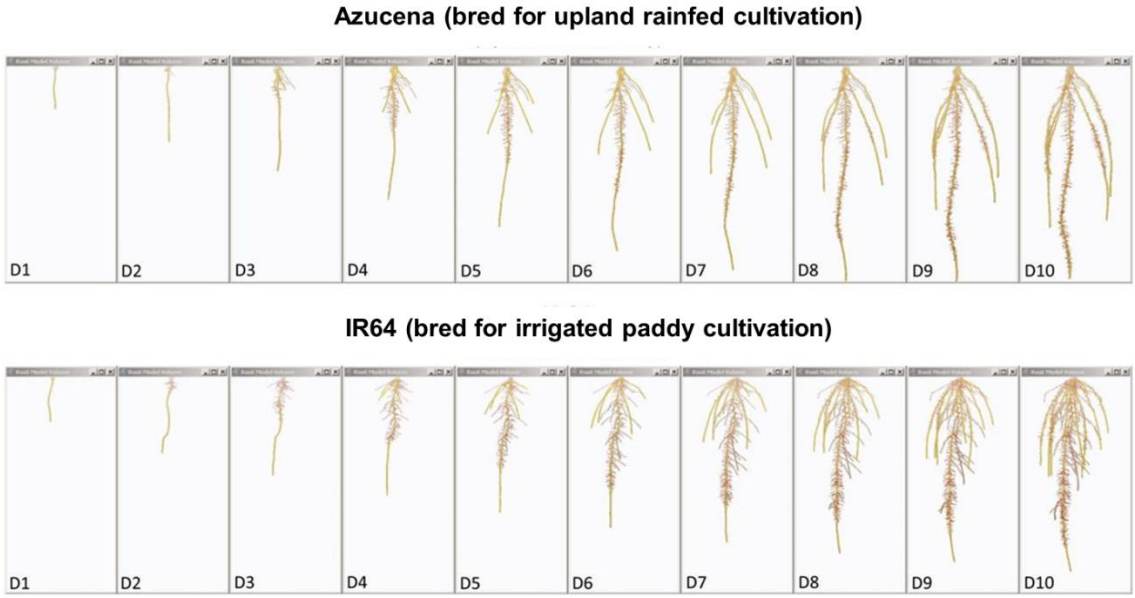


Figure 3.4. Root system models of two rice varieties bred for contrasting agricultural systems. These root system models, generated from image series of seedling rice root systems of cv. Azucena (top), and cv. IR64 (bottom) over ten days of growth (D1-D10) in a clear, gellan-gum nutrient media show how the breeding of crop varieties adapted to particular cultivation systems and agroecological environments has resulted in inadvertent selection for different crop root architectures. Azucena, a rice variety bred for an upland rainfed growing environment develops a deeply rooted primary and crown root architecture consistent with rapid growth in search of water, whereas IR64, bred for a water-sufficient, irrigated paddy system is more shallowly rooted, but develops longer, highly branched lateral roots in the top part of the root system to scavenge nutrient resources, such as nitrogen and phosphorus, from near the soil surface. Primary and crown roots are shown in yellow; the root system skeleton is shown in red. (Modified from Clark et al, 2011; models were generated using RootReader3D software).

While the genes and pathways involved in environmental perception and signaling may be unique to a particular stimulus, root growth response pathways often feed into the underlying genetic pathways by co-opting hormonal regulation. Current understanding of the genetic and hormonal regulation of RSA changes induced by tropisms, nutrient availability, toxic compounds, symbioses, and abiotic stresses are reviewed here and in Table 3.2.

Gravity

The downward growth of roots influences RSA traits such as root angle, length, and depth, is primarily controlled by a positive gravitropic response, hypothesized to be perceived through the sedimentation of statoliths (amyloplasts--starch-containing plastids, or other plastids, such as chloroplasts) within statocytes, specialized gravity-sensing cells in the root tip (reviewed in Morita, 2010). The mechanism of gravity sensing is yet unknown, but is postulated to be through statolith pressure or movement receptor perception, or pressure-based opening of mechano-sensitive ion channels (reviewed in Perrin et al., 2005). In Arabidopsis, ALTERED RESPONSE TO GRAVITY 1 (ARG1) and ALTERED RESPONSE TO GRAVITY-LIKE 2 (ARL2), J-domain proteins localized to endomembrane organelles, are thought to intact with one another to form a gravity signal transduction complex, promoting rapid, transient cytoplasmic alkalinization through Ca^{2+} influx, and the redistribution of auxin efflux carrier PIN3 to the lower membrane of the statocytes (Boonsirichai et al., 2003; Harrison and Masson, 2008a; Harrison and Masson, 2008b). The relocalization of PIN3 results in the asymmetric redistribution of auxin along the new, lowest side of the root tip (Young and Evans, 1996; Lucas et al., 2008; Nishimura et al., 2009), followed by basipetal auxin transport to the root distal elongation zone, mediated by the auxin influx carrier AUX1 and efflux carrier ETHYLENE

INSENSITIVE ROOT 1 (EIR1) (Blancaflor and Masson, 2003; Swarup et al., 2005; Band et al., 2012; Brunoud et al., 2012). This new auxin gradient is thought to signal the upregulation of genes leading to cell elongation along the top end of the distal elongation zone, thus producing root tip curvature downward (Blancaflor and Masson, 2003; reviewed in Petrásek and Friml, 2009). In addition to auxin, other phytohormones or signaling molecules, including cytokinins (Aloni et al., 2006), reactive oxygen species (Cervantes, 2001; Joo et al., 2001), flavonoids and ethylene (Buer et al., 2006; Edelmann and Roth, 2006) may be involved in gravitropic root tip curvature growth response by controlling differential cell elongation in parallel with auxin or as regulators of the auxin-mediated signaling pathway.

The aforementioned concept of a genetically-controlled measure of gravitropism, the gravitropic setpoint angle--the equilibrium angle (or range of angles) from vertical at which an organ shows no gravity-induced differential growth (Digby and Firn, 1995), has bearing on RSA traits such as crown root and lateral root angle of growth. Mutant analyses of Arabidopsis lines with a normal primary root gravitropic response but variations in LR GSA suggest that the genetic control of GSA may be independent between lateral and primary roots, and that GSA may be mediated by auxin signaling and a root phototropic response (Mullen and Hangarter, 2003).

Light

Although the root systems of most plants are largely underground and not exposed to light, plant roots may be exposed to light through ambient diffusion or soil upheaval and have been found to possess phytochromes, phototropins, and cryptochromes, including both red and blue light photoreceptors (Ruppel et al., 2001; Mullen et al., 2002; Galen et al., 2007; Molas and Kiss, 2008). Root responses to light have been studied mostly in Arabidopsis, which is found to

display a negative phototropic response to blue light, mediated by the root phototropin (PHOT1) (Galen et al., 2007; Huala et al., 1997; Christie et al., 1998), and a positive phototropic response to red light, mediated by the root-expressed phytochromes A and B (PhyA and PhyB) (Kiss et al., 2003). PhyA also promotes root elongation under exposure to far red (Kurata and Yamamoto 1997; Correll *et al.* 2003; Costigan *et al.* 2011) and blue light (Costigan *et al.* 2011). Auxin concentration differentials may be partially responsible for root growth responses to shoot light exposure, as the proper plasma membrane localization of the auxin efflux transporter PIN2 was found to be greatly increased in light-grown, but the protein was targeted for vesicular degradation in dark-grown seedlings (Laxmi et al., 2008). JA is also implicated in a root-localized light response, as one study has demonstrated that phytochromes, or more specifically, phytochrome chromophores are necessary for the JA-mediated root growth inhibition (Costigan *et al.* 2011).

Water availability

Given that one of the main functions of the root system is water uptake, soil water availability and soil hydraulic conductivity, especially in the extreme conditions of drought leading to water deficiency or flooding leading to soil saturation and hypoxia, is arguably the most important environmental factor influencing root growth and development. Studies in Arabidopsis have shown that decreasing osmotic potential as a representation of drought stress reduces the LR outgrowth and emergence from LRP of plants grown on agar plates (Deak and Malamy 2005). Similar research in maize has shown that small increases in negative water potential stimulate primary root elongation, but further water stress decreases the rate of PR growth (Sharp *et al.* 1988; Wiegers *et al.* 2009).

Hormonal signaling controlling root growth responses to water availability is not yet fully elucidated, however, ABA has been shown to stimulate PR elongation and LR emergence in response to drought (De Smet *et al.* 2006). In contrast, in flooded deepwater rice plants, a decreased internode ABA level and the concurrent accumulation of GA and the ethylene produced as a response to hypoxia and flooding stress, initiates the programmed cell death of adventitious root primordia epidermal cells, allowing the adventitious root development and outgrowth (Mergemann and Sauter 2000; Steffens *et al.* 2006). Similarly, the Arabidopsis *LONG CHAIN FATTY ACID SYNTHETASE 2 (LACS2*) gene essential for cutin biosynthesis was shown to be required in order for plants to be able to synthesize a cutin layer that suppresses lateral root emergence under low water availability (Macgregor *et al.* 2008). The rice ERF-like TF SUBMERGENCE 1 (SUB1) (Xu *et al.* 2006), a TF involved in mediating responses to both plant submergence and drought, may also be one of many genes involved in regulating root growth under water stress, as osmotic stress-induced inhibition of root growth was found to be slightly suppressed in rice varieties with a functional copy of the SUB1 gene (Fukao *et al.* 2011).

Growth in response to soil nutrients

Plant root adaptive growth in response to soil macro and micronutrients depends on a wide array of variables: nutrient forms, availability, concentration, localization, and nutrient behavior in soil, as well as the nutrient status of the plant. Similar to the tropic responses above, plant root growth in response to a nutrient stimulus requires four main steps: stimulus perception, signal transduction, target gene regulation, gene product mediation of growth response.

Nitrogen

Nitrogen, the most limiting nutrient to plant growth is an interesting example of these highly

plastic plant responses to nutrient availability, as it can inhibit LR outgrowth, development and elongation under high N conditions, or in soil with low inorganic nitrogen, soil patches with high inorganic nitrogen can have a local, stimulatory effect on LR elongation and branching within the high N area. Arabidopsis senses nitrate through the primary root tip, with downstream components of the nitrate LR growth response pathway include high and low-affinity Arabidopsis *NITRATE TRANSPORTERS 1.1* and *2.1*, (*AtNRT1.1* and *AtNRT2.1*) (Zhang et al., 1999; Malamy and Ryan, 2001), and nitrate-responsive TFs, including the MADS box TF *ARABIDOPSIS NITRATE REGULATED 1* (*ANR1*) (Zhang and Forde, 1998).

The nitrate transporters may be either nitrate sensors or, transporters that facilitate N movement for detection via another protein. AtNRT2.1 is necessary for LR growth repression in plants with a high external carbon to nitrogen value (Malamy and Ryan, 2001; Little et al., 2005; Remans et al., 2006), and AtNRT1.1 is a dual-affinity transporter induced by both auxin and nitrate and important for nitrate uptake under high N conditions (Liu et al., 1999; Guo et al., 2002; Munos et al., 2004). AtNRT1.1 is also an auxin influx facilitator, decreasing its auxin transport activity in response to nitrate sensing, and is proposed to repress lateral root development by promoting basipetal auxin transport out of LRP under low external nitrate conditions (Krouk et al., 2010). ANR1 mediates the localized N response, regulating the increased proliferation of LRs in N-dense patches, and may be a direct or indirect target of the signal perception/transduction pathway involving AtNRT1.1 (Zhang and Forde, 1998; Remans et al., 2006). ABA may also act in the same pathway as nitrate by inhibiting LR growth under high N conditions (Signora et al., 2001; De Smet et al., 2003). SLs appear to be upregulated in plants under low N conditions (Yoneyama et al, 2007b); however, whether increased these SL levels have a definite impact on root growth has yet to be determined.

Changes in RSA may also be induced depending on the prevailing available organic form of nitrogen, such as L-glutamate or carnitine. In Arabidopsis seedlings, the sensing of L-glutamate by the primary root tip inhibits cell division in the primary root meristem and induces LR formation and outgrowth. L-glutamate may act more as a signaling molecule as opposed to a nitrogen source, as several Arabidopsis auxin-signaling mutants display varying levels of sensitivity to L-glutamate (Walch-Liu et al., 2006), and a rice glutamate receptor mutant displays a host of RSA changes, with short primary and lateral roots, reduced cell division and RAM cell death (Li et al., 2006). Carnitine, transported in Arabidopsis by AtOCT1, has been shown to stimulate LR formation, perhaps by locally affecting the C:N ratio important in modulating LR development (Lelandais-Briere et al., 2007).

Phosphorus

Phosphorus is the second most limiting nutrient because of its high affinity to bind metals in acidic and alkaline topsoil layers, forming insoluble substrates. Phosphorus is taken up by plants as phosphate (Pi), either directly by the root system or, in arbuscular mycorrhizae host plants, may also be transferred through the fungal symbiont-- the genetic control of which will be explored in detail later in this paper.

Under high Pi conditions in Arabidopsis, primary root growth is promoted, while LR growth is inhibited (Linkohr et al., 2002). Under natural conditions where Pi is limiting, plants adopt a root foraging strategy to explore topsoil layers for phosphorus. This Pi foraging strategy may be accomplished through one of several different RSA and physiological changes. In Arabidopsis and rice, growth shifts to favor an increased root:shoot ratio, with a higher initiation and outgrowth of lateral roots, forming a shallow, highly branched root system (Williamson et al.,

2001; Lopez-Bucio et al., 2002; Shimizu et al., 2004). Under low Pi conditions, Arabidopsis primary root growth is inhibited (Williamson et al., 2001; Linkohr et al., 2002; Lopez-Bucio et al., 2002), while root hairs increase in density and length (Bates and Lynch, 1996; Bates and Lynch, 2000). In legumes, including soybean, pea, and common bean (*Phaseolus vulgaris*), basal root growth angle is shifted from a downward to a more horizontal direction (Bonser et al., 1996), though a recent study shows the opposite effect in Arabidopsis, with LR gravitropic setpoint angles shifting to a steeper, downward orientation under low Pi conditions (Bai et al, 2013). Several different families of plants develop proteoid or cluster roots-- highly branched bunches of lateral roots just below the soil surface that secrete phosphatases and organic acids which solubilize bound phosphate for uptake (Shane and Lambers, 2005; Schulze et al., 2006).

In Arabidopsis, the PR tip is the key organ involved in phosphate sensing, and the initial effect of low external Pi perception is the inhibition of PR growth by the loss of meristem activity and cell elongation (Williamson et al., 2001; Sanchez-Calderon et al., 2005). While a plant Pi-receptor has yet to be identified, studies suggest that the P_5 type ATPase PHOSPHATE DEFICIENCY RESPONSE 2 (PDR2), and multicopper oxidase LOW PHOSPHATE ROOT 1 (LPR1)function in an endoplasmic reticulum-localized Pi-signaling pathway (Ticconi et al., 2004; Reymond et al., 2006; Ticconi et al., 2009). PHOSPHATE STARVATION RESPONSE 1 (PHR1) (Bari *et al.* 2006), an Arabidopsis MYB-like TF that binds the promoter sequences of low-Pi induced genes, and its regulator SMALL UBIQUITIN-LIKE MODIFIER 1 (AtSIZ1) (Miura *et al.* 2005), a small ubiquitin modified E3 ligase, and the downstream PHOSPHATE 2 (PHO2), an E2 conjugase, and the microRNA *miR-399*, which regulates PHO2 expression, are all involved in Pi-deficiency-related transcriptional changes (Bates and Lynch, 2000; Bari et al., 2006). The Arabidopsis WRKY75 TF is also induced during Pi-deprivation and may modulate both

phosphate and non-phosphate induced LR development and control the transcription of genes such as high-affinity Pi transporters important for Pi uptake (Devaiah and Raghothama, 2007; Devaiah et al., 2007). The Pi-induced tobacco bZIP TF PHOSPHATE INDUCED 2 (PHI2) (Sano and Nagata, 2002) and rice bHLH TF PI STARVATION-INDUCED TRANSCRIPTION FACTOR 1 (OsPTF1) (Yi et al., 2005) may also have a role in modulating low-Pi induced changes in RSA.

Increased auxin sensitivity, decreased CK sensitivity, and changes in auxin transport and localization appear to be at least partially responsible for Pi stress-induced LR development. A shift in auxin overaccumulation from the PR apex to the LRP, or an increased sensitivity of LRP to auxin have been suggested as proposed mechaniums for increases in LRP emergence and LR density (Franco-Zorrilla et al., 2002; Lopez-Bucio et al., 2005; Nacry et al., 2005). TIR1 auxin receptor-dependent degradation of TF-repressing AUX/IAA proteins is essential for LR development in Pi-stressed seedlings (Perez-Torres et al., 2008). The effect of auxin under low-Pi conditions is also regulated by CK signaling, which represses auxin-induced gene transcription Pi-starved Arabidopsis plants display a decreased response to CK, partly due to the reduced expression of the CR receptor *CRE1* (Franco-Zorrilla et al., 2002). Ethylene perception is likely also necessary for increased root hair development and LR elongation and decreased primary root elongation under low Pi conditions (Schmidt and Schikora, 2001; Lopez-Bucio et al., 2002; Ma et al., 2003) and has additionally been shown to affect Pi stress-induced changes in basal root growth angle in bean (Lynch and Brown, 2001). Similar to CK, GA acts as a negative repressor of Pi-induced root architecture changes under low-Pi conditions; Pi-deficient plants accumulate DELLA-proteins, which repress GA-induced root growth suppression and thus allow for auxin-mediated LR initiation and elongation (Jiang et al., 2007). SL production is induced by

low Pi in many species including tomato, Arabidopsis, pea, and rice (López-Ráez and Bouwmeester, 2008; Kohlen et al., 2011; Mayzlish-Gati et al., 2012; Foo et al., 2012; Umehara et al., 2010). Some studies suggest that increased production and exudation of SLs under soil Pi or N deficiency is dependent on whether the plant 1) is an AMF-compatible host, and 2) whether it is dependent on the AMS for Pi and N uptake (Umehara et al., 2010; Yoneyama et al., 2007b, 2007a, 2008); however what effect, if any, this increased SL exudation has on root growth is unclear. Exogenously-applied GR24 appears to increase LR formation under low Pi or decrease LR formation under sufficient Pi though the F-box protein MORE AXILLARY GROWTH2 (MAX2), a putative component of the SL-signaling pathway (Kapulnik et al, 2011; Ruyter-Spira, 2011).

Sulfur

Sulfur, taken up by plant roots as sulfate, is another limiting plant macronutrient, and is essential for the synthesis of methionine and cysteine. Sulfur deficiency can have significant effects on RSA; sulfate limited Arabidopsis and maize plants increase their lateral root production, developing an extensive, highly branched root system, often at the expense of shoot growth (Kutz et al., 2002; Bouranis et al., 2008). Another conflicting Arabidopsis study found a decrease in LRP and emerged LR under low-sulfate growth conditions (Dan et al., 2007). To rectify these two opposing developmental outcomes, a two-state model was proposed wherein short-term sulfur limitation let to increased LR growth for sulfate foraging, but longer-term sulfate deficiency led to overall decreased growth and photosynthesis, ending in premature senescence (Hoefgen and Nikiforova, 2008; Lewandowska and Sirko, 2008).

While the genes involved in internal and external sulfate sensing and transcriptional regulation have not yet been cloned and characterized, several components of root sulfate import and signal transduction have been identified. Of the five major classes of sulfate transporters identified in Arabidopsis and rice (Takahashi et al., 1999; Buchner et al., 2004; reviewed in Takahashi, 2010), the Group 1 high-affinity transporters are essential for root sulfate uptake. Arabidopsis *SULFATE TRANSPORTER 1;2* (*SULTR1*;*2*) is expressed under both sulfate-sufficient and low-sulfate conditions and transcriptionally regulated by the ETHYLENE-INSENSITIVE3-LIKE3 TF SLIM1, whereas the *SULTR1;1* gene induced only under sulfate stress (Takahashi et al., 2000; Yoshimoto et al., 2002; Maruyama-Nakashita et al., 2004) and upregulated by O-acetylserine (thiol) lyase (OAS), a rate-limiting enzyme involved in sulfate assimilation into cysteine (Leustek et al., 2000; Saito, 2000).

Auxin may play a central role in LR production under sulfate stress. In Arabidopsis, sulfate deficiency activates the transcription of *NITRILASE 3* (*NIT3*), which converts indole-3-acetonitrile to the auxin indole-3-acetic acid (IAA) (Kutz et al., 2002). However, while NIT3 activity is especially upregulated in LRP under sulfate limitation, increased concentrations of auxin have not been proven (Kutz et al., 2002; Lewandowska and Sirko, 2008). Studies of sulfur-limitation regulated auxin signaling genes such as *BIG*, named for the huge 560 kD calossin-like protein it encodes, required for the polar transport of auxin (Gil *et al.* 2001), as well as the auxin TF genes *IAA13, IAA28* and *ARF-2*, indicate that auxin is likely involved in the indirect regulation of sulfur homeostasis and short to long-term sulfur deficiency responses (Hirai et al., 2003; Maruyama-Nakashita et al., 2003; Nikiforova et al., 2003; Hoefgen and Nikiforova, 2008; Lewandowska and Sirko, 2008). Jasmonic acid (JA) may also play a role in sulfur regulation, as demonstrated by research in Arabidopsis finding low sulfur JA biosynthesis

genes upregulated under low sulfur in (Maruyama-Nakashita et al., 2003), exogenous application of JA promoted sulfur assimilation and there is also evidence to suggest that CKs and sucrose may affect sulfur responsive gene transcription (Ohkama et al., 2002).

Toxic compounds

High soil concentrations of naturally occurring soluble salts, aluminum, and heavy metals, such as cadmium, lead, and chromium, can be highly phytotoxic and seriously impair plant root growth. Plants exhibit two main strategies to manage toxic soil compounds: 1) producing root exudates that bind and neutralize the toxin in the rhizosphere, and 2) actively transporting the compound into the root, but neutralizing and sequestering it in vacuoles for safe accumulation, or eliminating it through exudation.

Aluminum toxicity

Aluminum is the 3rd most abundant element and the most abundant metal in the Earth's crust. Aluminum toxicity is one of the major constraints to yield productivity worldwide, especially in the acid soils of the tropics and subtropics that comprise almost 50% of all non-irrigated arable land in those regions (Uexküll and Mutert, 1995). At a soil pH of 5.5 or less, Al is solubilized into Al^{3+}, its phytotoxic form, which has a high plant uptake affinity through diffusion (Kochian, 1995). $Al3^{+}$ is highly toxic to plant growth, causing a rapid inhibition of root apical cell expansion and elongation, and the eventual cessation of cell division, resulting in a stunted, brittle root system with swollen malformed tips, inhibited LR initiation and outgrowth, deformed root hairs, and a poor nutrient and water uptake capacity (Foy, 1984; Delhaize et al., 1993; Kochian, 1995; Matsumoto, 2000).

In addition to Arabidopsis, several cereal crops, such as, maize, rice, sorghum, and wheat have been used to examine the physiological and molecular mechanisms of aluminum tolerance, as members of the grass family appear to be among the most resistant to aluminum toxicity (Delhaize et al., 1993; Magalhaes et al., 2004; Mao et al., 2004; Caniato et al., 2007). The two most well-studied mechanisms of aluminum tolerance include external avoidance, through root secretion of organic acids such as malate, citrate, and oxalate, which chelate $Al3^{+}$ ions in the rhizosphere, preventing their diffusion into roots (Miyasaka et al., 1991; Delhaize et al., 1993; Ma and Furukawa, 2003), and true, internal tolerance, by the uptake, organic acid chelation, and sequestration of bound aluminum substrates (Matsumoto et al., 1996; Ma et al., 2001; Huang et al., 2009; Klug and Horst, 2010); however, only the molecular pathways involved in $Al3^{+}$-stress induced RSA changes will be discussed below.

The site of $Al3^{+}$ sensitivity in maize is the root apex (Ryan et al., 1993); however, exposure of only the distal transition zone of maize roots to $Al3^{+}$ was found to reduce cell elongation in the elongation zone (Sivaguru and Horst, 1998), suggesting the presence of a diffusible signal between the zones, later found to be the ethylene-mediated basipetal transport of auxin (Kollmeier et al., 2000; Sun et al., 2010). In Arabidopsis, the ethylene receptor gene *ETHYLENE RECEPTOR 1* (*ETR1I)* (O'Malley et al., 2005) and the ethylene signal transducer *ETHYLENE INSENSITIVE 2* (*EIN2*) (Alonso et al., 1999) were found to be necessary to the $Al3^{+}$ induced inhibition of root elongation (Sun et al., 2010). These genes, likely along with other members of the ethylene signaling pathway, are essential for $Al3^{+}$ induced upregulation of the Arabidopsis ethylene synthesis genes *1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID SYNTHASE 2*, *6*, and *8* (*AtACS2*, *AtACS6*, *AtACS8)* and *1-AMINOCYCLOPROPANE-1-CARBOXYLIC ACID OXIDASE 1* and *2* (*AtACO1*, and *AtACO2*) (Tsuchisaka and Theologis,

2004), followed by the upregulation of auxin transporters *AtPIN2* and *AUX1*, leading to auxin distribution changes that are likely responsible for the inhibition of root elongation (Sun et al., 2010).

The binding of $Al3^+$ to negative binding sites on root cell walls and plasma membranes, has also been proposed to inhibit root elongation by increasing wall and membrane rigidity leading to transverse ruptures between the dermal and outer cortical cell layers from the inner cortex, and causing root tip damage (Kopittke *et al.* 2007), as well as impaired membrane function (Miyasaka et al., 1989; Ahn et al., 2001; Jones et al., 2006; Sun et al., 2010). $Al3^+$ binds primarily to negatively charged pectin in cell walls; the degree of binding has been found to be determined not by the amount of pectin, but by its negative charge as modulated by methylation (Eticha et al., 2005) by pectin methylesterase (PME) (Schmohl et al., 2000).

Interestingly enough, the blocking of $Al3^+$ cell wall binding sites (Huang et al., 2009) may be a major mechanism of aluminum resistance in rice, which does not appear to secrete enough chelating organic acids to rely on an $Al3^+$ external avoidance strategy (Ma et al., 2002). Two genes, *SENSITIVE TO ALUMINUM RHIZOTOXICITY 1* and *2* (*STAR1* and *STAR2*) encode the nuclear binding domain and transmembrane domain, respectively, of an ABC transporter with specificity for UDP glucose that is upregulated following root exposure to $Al3^+$ (Huang et al., 2009). Both *STAR* genes are upregulated by the constitutively-expressed rice root *ALUMINUM RESISTANT TRANSCRIPTION FACTOR 1* (*ART1*), which also upregulates several other genes implicated in different aluminum tolerance mechanisms (Yamaji et al., 2009). Among these are rice homologs of genes encoding proteins implicated in modulating root elongation and cell wall elasticity: namely an α-expansin EXPA10, members of which have been shown to decrease cell

wall extension potential when exposed to Al3+,(Gao et al., 2008), and are additionally downregulated in response to $Al3^+$ (Lee and Kende, 2002), and an Arabidopsis cell wall-associated putative endochitinase CHITINASE A (AtCHIA) (Yokoyama and Nishitani 2004), likely involved in modulating cell wall extension by regulating chitin levels (Kwon *et al.* 2005).

The upregulation of 1,3-β-D-glucan synthase (Bhuja et al., 2004), resulting in callose deposition in root apices, especially in endodermal and cortical cell walls (Budíková, 1999), is another signal of $Al3^+$-induced injury, (Jones et al., 2006; Sivaguru et al., 2006). It is proposed that this abnormal callose deposition may inhibit both symplastic and apoplastic flow (Sivaguru et al., 2000; Sivaguru et al., 2006; reviewed in Horst et al., 2010), causing inhibition of root growth. It is not yet understood whether callose deposition actually represents $Al3^+$- induced injury, is a secondary cell-strengthening response to aluminum damage, or possibly even a defense response to block further $Al3^+$ binding.

Salinity

Salinity is estimated to affect at much as 20% of the world's agricultural land and 20% of the world's irrigated cropland, (Flowers and Yeo, 1995) due to a number of natural and man-made factors, including natural salinity and coastal proximity, poor water or fertilizer management, the clearing of vegetation, and prolonged cycles of drought and flooding. In most saline soils, sodium chloride (NaCl) is the most soluble and abundant salt, with calcium and magnesium chloride in lesser concentrations. The dominant causes of plant saline toxicity are complicated by the effects of saline soils on external root osmotic stress, which affects water and nutrient uptake, especially in competition with potassium (K^+) and calcium (Ca^{2+}), and internal ionic

stress most frequently from the buildup of high sodium (Na^+) concentrations (Munns and Tester, 2008).

Different species of plants have varying levels of salt tolerance, from the highly halophilic saltbush (*Atriplex* spp.) to highly sensitive species, such as rice and Arabidopsis (Munns and Tester, 2008). RSA is generally not affected as severely as shoot branching and leaf expansion under salt stress; in many plants, root growth decreases under NaCl treatment due to reduced epidermal cell division and elongation rates, likely in response to the osmotic stress (Kurth et al., 1986; Zidan et al., 1990). Salt stress also was shown to increase lateral root production and suppress primary root elongation in Arabidopsis (He et al., 2005), induce programmed cell death in rice root tips (Li et al., 2007), as well as raise the root death rate in sensitive tomato accessions (Snapp and Shennan, 1992).

Of the many mechanisms of salt tolerance—uptake inhibition, internal sequestration, leaf exclusion, root efflux, and osmotic stress tolerance (reviewed in Munns and Tester, 2008) —root uptake inhibition, efflux, and osmotic stress tolerance have probably the greatest local effect in mediating RSA changes and root growth responses. Na^+ is thought to enter the root by passive diffusion through either High Affinity K^+ transporters (HKTs), such as the rice OsHKT2;1 (Horie et al., 2007), or through non-selective cation channels (NSCCs); possibly glutamate activated receptors (GLRs), which complex with glutamate to form a channel (Demidchik et al., 2010), or cyclic nucleotide-gated channels (CNGC) (Leng et al., 2002; Tester and Davenport, 2003). In the current Arabidopsis model of Na+ stress signaling, internal Na+ presence is perceived by a yet unknown sensor, triggering cytosolic Ca2+ flux sensed by the Ca2+ sensor Salt Overly Sensitive 3 (SOS3) (Liu and Zhu, 1998), which complexes with and activates SOS2,

CBL-interacting protein kinase (Quintero et al., 2002). The SOS2/SOS3 complex is involved in controlling three different Na+ transporters to maintain a low cytoplasmic [Na+]. These include: SOS1, a plasma membrane Na+/H+ antiporter that increases Na+ efflux out of the cell (Zhu et al., 1998; Quintero et al., 2002), a vacuolar Na+/H+ exchanger (NHX1), which facilitates N+ sequestration in vacuoles (Apse et al., 1999; Gaxiola et al., 1999) and may negatively regulate HKTs , such as Arabidopsis HKT1, restricting Na+ buildup in the cytoplasm (Uozumi et al., 2000; Rus et al., 2001; Zhu, 2002; reviewed in Mahajan and Tuteja, 2005). Ionic balance between Na^+, H^+, Ca^{2+}, and K^+ is essential; under low K^+ conditions in rice, moderate levels of Na^+ influx into the roots through OsHKT2;1 transporters were found to be beneficial in partially maintaining root elongation otherwise inhibited under low K^+; however, the biochemical advantage to this phenomenon is not yet understood (Horie et al., 2007; Horie et al., 2009).

Symbiotic interaction with plant rhizobacteria and arbuscular or ectomycorrhizal fungi have also been shown to mitigate saline toxicity and alleviate salt stress, perhaps by modulation of root ion and nutrient levels (Sheng et al., 2008; Dimkpa et al., 2009; Evelin et al., 2009; Luo et al., 2009; Shilev et al., 2010). Internal fluctuations in the concentrations and transport of several hormones, including the stress-induced ABA, as well as ethylene, auxin, CKs, and possibly GAs, are observed in response to salinity stress and are mostly linked to shoot-to-root Na+ stress signaling (Kuiper et al., 1990; He et al., 2005; Khadri et al., 2006; Cao et al., 2008; Bano, 2010). Ethylene and auxin signaling were, however, found to be required for increased LR production in salt-stressed Arabidopsis seedlings in connection with the TF AtNAC2, induced by upstream EIN2 transduced ethylene signaling (He et al., 2005). Interestingly enough, auxin and ABA are also implicated in the opposite RSA response of *Medicago truncatula* under salt stress: decreased primary root elongation, LRP initiation, and LR emergence. In this study, ABA and

salt-stress both induced upregulation of HOMEOBOX 1 (HB1), a TF found to represses LRP emergence by repressing the downstream TF LBD1, which would otherwise activate downstream genes promoting LRP outgrowth (Ariel et al., 2010). Microarray comparative analysis of rice, Arabidopsis and ice plant (*Mesembryanthemum crystallinum*) revealed several dozen common genes with salinity-induced transcriptional changes, including genes involved in stress perception and osmotic regulation (Pareek et al., 2007). The precise identity of root architecture-related genes regulated by salt stress-induced TFs have yet to be determined.

Symbioses

Plant root symbiotic associations with microbes, most notably the mycorrhizal and rhizobial symbioses, have long been known to promote plant nutrient uptake efficiency. In order to support these symbioses, host plant root architecture may undergo a number of significant changes throughout the pre-contact root-microbe signaling, symbiosis development, and establishment processes detailed in the following sections on mycorrhizal and rhizobial symbioses below. Although both symbioses induce different changes in root architecture and plant nutrient status, they share some similar components in their signaling and early developmental pathways, the so-called 'SYM pathway' (Parniske, 2008). Recently, a set of seven common SYM genes/proteins required for both symbioses were identified (Parniske, 2008). These include: the Leu-rich repeat receptor kinase SYMRK/DOES NOT MAKE INFECTION 2 (DMI2), activated after nod-factor signal perception (Endre *et al.* 2002; Yoshida and Parniske 2005); two nuclear membrane-localized cation channels, CASTOR (Imaizumi-Anraku *et al.* 2004) and POLLUX/DMI1 (Ané *et al.* 2004; Imaizumi-Anraku *et al.* 2004); two nucleoporins, NUP85 (Saito *et al.* 2007) and NUP133 (Kanamori *et al.* 2006), all necessary for

inducing the Ca^{2+} spike signal (Kosuta et al., 2008); the calcium/calmodulin-dependent protein kinase CCaMK (Lévy *et al.* 2004; Mitra *et al.* 2004; Tirichine *et al.* 2006), which acts downstream of Ca^{2+} spiking and is thought to transduce the calcium signals, partly through the physical interaction and phosphorylation of CYCLOPS, a protein with a nuclear localization signal and carboxy-terminal coiled-coil domain protein of unknown function. Intersecting research on the arbuscular mycorrhizal (AM) and rhizobial symbioses have largely been carried out on the model legumes *Lotus japonicus* and *Medicago truncatula*, as neither Arabidopsis, nor any of the other non-leguminous model plants have the ability to host the rhizobial symbiosis.

Mycorrhizal symbioses

Over 90% of land plants form symbioses with mycorrhizal fungi. These symbioses improve plant nutrient capture through fungal mineral scavenging and transfer to the plant, and can be linked to significant changes in plant root architecture. Most of the research in this field, and subsequently in this review, is focused on the arbuscular mycorrhizal symbiosis (AMS), the most common type of mycorrhizal symbiosis, found in over 80% of plant species and involving the ~200 obligate biotroph fungal species of the *Glomeromycota* phylum (Schüßler et al., 2001; Strack et al., 2003). The AM symbiosis has ancient origins--estimated to be 400 million years old, it is suggested to have played a major role in the early colonization of land by plants (Pirozynski and Malloch, 1975; Simon et al., 1993). The AMS is characterized by precontact plant-fungal signaling, fungal contact and entry of the host plant root system, and the formation of arbuscules, highly branched fungal structures within root cortical cells that are the site of nutrient (primarily P, but also N, Zn, and Fe) transfer from the fungus to the plant and carbohydrate transfer from the plant to the fungus (reviewed in Parniske, 2008).

Pre-contact signaling, development, and maturation phases of the AM symbiosis all may induce changes in RSA, however, separating these changes from those induced indirectly as a result of improvements in plant nutrient status is challenging. Previous studies have generally reported increases in root branching as a result of colonization, yet a review of these studies reveal further complicating factors: plant root systems do not respond to AM fungal colonization in the same ways. Colonization-induced root responses appear to differ depending on host plant species, types (woody vs. non-woody; monocot vs. dicot), or varieties, soil water and nutrient status, especially of P, and possibly even the species of AM fungi (Hetrick et al., 1988; Berta et al., 1990; Berta et al., 1995; Olah et al., 2005; Gutjahr et al., 2009; reviewed in Hetrick, 1991; Berta et al., 2002; Parniske, 2008). Strigolactone synthesis and exudation from the roots triggers AM fungal hyphal branching, a key step in root colonization (Akiyama et al., 2005); however, the direct effect of Sls on AM symbiosis-related root growth and development is unclear and highly dependent on plant Pi and N status and concentration in the rhizosphere (see prior sections on nitrogen and phosphorus).

In maize, root thickness and overall root mass, but not LR formation, are increased by AM colonization, which also partially restores the lateral root growth completely absent in the *lateral rootless 1* (*lrt1*) mutant, possibly indicating the involvement of auxin signaling (Paszkowski and Boller, 2002). A partial hormonal influence in AM colonization-induced RSA changes may well be possible; studies have reported altered levels of auxin (Fitze et al., 2005), ethylene (Vierheilig et al., 2002), cytokinin (Dixon, 1990; Barker and Tagu, 2000), and ABA in colonized roots (Herrera-Medina et al., 2007), as well as specific roles for auxin, cytokinin, and ABA in AM symbiosis development (Barker and Tagu, 2000; Fitze et al., 2005; Ludwig-Muller and Guther, 2007). In contrast with maize, in which the AM symbiosis stimulates an increase in root

thickness, but not root number, AM colonization in rice was found to induce crown root elongation and both fine, determinate and long, indeterminate LR number (Gutjahr et al., 2009). Interestingly enough, while AMF-exposed three monogenic essential rice SYM gene mutants, *pollux-2, ccamk-2,* and *cyclops-1,* did not develop colonized roots, they showed a decrease in crown roots and an increase in LRs over non-AMF-mutant controls, indicating the presence of root growth pathways induced by AM fungi, but independent of the SYM pathway (Gutjahr et al., 2009).

The only definite example of AM fungi-induced RSA development is in the legume *Medicago truncatula,* where pre-fungal contact lateral root formation was discovered to be induced by a diffusible factor from AM fungi, the so-called 'Myc' factor of AM fungi that affects plant host signaling pathways (Olah et al., 2005). Induction of lateral root development by this pathway requires the proper function of two SYM pathway components, DOESN'T MAKE INFECTIONS 1, 2 (DMI1 and 2)(Endre *et al.* 2002; Stracke *et al.* 2002; Hogg *et al.* 2006), as well as the novel MtENOD11 protein, all of which have necessary but yet undetermined roles in pre-symbiont contact AM and rhizobium symbiosis signaling (Kosuta et al., 2003; Olah et al., 2005).

Rhizobium-legume symbiosis

The *Rhizobium*-legume symbiosis is the most prominent and well-studied of plant associations with N-fixing bacteria, and consists of a symbiotic association between the roots of legumes (Fabaceae) and root nodule-forming, N-fixing soil bacteria of the family Rhizobiaceae. Another similar, though lesser-studied, root nodule symbiosis is the actinorhizal symbiosis between plant species in three rosid orders, the Fagales, Cucurbitales, and Rosales, and N-fixing actinobacteria

of the genus Frankia (Swensen, 1996). Host plants in both symbioses benefit by gaining an internal supply of fixed-N, as well as potential increases in resistance to some disease and abiotic stresses, while the endosymbiotic bacteria gain a protected living environment and a carbon source supplied by plant photosynthate. Similar to the AMS, the Rhizobium-legume symbiosis starts with pre-contact signaling between the bacteria and host plant, followed by bacterial infection of root hairs, root hair curling, infection thread and nodule development, and bacterial colonization of nodules (reviewed in Provorov, 2000).

Colonization of legume roots may affect RSA in two ways: root nodule formation and changes in primary or lateral root growth. The two types of symbiotic nodules – determinate and indeterminate -- differ both structurally and developmentally, and are dependent on the host plant species. Cells of the tip meristem of determinate nodules fully differentiate at maturity and are not maintained resulting in spherical nodules at uniform developmental stages, whereas the tip of the meristem of indeterminate nodules is continuously active and producing new infected tissue, creating larger and longer cylindrical or bulbous nodules with different developmental zones (reviewed in Markmann and Parniske, 2009). Studies also suggest that there is a balance between lateral root and nodule formation, with nodule primordia initiation dependent on the suppression of lateral root emergence (Nutman, 1948; Lohar et al., 2004).

Given the ancient origin and near-universality of the AMS in the plant kingdom, and the familial specificity of the rhizobial symbiosis to only the Leguminosae, it has been proposed that the rhizobial symbiosis has recruited much of the key symbiotic development pathway from the AM symbiosis, then modified and evolved genes and pathways for nodulation specific functions (Markmann and Parniske, 2009). Although the functioning alleles of the seven aforementioned

known genes in the shared SYM pathway are necessary for the development of both the AM and rhizobial symbioses (Kistner et al., 2005), none of these are directly involved in symbiosis-related RSA changes. Each of these seven gene products is involved in only the early stages of the SYM signal reception and transduction pathway. The downstream, symbiosis-activated genes and networks feeding into intrinsic hormone-controlled and nutrient-modulated root growth pathways are what is actually involved in regulating *Rhizobium*-induced nodulation and lateral root development to balance plant nitrogen fixation needs with its carbon budget.

Cytokinin accumulation in root hairs and cortical cells after *Rhizobium* inoculation has been implicated as a key differentiation signal in stimulating root nodule organogenesis in response to Nod factor signaling (Lohar et al., 2004; Ferguson et al., 2010). CK suppresses pericycle cell division for lateral root primordia initiation, promotes cortical cell division for nodule primordia formation, and stimulates the expression of early *NODULIN* (*Nod*) genes (Bauer et al., 1996; Fang and Hirsch, 1998; Svistoonoff et al., 2010), a broad array of genes found to be transcriptionally activated or upregulated during nodulation, many of which are involved in cell wall synthesis (reviewed in Nap and Bisseling, 1990; Frugier et al., 2008). The prominent role of CK presence and/or perception in nodule formation is emphasized by studies showing pseudonodule formation in both legumes and non-legumes due to exogenously applied CK (Arora et al., 1959; Rodriguez-Barrueco and De Castro, 1973; Relic et al., 1993) and a cytokinin-like purine derivative secreted by a *Bradyrhizobium* strain that does not produce Nod factors (Giraud et al., 2007), as well as a gain-of-function mutation in a lotus histidine kinase cytokinin receptor *lhk1* that results in *Rhizobium* and CK-independent, spontaneous root nodule formation (Tirichine et al., 2007). CK receptors implicated in nodule development in *M. truncatula* include MtCRE1 (Gonzalez-Rizzo *et al.* 2006), an ortholog of Arabidopsis Cytokinin Receptor

1/Arabidopsis Histitine Kinase 4 (AHK4) (Yamada *et al.* 2001), and CK response regulators similar to the Arabidopsis CK-response proteins ARR4-5 (Gonzalez-Rizzo et al., 2006) and ARR 10-12 (Lohar et al., 2006). Transcription factors activated downstream of CK-signaling in root cortical cells include NODULATION SIGNALING PATHWAY 1 and 2 (NSP1 and NSP2) (Kalo et al., 2005; Smit et al., 2005) and NODULE INCEPTION (NIN) (Catoira 2000; Borisov *et al.* 2003; Marsh *et al.* 2007). All three of these TFs are essential for nodulation, and may regulate and coordinate nodule development by regulating the expression of downstream *NODULINs*—genes expressed specifically during nodulation (Nap and Bisseling 1990; Verma *et al.* 1992), such as EARLY NODULIN 11(ENOD11), a putative cell wall repetitive hydroxyl-proline-rich protein (Journet et al., 2001; Charron et al., 2004).

In addition to CK, a hormone network including auxin, JA, ABA, GA, SA, brassinosteroids, and ethylene are also tightly regulated during nodule organogenesis (reviewed in Ferguson et al., 2010). Auxin, brassinosteroids, and GA are reported to be positive regulators of nodule formation, while ABA, JA, and ethylene are reported to be negative regulators, possibly by their involvement in plant stress and defense response pathways (reviewed in Ding and Oldroyd, 2009). Several Medicago truncatula ethylene response factors (ERFs) have been found to be associated with Nod factor signal transduction, including the ERF REQUIRED FOR NODULATION (ERN) (Middleton et al., 2007) and ERF REQUIRED FOR NODULE DIFFERENTIATION (EFD) (Vernie et al., 2008). ABA is also thought to modulate the cytokinin response by promoting LR growth, suppressing nodule formation, and inhibiting *Rhizobium* and Nod factor-induced gene expression (Ding et al., 2008). Most studies done on

hormones and nodulation to date have only involved one to two hormone classes, thus a system-wide view of the interactions and effects of the major plant hormones on nodule organogenesis regulation has yet to be assembled.

Phenotyping platforms for further understanding of root architecture traits

High power, high resolution GWAS and sequencing methods have far outpaced phenotyping methods necessary for the discovery of regions and underlying genes involved in plant growth and development (McNally et al., 2006; Yan et al., 2009; Huang et al., 2010; Tung et al., 2010; Zhang et al., 2010). Precise, single-trait elucidation and accurate, efficient measurement are an absolute requirement for the replicated phenotyping of large panels of individuals necessary to resolve trait-genotype associations using GWA. Traditional methods used for root growth and architecture evaluation, such as field excavation, root bagging, plate culture, core sampling, and rhizotrons (reviewed in Shashidar et al., 2012) are poorly suited for the large number of individuals required by GWAS due to a range of issues including low volume and sampling size, poor trait complexity resolution and measurement accuracy, and high labor, time, space, and material costs. However, these traditional approaches provide invaluable information about plant growth and yield under relevant field conditions and can be productively integrated with results from newer phenotyping platforms to provide a strong rationale for prioritizing future research.

A host of new, minimally intrusive, non-destructive, whole-root-system growth systems and imaging platforms have now been developed that should revolutionize our ability to explore the genetic basic of RSA. Of these, hydroponics (Famoso et al., 2010) and gel (Fang et al., 2009; Iyer-Pascuzzi et al., 2010; Clark et al., 2011) growth systems are currently amongst those best suited for RSA trait measurement and analysis for their highly controlled and standardized

rooting environments, ease in whole root system visualization and adaptability for the imposition of environmental stresses and nutrient profiles. Both of these systems involve root growth in a non-natural, liquid or semi-solid rooting environment, however, they can require tailored adjustment for use with different plant species, and are somewhay spatially and thus developmentally limited to relatively simple root systems from small or young plants. X-ray computed tomography (Lontoc-Roy et al., 2006; Perret et al., 2007; Tracy et al., 2010), NMR (Menzel et al., 2007), laser (Braga et al., 2009), ground penetrating radar (GPR) and infrared (IR) and near-infrared (NIR) imaging systems (Dokken and Davis, 2007; Tirlapur and Konig, 1999) are advantageous in their ability to visualize plant root systems grown in soil or solid rooting media, but are currently limited by their small analysis volume and often low resolution and precision, as well as their cost, accessibility, and low-throughput.

With further advancements, NMR, GPR, and IR/NIR technologies have the greatest scale-up potential for the eventual non-destructive imaging and phenotyping of field-grown plant root systems. Although these current root growth systems and imaging technologies are still unable to accurately visualize and quantify complex, mature plant root systems grown under field conditions, they have contributed greatly to increase the precision and efficiency of 2D and 3D spatial and temporal imaging crucial for obtaining information about natural development of RSA in a solid rooting media (reviewed in Danjon and Reubens, 2007; Gregory et al., 2009). Comparative data analysis and integration, especially across controlled environment and field studies is necessary to determine whether QTLs detected by different phenotyping approaches are colocalized along the chromosomes. These regions can be targeted for further investigation to elucidate the genes and molecular mechanisms underlying the trait or phenotype(s) of interest.

The concurrent design of automated or semiautomated image capture systems and software for automated image processing, analysis, and root phenotype quantification (Armengaud et al., 2009; French et al., 2009; Famoso et al., 2010; Clark et al., 2011, 2012) are absolutely essential for simple, precise, and efficient root phenotyping with whole-root system growth platforms. These automated image capture and quantification software systems are also often easily adaptable to an array of low and high-tech growth systems, providing the potential to enhance the throughput and accuracy of root trait measurement from plants grown in a variety of growth systems. Sustained innovation in accurate, efficient, large-scale, high-throughput root growth and analysis systems, especially those tailored toward more the complex and natural soil and field environments will continue to be essential for future studies on the association and linkage mapping of RSA traits.

Understanding the genetic and environmental control of whole system architecture

Recent development of new, non-invasive, controlled, root phenotyping techniques and the ability to accurately visualize and quantify root system architecture paves the way for the further development of higher throughput technologies to assist with linkage and association mapping and mutant analysis. Concurrent advances in the development of informative populations and use of the latest genotyping/sequencing techniques can allow for the faster determination of genes involved in root architectural components and the molecular mechanisms underlying the intrinsic and extrinsic pathways which control root growth and development.

The next step will be to look at this new root phenotypic data in combination with the well-studied above-ground shoot and yield related traits to determine whether any correlations may be made between root architectural traits and plant performance in different environments. Progress

is being made on root-shoot hormone synthesis and signaling pathways (De Kroon *et al.* 2009; Puig *et al.* 2012), but the elucidation and integration of the complexes of molecular and hormonal networks that coordinate the developmental regulation with environmental perception and response remains an intriguing opportunity for the plant biology community and a compelling goal for plant breeders who seek new strategies for enhancing crop performance in the face of water and land shortages in the decades to come.

Supplemental Tables

All supplemental tables cited in this chapter (Supplemental Tables 3.1, 3.2, and 3.3)are available online at: http://dx.doi.org/10.3389/fpls.2013.00186

REFERENCES

Abel, S. (1994) Early auxin-induced genes encode short-lived nuclear proteins. *Proceedings of the National Academy of Sciences* 91:326–330

Ahn, S.J., Sivaguru, M., Osawa, H., Chung, G.C., Matsumoto, H. (2001) Aluminum inhibits the H(+)-ATPase activity by permanently altering the plasma membrane surface potentials in squash roots. *Plant Physiol*. 126:1381-1390

Aida, M., Beis, D., Heidstra, R., Willemsen, V.A., Blilou, I., Reis Galinha, C.I., Nussaume, L., Noh, Y.-S., Amasino, R., Scheres, B.J.G. (2004) The PLETHORA genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* 119, 109-120

Aida, M., Vernoux, T., Furutani, M., Traas, J., Tasaka, M. (2002) Roles of PIN-FORMED1 and MONOPTEROS in pattern formation of the apical region of the *Arabidopsis* embryo. *Developmen*t 129:3965

Akiyama, K., Matsuzaki, K., and Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 435: 824–827

Akiyama, K., Ogasawara, S., Ito, S., and Hayashi, H. (2010) Structural requirements of strigolactones for hyphal branching in AM fungi. *Plant & cell physiology* 51:1104–1117

Aloni, R., Aloni, E., Langhans, M., Ullrich, C.I. (2006) Role of cytokinin and auxin in shaping root architecture: regulating vascular differentiation, lateral root initiation, root apical dominance and root gravitropism. *Ann. Bot.* 97:883-893

Alonso, J.M., Hirayama, T., Roman, G., Nourizadeh, S., Ecker, J.R. (1999) EIN2, a bifunctional transducer of ethylene and stress responses in *Arabidopsis*. *Science* 284:2148-2152

Ané, J., Kiss, G., and Riely, B. (2004) *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. *Science* 303:1364-1367

Apse, M.P., Aharon, G.S., Snedden, W.A., Blumwald, E. (1999) Salt tolerance conferred by overexpression of a vacuolar Na+/H+ antiport in *Arabidopsis*. *Science* 285:1256-1258

Ariel, F., Diet, A., Verdenaud, M., Gruber, V., Frugier, F., Chan, R., Crespi, M. (2010) Environmental regulation of lateral root emergence in *Medicago truncatula* requires the HD-Zip I transcription factor HB1. *Plant Cell* 22:2171-2183

Arite, T., Kameoka, H., and Kyozuka, J. (2012) Strigolactone Positively Controls Crown Root Elongation in Rice. *Journal of Plant Growth Regulation* 31:165–172

Armengaud, P., Zambaux, K., Hills, A., Sulpice, R., Pattison, R.J., Blatt, M.R., Amtmann, A. (2009) EZ-Rhizo: integrated software for the fast and accurate measurement of root system architecture. *Plant J*. 57:945-956

Arora, N., Skoog, F., Allen, O.N. (1959) Kinetin-induced pseudonodules on tobacco roots. *Am. J. Bot*. 46:610-613

Badescu, G.O. and Napier, R.M. (2006) Receptors for auxin: will it all end in TIRs? *Trends Plant Sci*. 11:217-223

Bai, H., Murali, B., Barber, K., and Wolverton, C. (2013) Low phosphate alters lateral root setpoint angle and gravitropism. *American journal of botany* 100:175–182

Band, L., Wells, D., and Larrieu, A. (2012) Root gravitropism is regulated by a transient lateral auxin gradient controlled by a tipping-point mechanism. *Proc. Natl. Acad. Sci.* doi:10.1073/pnas.1201498109

Bao, F., Shen, J., Brady, S. R., Muday, G. K., Asami, T., and Yang, Z. (2004) Brassinosteroids interact with auxin to promote lateral root development in Arabidopsis. *Plant physiology* 134:1624–1631

Bano, A. (2010) Root-to-shoot signal transduction in rice under salt stress. *Pak. J. Bot*. 42:329-339

Bari, R., Pant, B.D., Stitt, M., Scheible, W-R. (2006) PHO2, microRNA399, and PHR1 define a phosphate-signaling pathway in plants. *Plant Physiol*. 141:988-999

Barker, S.J. and Tagu, D. (2000) The roles of auxins and cytokinins in mycorrhizal symbioses. *J. Plant Growth Regul*. 19:144-154

Bates, T.R. and Lynch, J.P. (1996) Stimulation of root hair elongation in *Arabidopsis thaliana* by low phosphorus availability. *Plant, Cell Environ*. 19:529-538

Bates, T.R. and Lynch, J.P. (2000) Plant growth and phosphorus accumulation of wild type and two root hair mutants of *Arabidopsis thaliana* (Brassicaceae). *Am. J. Bot*. 87:958-963

Bauer, P., Ratet, P., Crespi, M.D., Schultze, M., Kondorosi, A. (1996) Nod factors and cytokinins induce similar cortical cell division, amyloplast deposition and MsEnod12A expression patterns in alfalfa roots. *The Plant Journal* 10:91-105

Beeckman, T., Burssens, S., Inze, D. (2001) The peri-cell-cycle in *Arabidopsis*. *J. Exp. Bot*. 52:403-411

Benkova, E. and Hejatko, J. (2009) Hormone interactions at the root apical meristem. *Plant Mol. Biol*. 69:383-396

Bennett, M.J., Marchant, A., Green, H.G., May, S.T., Millner, P.A., Walker, A.R., Schulz, B., Feldmann, K.A. (1996) *Arabidopsis* AUX1 gene: a permease-like regulator of root gravitropism. *Science* 273:948-950

Berta, G., Fusconi, A., Hooker, J.E. (2002) Arbuscular mycorrhizal modifications to plant root systems: scale, mechanisms and consequences. In Mycorrhizal technology in agriculture: from genes to bioproducts. (Gianinazzi, S., et al, eds), pp. 71-85, Birkhäuser Verlag

Berta, G., Fusconi, A., Trotta, A., Scannerini, S. (1990) Morphogenetic modifications induced by the mycorrhizal fungus glomus strain E3 in the root system of *Allium porrum* L. *New Phytol*. 114: 207-215

Berta, G., Trotta, A., Fusconi, A., Hooker, J.E., Munro, M., Atkinson, D., Giovannetti, M., Morini, S., Fortuna, P., Tisserant, B., Gianinazzi-Pearson, V., Gianiazzi, S. (1995) Arbuscular mycorrhizal induced changes to plant growth and root system morphology in *Prunus cerasifera*. *Tree Physiol*. 15:281-293

Bhalerao, R.P., Eklof, J., Ljung, K., Marchant, A., Bennett, M., Sandberg, G. (2002) Shoot-derived auxin is essential for early lateral root emergence in *Arabidopsis* seedlings. *Plant J*. 29:325-332

Bhuja, P., McLachlan, K., Stephens, J., Taylor, G. (2004) Accumulation of 1,3-beta-D-glucans, in response to aluminum and cytosolic calcium in *Triticum aestivum*. *Plant Cell Physiol*. 45:543-549

Bishopp, A., Help, H., and Helariutta, Y. (2009) Cytokinin signaling during root development. *In*: International Review of Cell and Molecular Biology, Elsevier Inc. 276:1–48

Blancaflor, E.B. and Masson, P.H. (2003) Plant gravitropism. Unraveling the ups and downs of a complex process. *Plant Physiol*. 133:1677-1690

Blilou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Friml, J., Heidstra, R., Aida, M., Palme, K., Scheres, B. (2005) The PIN auxin efflux facilitator network controls growth and patterning in *Arabidopsis* roots. *Nature* 433:39-44

Boerjan, W., Cervera, M.T., Delarue, M., Beeckman, T., Dewitte, W., Bellini, C., Caboche, M., Van Onckelen, H., Van Montagu, M., Inze, D. (1995) Superroot, a recessive mutation in *Arabidopsis*, confers auxin overproduction. *Plant Cell* 7:1405-1419

Boniotti, M. B., and Gutierrez, C. (2001) A cell-cycle-regulated kinase activity phosphorylates plant retinoblastoma protein and contains, in *Arabidopsis*, a CDKA/cyclin D complex. *The Plant Journal* 28:341–350

Bonser, A.M., Lynch, J., Snapp, S. (1996) Effect of phosphorus deficiency on growth angle of basal roots in *Phaseolus vulgaris*. *New Phytol*. 132:281-288

Boonsirichai, K., Sedbrook, J.C., Chen, R., Gilroy, S., Masson, P.H. (2003) ALTERED RESPONSE TO GRAVITY is a peripheral membrane protein that modulates gravity-induced cytoplasmic alkalinization and lateral auxin transport in plant statocytes. *Plant Cell* 15:2612-2625

Borisov, A. Y., Madsen, L. H., Tsyganov, V. E., Umehara, Y., Voroshilova, V. A., Batagov, A. O., Sandal, N., Mortensen, A., Schauser, L., Ellis, N., Tikhonovich, I. A., Stougaard, J. (2003) The Sym35 gene required for root nodule development in pea is an ortholog of *Nin* from *Lotus japonicus*. *Plant Physiology* 131:1009–1017

Bouranis, D., Buchner, P., Chorianopoulou, S.N., Hopkins, L., Protonotarios, V.E., Siyiannis, V.F., Hawkesford, M.J. (2008). Responses to sulfur limitation in maize. Sulfur assimilation and abiotic stress in plants. N. Khan, S. Singh and S. Umar, Springer Berlin Heidelberg: pp.1-19

Brady, S. M., Sarkar, S. F., Bonetta, D., and McCourt, P. (2003) The ABSCISIC ACID INSENSITIVE 3 (ABI3) gene is modulated by farnesylation and is involved in auxin signaling and lateral root development in Arabidopsis. *The Plant Journal* 34:67–75

Braga, R.A., Dupuy, L., Pasqual, M., Cardoso, R.R. (2009) Live biospeckle laser imaging of root tissues. *Eur. Biophys*. J. 38:679-686

Brunoud, G., Wells, D. M., Oliva, M., Larrieu, A., Mirabet, V., Burrow, A. H., Beeckman, T., Kepinski, S., Traas, J., Bennett, M. J., Vernoux, T. (2012) A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature* 482:103–106

Buchner, P., Takahashi, H., Hawkesford, M.J. (2004) Plant sulfate transporters: co-ordination of uptake, intracellular and long-distance transport. *J. Exp. Bot*. 55:1765-1773

Budíková, S. (1999) Structural Changes and Aluminium Distribution in Maize Root Tissues. *Biologia Plantarum* 42:259-266

Buer, C.S., Sukumar, P., Muday, G.K. (2006) Ethylene modulates flavonoid accumulation and gravitropic responses in roots of *Arabidopsis*. *Plant Physiol*. 140:1384-1396

Campanoni, P., and Nick, P. (2005) Auxin-dependent cell division and cell elongation. 1-Naphthaleneacetic acid and 2,4-dichlorophenoxyacetic acid activate different pathways. *Plant Physiology* 137:939–948

Caniato, F.F., Guimaraes, C.T., Schaffert, R.E., Alves, V.C.M., Kochian, L.V., Borem, A., Klein, P.E., Magalhaes, J.V. (2007) Genetic diversity for aluminum tolerance in sorghum. Th*eor. Appl. Genet*. 114:863-876

Cao, Y.R., Chen, S-Y., Zhang, J-S. (2008) Ethylene signaling regulates salt stress response: An overview. *Plant. Signal. Behav*. 3:761-763

Carraro, N., Forestan, C., Canova, S., Traas, J., Varotto, S. (2006) ZmPIN1a and ZmPIN1b encode two novel putative candidates for polar auxin transport and plant architecture determination of maize. *Plant Physiol*. 142:254-264

Casimiro, I., Marchant, A., Bhalerao, R.P., Beeckman, T., Dhooge, S., Swarup, R., Graham, N., Inze, D., Sandberg, G., Casero, P.J., Bennett, M. (2001) Auxin transport promotes *Arabidopsis* lateral root initiation. *Plant Cell* 13:843-852

Catoira, R. (2000) Four genes of *Medicago truncatula* controlling components of a nod factor transduction pathway. *The Plant Cell* 12:1647–1666Celenza, J.L., Grisafi, P.L., Fink, G.R. (1995) A pathway for lateral root formation in *Arabidopsis thaliana*. *Genes Dev*. 9:2131-2142

Cervantes, E. (2001) ROS in root gravitropism: the auxin messengers? *Trends Plant Sci*. 6:556-556

Chandler, P.M., Marion-Poll, A., Ellis, M., Gubler, F. (2002) Mutants at the Slender1 locus of barley cv Himalaya. Molecular and physiological characterization. *Plant Physiol*. 129:181-190

Chang, S., Kim, Y., and Lee, J. (2004) Brassinolide interacts with auxin and ethylene in the root gravitropic response of maize (*Zea mays*). *Physiologia Plantarum* 121:666-673

Charron, D., Pingret, J-L., Chabaud, M., Journet, E-P., Barker, D.G. (2004) Pharmacological evidence that multiple phospholipid signaling pathways link Rhizobium nodulation factor perception in *Medicago truncatula* root hairs to intracellular responses, including Ca2+ spiking and specific ENOD gene expression. *Plant Physiol*. 136:3582-3593

Chen, C., DeClerck, G., Tian, F., Spooner, W., McCouch, S., and Buckler, E. (2012) PICARA, an analytical pipeline providing probabilistic inference about *a priori* candidates genes underlying genome-wide association QTL in plants. *PLoS ONE* 7:e46596

Christie, J.M., Reymond, P., Powell, G.K., Bernasconi, P., Raibekas, A.A., Liscum, E., Briggs, W.R. (1998) *Arabidopsis* NPH1: a flavoprotein with the properties of a photoreceptor for phototropism. *Science* 282:1698-1701

Clark, R. T., Famoso, A. N., Zhao, K., Shaff, J. E., Craft, E. J., Bustamante, C. D., McCouch, S. R., Aneshansley, D. J., and Kochian, L. V (2013) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. *Plant, cell & environment* doi: 10.1111/j.1365-3040.2012.02587.x

Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J., and Kochian, L. V (2011). Three-dimensional root phenotyping with a novel imaging and software platform. *Plant physiology* 156:455–465

Correll, M. J., Coveney, K. M., Raines, S. V, Mullen, J. L., Hangarter, R. P., and Kiss, J. Z. (2003) Phytochromes play a role in phototropism and gravitropism in *Arabidopsis* roots.

Advances in Space Research 31:2203–2210

Cosgrove, D.J. (2000) Loosening of plant cell walls by expansins. *Nature* 407:321-326

Coudert, Y., Perin, C., Courtois, B., Khong, N.G., Gantet, P. (2010) Genetic control of root development in rice, the model cereal. *Trends Plant S*ci. 15:219-226

Costigan, S. E., Warnasooriya, S. N., Humphries, B. A., and Montgomery, B. L. (2011) Root-localized phytochrome chromophore synthesis is required for photoregulation of root elongation and impacts root sensitivity to jasmonic acid in *Arabidopsis*. *Plant Physiology* 157:1138–1150

Cui, H., Levesque, M.P., Vernoux, T., Jung, J.W., Paquette, A.J., Gallagher, K.L., Wang., J.Y., Blilou, I., Scheres, B., Benfey P.N. (2007) An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* 316: 421-425

Dan, H., Yang, G., Zheng, A-L. (2007) A negative regulatory role for auxin in sulfate deficiency response in *Arabidopsis thaliana*. *Plant Mol. Biol*. 63:221-235

Deak, K. I., and Malamy, J. (2005) Osmotic regulation of root system architecture. *The Plant Journal for Cell and Molecular Biology* 43:17–28

Danjon, F., and Reubens, B. (2007) Assessing and analyzing 3D architecture of woody root systems, a review of methods and applications in tree and soil stability, resource acquisition and allocation. *Plant and Soil* 303:1–34

De Kroon, H., Visser, E. J. W., Huber, H., Mommer, L., and Hutchings, M. J. (2009) A modular concept of plant foraging behaviour: the interplay between local responses and systemic control. *Plant, Cell & Environment* 32:704–712

De Rybel, B., Vassileva, V., Grunewald, W., Naudts, M., Levesque, M.P., Ehrismann, J.S., Inze, D., Luschnig, C., Benfey, P.N., Weijers, D., Van Montagu, M.C.E., Bennett, M.J., Jurgens, G., Beeckman, T. (2010) Bimodular auxin response controls organogenesis in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the U S A* 107:2705–2710

De Smet, I., Signora, L., Beeckman, T., Inze, D., Foyer, C., Zhang, H. (2003) An abscisic acid-sensitive checkpoint in lateral root development of *Arabidopsis*. *The Plant Journal* 33:543-555

De Smet, I., Zhang, H., Inzé, D., and Beeckman, T. (2006) A novel role for abscisic acid emerges from underground. *Trends in Plant Science* 11:434–439

De Smet, I., Tetsumura, T., De Rybel, B., Frei dit Frey, N., Laplaze, L., Casimiro, I., Swarup, R., Naudts, M., Vanneste, S., Audenaert, D., Inze, D., Bennett, M.J., Beeckman, T. (2007) Auxin-dependent regulation of lateral root positioning in the basal meristem of *Arabidopsis*. *Development* 134:681-690

De Smet, I., Lau, S., Voss, U., Vanneste, S., Benjamins, R., Rademacher, E. H., Schlereth, A., De Rybel, B., Vassileva, V., Grunewald, W., Naudts, M., Levesque, M.P., Ehrismann, J.S., Inze, D., Luschnig, C., Benfey, P.N., Weijers, D., Van Montagu, M.C.E., Bennett, M.J., Jurgens, G., Beeckman, T. (2010) Bimodular auxin response controls organogenesis in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the U S A* 107:2705–2710

Delhaize, E., Ryan, P.R., Randall, P.J. (1993) Aluminum tolerance in wheat (*Triticum aestivum* L.) (II. Aluminum-stimulated excretion of malic acid from root apices). *Plant Physiol*. 103:695-702

Dello Ioio, R., Linhares, F.S., Sabatini, S. (2008) Emerging role of cytokinin as a regulator of cellular differentiation. *Curr. Opin. Plant Biol*. 11:23-27

Demidchik, V., Cuin, T.A., Svistunenko, D., Smith, S.J., Miller, A.J., Shabala, S., Sokolik, A., Yurin, V. (2010) *Arabidopsis* root K+-efflux conductance activated by hydroxyl radicals: single-channel properties, genetic basis and involvement in stress-induced cell death. *J. Cell. Sci*. 123:1468-1479

Devaiah, B.N. and Raghothama, K.G. (2007) Transcriptional regulation of Pi starvation responses by WRKY75. *Plant. Signal. Behav*. 2:424-425

Devaiah, B.N., Karthikeya, A.S., Raghothama, K.G. (2007) WRKY75 transcription factor is a modulator of phosphate acquisition and root development in *Arabidopsis*. *Plant Physiol*. 143:1789-1801

Dharmasiri, N., Dharmasiri, S., Estelle, M. (2005) The F-box protein TIR1 is an auxin receptor. *Nature* 435:441-445

Di Laurenzio, L., Wysocka-Diller, J., Malamy, J.E., Pysh, L., Helariutta, Y., Freshour, G., Hahn, M.G., Feldmann, K.A., Benfey, P.N. (1996) The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* 86:423-433

DiDonato, R.J., Arbuckle, E., Buker, S., Sheets, J., Tobar, J., Totong, T., Grisafi, P., Fink, G.R., Celenza, J.L. (2004) *Arabidopsis* ALF4 encodes a nuclear-localized protein required for lateral root formation. *Plant J.* 37:340-353

Digby, J. and Firn, R.D. (1995) The gravitropic set-point angle (GSA): the identification of an important developmentally controlled variable governing plant architecture. *Plant. Cell. Environ*. 18:1434-1440

Digby, J. and Firn, R.D. (2002) Light modulation of the gravitropic set-point angle (GSA). J. *Exp. Bot*. 53:377-381

Dill, A. and Sun, T. (2001) Synergistic derepression of gibberellin signaling by removing RGA

and GAI function in *Arabidopsis thaliana*. *Genetics* 159:777-785

Dimkpa, C., Weinand, T., Asch, F. (2009) Plant-rhizobacteria interactions alleviate abiotic stress conditions. *Plant Cell and Environment* 32:1682-1694

Ding, Y. and Oldroyd, G.E. (2009) Positioning the nodule, the hormone dictum. *Plant. Signal. Behav*. 4:89-93

Ding, Y., Kalo, P., Yendrek, C., Sun, J., Liang, Y., Marsh, J.F., Harris, J.M., Oldroyd, G.E.D. (2008) Abscisic acid coordinates Nod factor and cytokinin aignaling during the regulation of nodulation in *Medicago truncatula*. Plant Cell 20:2681-2695

Dixon, R.K. (1990) Cytokinin activity in citrus jambhiri seedlings colonized by mycorrhizal fungi. *Agric. Ecosyst. Environ.* 29:103-106

Dokken, K.M. and Davis, L.C. (2007) Infrared imaging of sunflower and maize root anatomy. J. *Agric. Food Chem*. 55:10517-10530

Edelmann, H.G. and Roth, U. (2006) Gravitropic plant growth regulation and ethylene: an unsought cardinal coordinate for a disused model. *Protoplasma* 229:183-191

Endre, G., Kereszt, A., Kevei, Z., Mihacea, S., Kaló, P., and Kiss, G. B. (2002) A receptor kinase gene regulating symbiotic nodule development. *Nature* 417:962–966

Eticha, D., Stass, A., Horst, W.J. (2005) Cell-wall pectin and its degree of methylation in the maize root-apex: significance for genotypic differences in aluminium resistance. *Plant Cell Environ*. 28:1410-1420

Evelin, H., Kapoor, R., Giri, B. (2009) Arbuscular mycorrhizal fungi in alleviation of salt stress: a review. *Ann. Bot.* 104:1263-1280

Fahn, A. (1990) Plant Anatomy, Fourth Edition , Pergamon

Famoso, A.N., Clark, R.T., Shaff, J.E., Craft, E., McCouch, S.R., Kochian, L.V. (2010) Development of a novel aluminum tolerance phenotyping platform used for comparisons of cereal aluminum tolerance and investigations into rice aluminum tolerance mechanisms. *Plant Physiol*. 153:1678-1691

Fang, S., Yan, X., Liao, H. (2009) 3D reconstruction and dynamic modeling of root architecture in situ and its application to crop phosphorus research. *Plant J*. 60:1096-1108

Fang, Y. and Hirsch, A.M. (1998) Studying early nodulin gene ENOD40 expression and induction by nodulation factor and cytokinin in transgenic alfalfa. *Plant Physiol*. 116:53-68

Farrás, R., Ferrando, A., Jásik, J., Kleinow, T., Okrész, L., Tiburcio, A., Salchert, K., Del Pozo,

C., Schell, J., and Koncz, C. (2001) SKP1-SnRK protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *The EMBO Journal* 20:2742–2756

Ferguson, B.J., Indrasumunar, A., Hayashi, S., Lin, M-H., Lin, Y-H., Reid, D.E., Gresshoff, P.M. (2010) Molecular analysis of legume nodule development and autoregulation. *J. Integr. Plant. Biol.* 52:61-76

Fiers, M., Hause, G., Boutilier, K., Casamitjana-Martinez, E., Weijers, D., Offringa, R., van der Geest, L., van Lookeren Campagne, M., Liu, C-M. (2004) Mis-expression of the CLV3/ESR-like gene CLE19 in *Arabidopsis* leads to a consumption of root meristem. *Gene* 327:37-49

Fiers, M., Golemiec, E., Van der Schors, R., Van der Geest, L., Li, K. W., Stiekema, W. J., and Liu, C.-M. (2006) The CLAVATA3/ESR motif of CLAVATA3 is functionally independent from the nonconserved flanking sequences. *Plant Physiology* 141:1284–1292

Fitter, A.H. (1991) The ecological significance of root system architecture: an economic approach. In Plant Root Growth: an Ecological Perspective (Atkinson, D., ed), pp. 229-243, Blackwell Scientific Publications

Fitze, D., Wiepning, A., Kaldorf, M., Lidwig-Muller, J. (2005) Auxins in the development of an arbuscular mycorrhizal symbiosis in maize. *J. Plant Physiol*. 162:1210-1219

Fleet, C.M. and Sun, T. (2005) A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Curr. Opin. Plant Biol*. 8:77-85

Flowers, T.J. and Yeo, A.R. (1995) Breeding for salinity resistance in crop plants: Where next? *Aust. J. Plant Physiol*. 22:875-884

Foo, E., Yoneyama, K., Hugill, C. J., Quittenden, L. J., and Reid, J. B. (2012) Strigolactones and the regulation of pea symbioses in response to nitrate and phosphate deficiency. *Molecular Plant* 6:76–87

Foy, C.D. (1984) Adaptation of plants to mineral stress in problem soils. Ciba found. *Symp*. 102:20-39

Franco-Zorrilla, J.M., Martin, A.C., Solano, R., Rubio, V., Leyva, A., Paz-Ares, J. (2002) Mutations at CRE1 impair cytokinin-induced repression of phosphate starvation responses in *Arabidopsis*. *Plant J*. 32:353-360

French, A., Ubeda-Tomas, S., Holman, T.J., Bennett, M.J., Pridmore, T. (2009) High-throughput quantification of root growth using a novel image-analysis tool. *Plant Physiol*. 150:1784-1795

Frigerio, M., Alabadi, D., Perez-Gomez, J., Garica-Carcel, L., Phillips, A.L., Hedden, P., Blazquez, M.A. (2006) Transcriptional regulation of gibberellin metabolism genes by auxin signaling in *Arabidopsis*. *Plant Physiol*. 142:553-563

Frugier, F., Kosuta, S., Murray, J.D., Crespi, M., Szczyglowshi, K. (2008) Cytokinin: secret agent of symbiosis. *Trends Plant Sci.* 13:115-120

Fu, X., and Harberd, N. P. (2003) Auxin promotes Arabidopsis root growth by modulating gibberellin response. *Nature* 421:740–743

Fukaki, H. and Tasaka, M. (2009) Hormone interactions during lateral root formation. *Plant Mol. Biol*. 69:437-449

Fukao, T., Yeung, E., and Bailey-Serres, J. (2011) The submergence tolerance regulator SUB1A mediates crosstalk between submergence and drought tolerance in rice. *The Plant cell* 23:412–427

Fukaki, H., Okushima, Y., and Tasaka, M. (2007) Auxin-mediated lateral root formation in higher plants. *International Review of Cytology* 256:111–137

Gagne, J. M., Downes, B. P., Shiu, S.-H., Durski, A. M., and Vierstra, R. D. (2002) The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proceedings of the National Academy of Sciences* 99:11519–11524

Galen, C., Rabenold, J.J., Liscum, E. (2007) Functional ecology of a blue light photoreceptor: effects of phototropin-1 on root growth enhance drought tolerance in *Arabidopsis thaliana*. *New Phytol*. 173:91-99

Gao, Q., Zhao, M., Li, F., Guo, Q., Xing, S., Wang, W. (2008) Expansins and coleoptile elongation in wheat. *Protoplasma* 233:73-81

Gaxiola, R.A., Rao, R., Sherman, A., Grisafi, P., Alper, S.L., Fink. G.R. (1999) The *Arabidopsis thaliana* proton transporters, AtNhx1 and Avp1, can function in cation detoxification in yeast. *Proc. Natl. Acad. Sci. U. S. A.* 96:1480-1485

Geldner, N., Anders, N., Wolters, H., Keicher, J., Kornberger, W., Muller, P., Delbarre, A., Ueda, T., Nakano, A., Jurgens, G.(2003) The *Arabidopsis* GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell* 112:219-230

Gil, P., Dewey, E., Friml, J., Zhao, Y., Snowden, K. C., Putterill, J., Palme, K., Estelle, M., and Chory, J. (2001) BIG: a calossin-like protein required for polar auxin transport in *Arabidopsis*. *Genes & development* 15:1985–1997

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J-C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W.L., Chang, W-S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Verméglio, A., Médigue, C., Sadowsky, M. (2007) Legumes symbioses: Absence of nod genes in photosynthetic bradyrhizobia. *Science* 316:1307-1312

Goh, T., Joi, S., Mimura, T., and Fukaki, H. (2012a) The establishment of asymmetry in Arabidopsis lateral root founder cells is regulated by LBD16/ASL18 and related LBD/ASL proteins. *Development (Cambridge, England)* 139, 883–893

Goh, T., Kasahara, H., Mimura, T., Kamiya, Y., and Fukaki, H. (2012b) Multiple AUX/IAA-ARF modules regulate lateral root formation: the role of *Arabidopsis* SHY2/IAA3-mediated auxin signalling. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 367:1461–1468

Gonzalez-Rizzo, S., Crespi, M., and Frugier, F. (2006) The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. *The Plant Cell* 18:2680-2693

Gou, J., Strauss, S. H., Tsai, C. J., Fang, K., Chen, Y., Jiang, X., and Busov, V. B. (2010) Gibberellins regulate lateral root formation in Populus through interactions with auxin and other hormones. *The Plant Cell* 22:623–639

Gowda, V. R. P. V., Henry, A., Yamauchi, A., Shashidhar, H. E., and Serraj, R. (2011) Root biology and genetic improvement for drought avoidance in rice. *Field Crops Research* 122:1–13

Gray, W.M., Kepinski, S., Rouse, D., Leyser, O., Estelle, M. (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* 414:271-276

Gray, W. M., Hellmann, H., Dharmasiri, S., and Estelle, M. (2002) Role of the *Arabidopsis* RING-H2 protein RBX1 in RUB modification and SCF function. *The Plant Cell* 14:2137–2144

Gregory, P.J., Bengough, A.G., Grinev, D., Schmidt, S., Thomas, W.T.B., Wojciechowski, T., Young, I.M. (2009) Root phenomics of crops: opportunities and challenges. *Funct. Plant Biol*. 36:922-929

Groth, P., Kalev, I., Kirov, I., Traikov, B., Leser, U., Weiss, B. (2010) Phenoclustering: online mining of cross-species phenotypes. *Bioinformatics* 26:1924–1925

Guo, H.-S., Xie, Q., Fei, J.-F., and Chua, N.-H. (2005) MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *arabidopsis* lateral root development. *The Plant cell* 17:1376–1386

Guo, F.Q., Wang, R., Crawford, N.M. (2002) The *Arabidopsis* dual-affinity nitrate transporter gene AtNRT1.1 (CHL1) is regulated by auxin in both shoots and roots. *J. Exp. Bot.* 53:835-844

Gutjahr, C., Casieri, L., Paszkowski, U. (2009) Glomus intraradices induces changes in root system architecture of rice independently of common symbiosis signaling. *New Phytol*. 184:829-837

Hardtke, C.S., Ckurshumova, W., Vidaurre, D.P., Singh, S.A., Stamatiou, G., Tiwari, S.b.,

Hagen, G., Guilfoyle, T.J., Berleth, T. (2004) Overlapping and non-redundant functions of the *Arabidopsis* auxin response factors MONOPTEROS and NONPHOTOTROPIC HYPOCOTYL 4. *Development* 131:1089-1100

Harrison, B. and Masson, P.H. (2008a) ARG1 and ARL2 form an actin-based gravity-signaling chaperone complex in root statocytes? *Plant. Signal. Behav*. 3:650-653

Harrison, B.R. and Masson, P.H. (2008b) ARL2, ARG1 and PIN3 define a gravity signal transduction pathway in root statocytes. *Plant J*. 53:380-392

He, X.J., Mu, R-L., Cao, W-H., Zhang, Z-G., Zhang, J-S., Chen, S-Y. (2005) AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. *Plant J*. 44:903-916

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M-T., Benfey, P.N. (2000) The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* 101:555-567

Hellmann, H., Hobbie, L., Chapman, A., Dharmasiri, S., Dharmasiri, N., Del Pozo, C., Reinhardt, D., and Estelle, M. (2003) *Arabidopsis* AXR6 encodes CUL1 implicating SCF E3 ligases in auxin regulation of embryogenesis. *The EMBO Journal* 22:3314–3325

Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol*. 7:637-644

Herrera-Medina, M.J., Steinkellner, S., Vierheilig, H., Bote, J.A.O., Garrido, J.M.G. (2007) Abscisic acid determines arbuscule development and functionality in the tomato arbuscular mycorrhiza. *New Phytol*. 175:554-564

Hetrick, B.A.D. (1991) Mycorrhizas and root architecture. *Cellular and Molecular Life Sciences* 47:355-362

Hetrick, B.A.D., Leslie, J.F., Thompson Wilson, G., Gerschefske Kitt, D. (1988) Physical and topological assessment of effects of a vesicular-arbuscular mycorrhizal fungus on root architecture of big bluestem. *New Phytol*. 110:85-96

Hetz, W., Hochholdinger, F., Schwall, M., Feix, G. (1996) Isolation and characterization of rtcs, a maize mutant deficient in the formation of nodal roots. *The Plant Journal* 10:845-857

Himanen, K., Boucheron, E., Vanneste, S., De Almeida Engler, J., Inze, D., Beechmann, T. (2002) Auxin-mediated cell cycle activation during early lateral root initiation. *Plant Cell* 14: 2339-2351

Hirai, M.Y., Fujiwara, T., Awazuhara, M., Kimura, T., Noji, M., Siato, K.(2003) Global expression profiling of sulfur-starved *Arabidopsis* by DNA macroarray reveals the role of O-

acetyl-l-serine as a general regulator of gene expression in response to sulfur nutrition. *Plant J.* 33:651-663

Hirose, N., Makita, N., Kojima, M., Kamada-Nobusada, T., and Sakakibara, H. (2007) Overexpression of a type-A response regulator alters rice morphology and cytokinin metabolism. *Plant & Cell Physiology* 48:523–539

Hochholdinger, F. (2009) The maize root system: Morphology, anatomy and genetics. In Handbook of maize: Its biology. (Bennetzen, J.L., and Hake, S.C. eds) *Springer, NY* pps.145–160 10.1007/978-0-387-79418-1_8

Hochholdinger, F. and Tuberosa, R. (2009) Genetic and genomic dissection of maize root development and architecture. *Curr Opin Plant Biol* 12:172-177

Hochholdinger, F. and Zimmermann, R. (2008) Conserved and diverse mechanisms in root development. *Curr. Opin. Plant Biol*. 11:70-74

Hochholdinger, F., Park, W. J., Sauer, M., and Woll, K. (2004) From weeds to crops: genetic analysis of root development in cereals. *Trends in Plant Science* 9:42–48

Hochholdinger, F., Park, W.J., Feix, F.H. (2001) Cooperative action of SLR1 and SLR2 is required for lateral root-specific cell elongation in maize. *Plant Physiol*. 125:1529-1539development. *The Plant cell* 17:1376–1386

Hochholdinger, F. and Feix, G. (1998) Early post-embryonic root formation is specifically affected in the maize mutant Irt1. *Plant J*. 16, 247-255

Hoefgen, R. and Nikiforova, V.J. (2008) Metabolomics integrated with transcriptomics: assessing systems response to sulfur-deficiency stress. *Physiol. Plant*. 132:190-198

Hogg, B. V, Cullimore, J. V, Ranjeva, R., and Bono, J.-J. (2006) The DMI1 and DMI2 early symbiotic genes of *medicago truncatula* are required for a high-affinity nodulation factor-binding site associated to a particulate fraction of roots. *Plant Physiology* 140:365–373

Horie, T., Costa, A., Kim, T.H., Han, M.J., Horie, R., Leung, H-Y., Miyao, A., Hirochika, H., An, G., Schroeder, J.I. (2007) Rice OsHKT2;1 transporter mediates large Na+ influx component into K+-starved roots for growth. *EMBO J*. 26:3003-3014

Horie, T., Hauser, F., Schroeder, J.I. (2009) HKT transporter-mediated salinity resistance mechanisms in *Arabidopsis* and monocot crop plants. *Trends Plant Sci*. 14:660-668

Horst, W.J., Wang, Y., Eticha, D. (2010) The role of the root apoplast in aluminium-induced inhibition of root elongation and in aluminium resistance of plants: a review. *Ann. Bot.* 106:185-197

Horváth, B., Yeun, L. H., Domonkos, A., Halász, G., Gobbato, E., Ayaydin, F., Miró, K., Hirsch, S., Sun, J., Tadege, M., Ratet, P., Mysore, K.S., Ane, J-M., Oldroyd, G.E.D., Kalo, P. (2011) *Medicago truncatula IPD3* is a member of the common symbiotic signaling pathway required for rhizobial and mycorrhizal symbioses. *Molecular Plant-microbe Interactions* 24:1345–1358

Huala, E., Oeller, P.W., Liscum, E., Han, I-S., Larsen, E., Briggs, W.R. (1997) *Arabidopsis* NPH1: a protein kinase with a putative redox-sensing domain. *Science* 278:2120-2123

Huang, C.F., Yamaji, N., Mitani, N., Yano, M., Nagamura, Y., Ma, F. (2009) A bacterial-type ABC transporter is involved in aluminum tolerance in rice. *Plant Cell* 21:655-667

Huang, J., Xia, H., Li, Z., Xiong, Y., Kong, G. (2009) Soil aluminum uptake and accumulation by *Paspalum notatum*. *Waste Manag. Res*. 27:668-675

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Ahang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., Liu, T., Zhou, T., Jing, Y., Li, W., Lin, Z., Buckler, E.S., Qian, Q., Zhang, G-F., Li, J., Han, B. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet*. 42:961-967

Ikeda, A., Sonoda, Y., Vernieri, P., Perata, P., Hirochika. H., Yamaguchi, J. (2002) The slender rice mutant, with constitutively activated gibberellin signal transduction, has enhanced capacity for abscisic acid level. *Plant Cell Physiol*. 43:974-979

Ikeda, A., Ueguchi-Tanaka, M., Sonoda, Y., Kitano, H., Koshioka, M., Futsuhara, Y., Matsuoka, M., Yamaguchi, J. (2001) Slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. *Plant Cell* 13:999-1010

Imaizumi-Anraku, H., Takeda, N., and Charpentier, M. (2004) Plastid proteins crucial for symbiotic fungal and bacterial entry into plant roots. *Nature* 433:527-531

Inukai, Y., Miwa, M., Nagato, Y., Kitano, H., Yamauchi, A. (2001) Characterization of rice mutants deficient in the formation of crown roots. *Breed. Sci*. 51:123-129

Inukai, Y., Sakamoto, T., Ueguchi-Tanaka, M., Shibata, Y., Gomi, K., Umemura, I., Hasegawa, Y., Ashikari, M., Kitano, H., Matsuoka, M. (2005) Crown rootless1, which is essential for crown root formation in rice, is a target of an AUXIN RESPONSE FACTOR in auxin signaling. *Plant Cell* 17, 1387-1396

Itoh, J., Nonomura, K-I, Ikeda, K., Yamaki, S., Inukai, Y., Yamagushi, H., Kitano, H., Nagato, Y. (2005) Rice plant development: from zygote to spikelet. *Plant Cell Physiol*. 46:23-47

Iwama, A., Yamashino, T., Tanaka, Y., Sakakibara, H., Kakimoto, T., Sato, S., Kato, T., Tabata, S., Nagatani, A., Mizuno, T. (2007) AHK5 histidine kinase regulates root elongation through an ETR1-dependent abscisic acid and ethylene signaling pathway in *Arabidopsis thaliana*. *Plant Cell Physiol*. 48:375-380

Iyer-Pascuzzi, A.S., Symonova, O., Mileyko, Y., Hao, Y., Belcher, H., Harer, J., Weitz, J.S., Benfey, P.N. (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiol*. 152:1148-1157

Jain, M., Tyagi, A. K., and Khurana, J. P. (2006) Molecular characterization and differential expression of cytokinin-responsive type-A response regulators in rice (*Oryza sativa*). *Plant Biology* 6:1

Jebanathirajah, J. A., Peri, S., and Pandey, A. (2002) Toll and interleukin-1 receptor (TIR) domain-containing proteins in plants: a genomic perspective. *Trends in Plant Science* 7:388–391

Jiang, C., Gao, X., Liao, L., Harberd, N.P., Fu, X. (2007) Phosphate starvation root architecture and anthocyanin accumulation responses are modulated by the gibberellin-DELLA signaling pathway in *Arabidopsis*. *Plant Physiol*. 145:1460-1470

Jones, D.L., Blancaflor, E.B., Kochian, L.V., Gilroy, S. (2006) Spatial coordination of aluminium uptake, production of reactive oxygen species, callose production and wall rigidification in maize roots. *Plant Cel. Environ*. 29:1309-1318

Joo, J.H., Bae, Y.S., Lee, J.S. (2001) Role of auxin-induced reactive oxygen species in root gravitropism. *Plant Physiol*. 126:1055

Journet, E.P., El-Gachtouli, N., Vernoud, V., de Billy, F., Pichon, M., Dedieu, A., Morandi, D., Barker, D.G., Gianinazzi-Pearson, V. (2001) *Medicago truncatula* ENOD11: a novel RPRP-encoding early nodulin gene expressed during mycorrhization in arbuscule-containing cells. *Mol. Plant Microbe Interac*t. 14:737-748

Kalo, P., Gleason, C., Edwards, A., Marsh, J.M., Mitra, R.M., Hirsch, S., Jakab, J., Sims, S., Long, S.R., Rogers, J., Kiss, G.B., Downie, J.A., Oldroyd, G.E.D. (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science* 308:1786-1789

Kanamori, N., Madsen, L.H., Radutoiu, S., Frantescu, M., Quistgaard, E. M.H., Miwa, H., Downie, J.A., James, E.K., Felle, H.H., Haaning, L.L., Jensen, T.H., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., Stougaard, J. (2006) A nucleoporin is required for induction of Ca2+ spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences* 103:359–364

Kapulnik, Y., Delaux, P. P.-M., Resnick, N., Mayzlish-Gati, E., Wininger, S., Bhattacharya, C., Séjalon-Delmas, N., Combier, J.-P., Bécard, G., Belausov, E., Beeckman, T., Dor, E., Hershenhorn, J., Koltai, H. (2011) Strigolactones affect lateral root formation and root-hair

elongation in Arabidopsis. *Planta* 233:209–216

Kartal, G., Temel, A., Arican, E., Gozukirmizi, N. (2009) Effects of brassinosteroids on barley root growth, antioxidant system and cell division. *Plant Growth Regulation* 58:261-267. DOI: 10.1007/s10725-009-9374-z

Kepinski, S. and Leyser, O. (2005) The *Arabidopsis* F-box protein TIR1 is an auxin receptor. *Nature* 435:446-451

Khadri, M., Tejera, N.A., Lluch, C. (2006) Alleviation of salt stress in common bean (*Phaseolus vulgaris*) by exogenous abscisic acid supply. *Journal of Plant Growth Regulation* 25:110-119

Khan, G., Declerck, M., and Sorin, C. (2011) MicroRNAs as regulators of root development and architecture. *Plant MolecularBio* 77:47-58

Kim, T.-W., Lee, S. M., Joo, S.-H., Yun, H. S., Lee, Y., Kaufman, P. B., Kirakosyan, A., Kim, S.-H., Nam, K. H., Lee, J. S., Chang, S. C., Kim, S-K. (2007) Elongation and gravitropic responses of Arabidopsis roots are regulated by brassinolide and IAA. *Plant, Cell & Environment* 30:679–689

Kinoshita, A., Nakamura, Y., Sasaki, E., Kyozuka, J., Fukuda, H., and Sawa, S. (2007) Gain-of-function phenotypes of chemically synthetic CLAVATA3/ESR-related (CLE) peptides in *Arabidopsis thaliana* and *Oryza sativa*. *Plant and Cell Physiology* 48:1821–1825

Kiss, J.Z., Mullen, J.L., Correll, M.J., Hangarter, R.P. (2003) Phytochromes A and B mediate red-light-induced positive phototropism in roots. *Plant Physiol*. 131:1411-1417

Kistner, C., Winzer, T., Pitzschke, A., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Webb, K.J., Szczyglowski, K., Parniske, M. (2005) Seven Lotus japonicus genes required for transcriptional reprogramming of the root during fungal and bacterial symbiosis. *Plant Cell* 17:2217-2229

Klug, B. and Horst, W.J. (2010) Spatial characteristics of aluminum uptake and translocation in roots of buckwheat (*Fagopyrum esculentum*). *Physiol. Plant*. 139:181-191

Kochian, L.V. (1995) Cellular Mechanisms of Aluminum Toxicity and Resistance in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 46:237

Kohlen, W., Charnikhova, T., and Liu, Q. (2011) Strigolactones are transported through the xylem and play a key role in shoot architectural response to phosphate deficiency in nonarbuscular mycorrhizal host. *Plant Physiology* 155:974-987

Kollmeier, M., Felle, H.H., Horst, W.J. (2000) Genotypical differences in aluminum resistance of maize are expressed in the distal part of the transition zone. Is reduced basipetal auxin flow involved in inhibition of root elongation by aluminum? *Plant Physiol*. 122:945-956

Kopittke, P., Blamey, F.P.C., Menzies, N.W. (2008) Toxicities of soluble Al, Cu, and La include ruptures to rhizodermal and root cortical cells of cowpea. *Plant and Soil* 303:217-227

Kopittke, P. M., Blamey, F. P. C., and Menzies, N. W. (2007) Toxicities of soluble Al, Cu, and La include ruptures to rhizodermal and root cortical cells of cowpea. *Plant and Soil* 303:217–227

Koltai, H., Dor, E., Hershenhorn, J., Joel, D. M., Weininger, S., Lekalla, S., Shealtiel, H., Bhattacharya, C., Eliahu, E., Resnick, N., Barg, R., Kapulnik, Y. (2010) Strigolactones' effect on root growth and root-hair elongation may be mediated by auxin-efflux carriers. *Journal of Plant Growth Regulation* 29:129–136

Kosuta, S., Hazledine, S., Sun, J., Miwa, H., Morris, R. J., Downie, J. A., and Oldroyd, G. E. D. (2008) Differential and chaotic calcium signatures in the symbiosis signaling pathway of legumes. *Proceedings of the National Academy of Sciences* 105:9823–9828

Kosuta, S., Chabaud, M., Lougnon, G., Cough, C., Denarie, J., Barker, D.G., Becard, G. (2003) A diffusible factor from arbuscular mycorrhizal fungi induces symbiosis-specific MtENOD11 expression in roots of *Medicago truncatula*. *Plant Physiol*. 131:952-962

Kramer, E.M. (2004) PIN and AUX/LAX proteins: their role in auxin accumulation. Trends *Plant Sci*. 9:578-582

Krouk, G., Lacombe, B., Bielach, A., Perrine-Walker, F., Malinska, K., Mounier, E., Hoyerova, K., Tillard, P., Leon, S., Ljung, K., Zazimalova, E., Benkova, E., Nacry, P., Gojon, A. (2010) Nitrate-regulated Auxin transport by NRT1.1 defines a mechanism for nutrient sensing in plants. *Developmental Cell* 18:927-937

Kuderová, A., Urbánková, I., Válková, M., Malbeck, J., Brzobohaty, B., Némethová, D., and Hejátko, J. (2008) Effects of conditional IPT-dependent cytokinin overproduction on root architecture of Arabidopsis seedlings. *Plant & Cell Physiology* 49:570–582

Kuiper, D., Schuit, J., Kuiper, P.J.C. (1990) Actual cytokinin concentrations in plant tissue as an indicator for salt resistance in cereals. *Plant and Soil* 123:243-250

Kurata, T., and Yamamoto, K. T. (1997) Light-stimulated root elongation in *Arabidopsis thaliana*. *Journal of Plant Physiology* 151:346–351

Kurth, E., Cramer, G.R., Lauchli, A., Epstein, E. (1986) Effects of NaCl and CaCl(2) on cell enlargement and cell production in cotton roots. *Plant Physiol*. 82:1102-1106

Kutz, A., Muller, A., Hennig, P., Kaiser, W.M., Piotrowski, M., Weiler, E.W. (2002) A role for nitrilase 3 in the regulation of root morphology in sulphur-starving *Arabidopsis thaliana*. *Plant J*. 30:95-106

Kwon, H.-K., Yokoyama, R., and Nishitani, K. (2005) A proteomic approach to apoplastic proteins involved in cell wall regeneration in protoplasts of *Arabidopsis* suspension-cultured cells. *Plant & Cell Physiology* 46:843–857

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., et al. (2012) The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 40:D1202–1210

Laplaze, L., Benkova, E., Casimiro, I., Maes, L., Vanneste, S., Swarup, R., Weijers, D., Calvo, V., Parizot, B., Herrera-Rodriguez, M.B., Offringa, R., Graham, N., Doumas, P., Friml, J., Bogusz, D., Beeckmann, T., Bennett, M. (2007) Cytokinins act directly on lateral root founder cells to inhibit root initiation. *Plant Cell* 19:3889-3900

Larsen, P.B., Cancel, J., Rounds, M., Ochoa, V. (2007) *Arabidopsis* ALS1 encodes a root tip and stele localized half type ABC transporter required for root growth in an aluminum toxic environment. *Planta* 225:1447-1458

Laskowski, M., Biller, S., Stanley, K., Kajstura, T., Prusty R. (2006) Expression profiling of auxin-treated *Arabidopsis* roots: toward a molecular analysis of lateral root emergence. *Plant Cell Physiol*. 47: 788-792

Laskowski, M.J., Williams, M.E., Nusbaum, H.C., Sussex, I.E. (1995) Formation of lateral root meristems is a two-stage process. *Development* 121:3303-3310

Lawrence, C., and Harper, L. (2008) MaizeGDB: the maize model organism database for basic, translational, and applied research. *International Journal of Plant Genomics* doi:10.1155/2008/496957

Laxmi, A., Pan, J., Morsy, M., Chen, R. (2008) Light plays an essential role in intracellular distribution of auxin efflux carrier PIN2 in *Arabidopsis thaliana*. *PLoS One* 3:e1510

Lee, H.W., Kim, N.Y., Lee, D.J., Kim, J. (2009) LBD18/ASL20 regulates lateral root formation in combination with LBD16/ASL18 downstream of ARF7 and ARF19 in *Arabidopsis*. *Plant Physiol.* 151:1377-1389

Lee, Y. and Kende, H. (2002) Expression of alpha -expansin and expansin-like genes in deepwater rice. *Plant Physiol*. 130:1396-1405

Lelandais-Briere, C., Jovanovic, M., Torres, G.M., Perrin, Y., Lemoine, R., Corre-Menguy, F., Hartmann, C. (2007) Disruption of AtOCT1, an organic cation transporter gene, affects root development and carnitine-related responses in *Arabidopsis*. *Plant J*. 51:154-164

Leng, Q., Mercier, R.W., Hua, B-G., Fromm, H., Berkowitz, G.A. (2002) Electrophysiological analysis of cloned cyclic nucleotide-gated ion channels. *Plant Physiol*. 128:400-410

Leustek, T., Martin, M.N., Bick, J-A., Davies, J.P. (2000) Pathways and regulation of sulfur metabolism revealed through molecular and genetic studies. *Annu. Rev. Plant Physiol. Plant Mol. Biol*. 51:141-165

Lévy, J., Bres, C., and Geurts, R. (2004) A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* 303:1361-1364

Lewandowska, M. and Sirko, A. (2008) Recent advances in understanding plant response to sulfur-deficiency stress. *Acta Biochim. Pol*. 55:457-471

Li, J., Zhu, S., Song, X., Shen, Y., Chen, H., Yu, J., Yi, K., Liu, Y., Karplus, V.J., Wu, P., Deng, X.D. (2006) A rice glutamate receptor-like gene is critical for the division and survival of individual cells in the root apical meristem. *Plant Cell* 18:340-349

Li, J.-Y., Jiang, A-L., Zhang, W. (2007) Salt stress-induced programmed cell death in rice root tip cells. *Journal of Integrative Plant Biology* 49:481-486

Linkohr, B.I., Williamson, L.C., Fitter, A.H., Leyser, O. (2002) Nitrate and phosphate availability and distribution have different effects on root system architecture of *Arabidopsis*. *The Plant Journal* 29:751-760

Little, D.Y., Rao, H., Oliva, S., Daniel-Vedele, F., Krapp, A., Malamy, J.E. (2005) The putative high-affinity nitrate transporter NRT2.1 represses lateral root initiation in response to nutritional cues. *Proc. Natl. Acad. Sci. U. S. A.* 102:13693-13698

Liu, H., Wang, S., Yu, X., Yu, J., He, X., Zhang, S., Shou, H., Wu, P. (2005) ARL1, a LOB-domain protein required for adventitious root formation in rice. *Plant J*. 43:47-56

Liu, J. and Zhu, J.K. (1998) A calcium sensor homolog required for plant salt tolerance. *Science* 280:1943-1945

Liu, K-H., Huang, C-Y., Tsay, Y-F. (1999) CHL1 is a dual-affinity nitrate transporter of *Arabidopsis* involved in multiple phases of nitrate uptake. *Plant Cell* 11:865-874

Liu, S., Wang, J., Wang, L., Wang, X., Xue, Y., Wu, P., Shou, H. (2009) Adventitious root formation in rice requires OsGNOM1 and is mediated by the OsPINs family. *Cell Res*. 19, 1110-1119

Lohar, D.P., Schaff, J.E., Laskey, J.G., Kieber, J.J., Bilyeu, K.D., Bird, D.M. (2004) Cytokinins play opposite roles in lateral root formation, and nematode and Rhizobial symbioses. *Plant Journal* 38:203-214

Lohar, D.P., Sharopova, N., Endre, G., Penuela, S., Samac, D., Town, C., Silverstein, K.A.T., VandenBosch, K.A. (2006) Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiol*. 140:221-234

Luschnig, C. (2002) Auxin transport: ABC proteins join the club. *Trends in Plant Science* 7:329–332

Lontoc-Roy, M., Dutilleul, P., Prasher, S.O., Han, L., Brouillet, T., Smith, D. L. (2006) Advances in the acquisition and analysis of CT scan data to isolate a crop root system from the soil medium and quantify root system complexity in 3-D space. *Geoderma* 137:231-241

López-Bucio, J., Hernandez-Abreu, E., Sanchez-Calderon, L., Perez-Torres, A., Rampey, R. A., Bartel, B., and Herrera-Estrella, L. (2005) An auxin transport independent pathway is involved in phosphate stress-induced root architectural alterations in arabidopsis. Identification of BIG as a mediator of auxin in pericycle cell activation. *Plant Physiology* 137:681–691

Lopez-Bucio, J., Hernandez-Abreu, E., Sanchez-Calderon, L., Nieto-Jacobo, M.F., Simpson, J., Herrera-Estrella, L. (2002) Phosphate availability alters architecture and causes changes in hormone sensitivity in the *Arabidopsis* root system. *Plant Physiol*. 129:244-256

López-Ráez, J. A., and Bouwmeester, H. (2008) Fine-tuning regulation of strigolactone biosynthesis under phosphate starvation. *Plant Signaling & Behavior* 3: 963–965

Lorbiecke, R., and Sauter, M. (1999) Adventitious root growth and cell-cycle induction in deepwater rice. *Plant Physiology* 119:21–30

Lucas, M., Swarup, R., and Paponov, I. (2011) Short-Root regulates primary, lateral, and adventitious root development in Arabidopsis. *Plant Physiology* 155:384-398

Lucas, M., Godin, C., Jay-Allemand, C., Laplaze, L. (2008) Auxin fluxes in the root apex co-regulate gravitropism and lateral root initiation. *J. Exp. Bot.* 59:55-66

Ludwig-Muller, J. and Guther, M. (2007) Auxins as signals in arbuscular mycorrhiza formation. *Plant Signal Behav.* 2:194-196

Luo, Z.B., Janz, D., Jiang, X., Gobel, C., Wildhagen, H., Tan, Y., Rennenberg, H., Feussner, I., Polle, A. (2009) Upgrading root physiology for stress tolerance by ectomycorrhizas: Insights from metabolite and transcriptional profiling into reprogramming for stress anticipation. *Plant Physiol*. 151:1902-1917

Lynch, J.P. and Brown, K.M. (2001) Topsoil foraging – an architectural adaptation of plants to low phosphorus availability. *Plant Soil* 237:225-237

Lynch, J. P., and Brown, K. M. (2012) New roots for agriculture: exploiting the root phenome. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367:1598–1604

Ma, J.F. and Furukawa, J. (2003) Recent progress in the research of external Al detoxification in higher plants: a minireview. *J. Inorg. Biochem*. 97:46-51

Ma, J.F., Ryan, P.R., Delaize, E. (2001) Aluminium tolerance in plants and the complexing role of organic acids. *Trends Plant Sci*. 6:273-278

Ma, J.F., Shen, R., Zhao, Z., Wissuwa, M., Takeuchi, Y., Ebitani,T., Yano, M. (2002) Response of rice to Al stress and identification of quantitative trait Loci for Al tolerance. *Plant Cell Physiol*. 43:652-659

Ma, Z., Baskin, T.I., Brown, K.M., Lynch, J.P. (2003) Regulation of root elongation under phosphorus stress involves changes in ethylene responsiveness. *Plant Physiol*. 131:1381-1390

Macgregor, D. R., Deak, K. I., Ingram, P. A., and Malamy, J. E. (2008) Root system architecture in *Arabidopsis* grown in culture is regulated by sucrose uptake in the aerial tissues. *The Plant Cell* 20:2643–2660

Magalhaes, J.V., Garvin, D.F., Wang, Y., Sorrells, M.E., Klein, P.E., Schaffert, R.E., Li, L., Kochian, L.V. (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the poaceae. *Genetics* 167:1905-1914

Mahajan, S. and Tuteja, N. (2005) Cold, salinity and drought stresses: An overview. *Arch. Biochem. Biophys*. 444:139-158

Malamy, J.E. (2005) Intrinsic and environmental response pathways that regulate root system architecture. *Plant Cell Environ.* 28:67-77

Malamy, J.E. and Benfey, P.N. (1997) Down and out in *Arabidopsis*: the formation of lateral roots. *Trends Plant Sci*. 2:390-396

Malamy, J.E. and Ryan, K.S. (2001) Environmental regulation of lateral root initiation in *Arabidopsis*. *Plant Physiol*. 127:899-909

Mao, C.Z., Yang, L., Zheng, B-S., Wu, Y-R., Liu, F-Y., Wu, P. (2004) Comparative mapping of QTLs for Al tolerance in rice and identification of positional Al-induced genes. *J. Zhejiang Univ. Sci*. 5:634-643

Maraschin, F.D.S., Memelink, J., Offringa, R. (2009) Auxin-induced, SCF(TIR1)-mediated poly-ubiquitination marks AUX/IAA proteins for degradation. *Plant J*. 59:100-109

Marin-Rodriguez, M.C., Orchard, J., Seymour G.B. (2002) Pectate lyases, cell wall degradation and fruit softening. *J. Exp. Bot.* 53:2115-2119

Markmann, K. and Parniske, M. (2009) Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* 14:77-86

Marsh, J.F., Rakocevic, A., Mitra, R.M., Brocard, L., Sun, J., Eschstruth, A., Long, S.E., Schultze, M., Ratet, P., Oldroyd, G.E.D. (2007) *Medicago truncatula* NIN is essential for

rhizobial-independent nodule organogenesis induced by autoactive calcium/calmodulin-dependent protein kinase. *Plant Physiol*. 144:324-335

Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T., Takahashi, H. (2003) Transcriptome profiling of sulfur-responsive genes in *Arabidopsis* reveals global effects of sulfur nutrition on multiple metabolic pathways. *Plant Physiol*. 132:597-605

Maruyama-Nakashita, A., Nakamura, Y., Yamaya, T., Takahashi, H. (2004) Regulation of high-affinity sulfate transporters in plants: towards systematic analysis of sulphur signaling and regulation. *J. Exp. Bot*. 55:1843-1849

Matsumoto, H. (2000) Cell biology of aluminum toxicity and tolerance in higher plants. *Int. Rev. Cytol*. 200:1-46

Matsumoto, H., Senoo, Y., Kasai, M., Maeshima, M. (1996) Response of the plant root to aluminum stress: Analysis of the inhibition of the root elongation and changes in membrane function. *Journal of Plant Research* 109:99-105

Mayzlish-Gati, E., De-Cuyper, C., Goormachtig, S., Beeckman, T., Vuylsteke, M., Brewer, P. B., Beveridge, C. A., Yermiyahu, U., Kaplan, Y., Enzer, Y., Wininger, S., Resnick, N., Cohen, M., Kapulnik, Y., Koltai, H. (2012) Strigolactones are involved in root response to low phosphate conditions in Arabidopsis. *Plant Physiology* 160:1329–1341

McNally, K.L., Bruskiewich, R., Mackill, D., Buell, C.R., Leach, J.E., Leung, H. (2006) Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol*. 141:26-31

Meijer, M., and Murray, J. (2000) The role and regulation of D-type cyclins in the plant cell cycle. *Plant Molecular Biology* 43:621-633

Meng, Y., Ma, X., Chen, D., Wu, P., and Chen, M. (2010) MicroRNA-mediated signaling involved in plant root development. *Biochemical and Biophysical Research Communications* 393:345–349

Menzel, M.I., Oros-Peusquens, A-M., Pohlmeier, A., Shah, J., Schurr, U., Schneider, H.U. (2007) Comparing 1H-NMR imaging and relaxation mapping of German white asparagus from five different cultivation sites. *Journal of Plant Nutrition and Soil Science* 170:24-38

Mergemann, H. and Sauter, M. (2000) Ethylene induces epidermal cell death at the site of adventitious root emergence in rice. *Plant Physiol*. 124:609-614

Middleton, P.H., Jakab, J., Penmetsa, R.V., Starker, C.G., Doll, J., Kalo, P., Prabhu, R., Marsh, J.F., Mitra, R.M., Kereszt, A., Dudas, B., VendenBosch, K., Long, S.R., Cook, D.R., Kiss, G.B., Oldroyd, G.E.D. (2007) An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction. *Plant Cell* 19:1221-1234

Mitra, R., Gleason, C., and Edwards, A. (2004) A Ca 2 calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. *Proceedings of the National Academy of Sciences* 101:4701-4705

Miura, K., Rus, A., Sharkhuu, A., Yokoi, S., Karthikeyan, A. S., Raghothama, K. G., Baek, D., Koo, Y. D., Jin, J. B., Bressan, R. A., Yun, D-J., Hasegawa, P.M. (2005) The *Arabidopsis* SUMO E3 ligase SIZ1 controls phosphate deficiency responses. *Proceedings of the National Academy of Sciences* 102:7760–7765

Miwa, H., Kinoshita, A., Fukuda, H., and Sawa, S. (2009) Plant meristems: CLAVATA3/ESR-related signaling in the shoot apical meristem and the root apical meristem. *Journal of Plant Research* 122:31–39

Miyasaka, S.C., Buta, G.J., Howell, R.K., Foy, C.D. (1991) Mechanism of aluminum tolerance in snapbeans : root exudation of citric acid. *Plant Physiol*. 96:737-743

Miyasaka, S.C., Kochian, L.V., Shaff, J.E., Foy, C.D. (1989) Mechanisms of aluminum tolerance in wheat : An investigation of genotypic differences in rhizosphere pH, K, and H transport, and root-cell membrane potentials. *Plant Physiol*. 91: 1188-1196

Molas, M.L. and Kiss, J.Z. (2008) PKS1 plays a role in red-light-based positive phototropism in roots. *Plant Cell Environ.* 31:842-849

Moreno-Risueno, M. A., Van Norman, J. M., Moreno, A., Zhang, J., Ahnert, S. E., and Benfey, P. N. (2010) Oscillating gene expression determines competence for periodic Arabidopsis root branching. *Science* 329:1306–1311

Morita, M.T. (2010) Directional gravity sensing in gravitropism. *Annual Review of Plant Biology* 61:705-720

Moubayidin, L., Di Mambro, R., Sabatini, S. (2009) Cytokinin-auxin crosstalk. *Trends Plant Sci*. 14:557-562

Mullen, J.L. and Hangarter, R.P. (2003) Genetic analysis of the gravitropic set-point angle in lateral roots of *Arabidopsis*. *Adv. Space Res.* 31:2229-2236

Mullen, J.L., Wolverton, C., Ishikawa, H., Hangarter, R.P., Evans, M.L. (2002) Spatial separation of light perception and growth response in maize root phototropism. *Plant Cell Environ*. 25:1191-1196

Muller, B. and Sheen, J. (2008) Cytokinin and auxin interaction in root stem-cell specification during early embryogenesis. *Nature* 453:1094-1097

Munns, R. and Tester, M. (2008) Mechanisms of salinity tolerance. *Annual Review of Plant Biology* 59:651-681

Munos, S., Cazettes, C., Fizames, C., Gaymard, F., Tillard, P., Lepetit, M., Gojon, A. (2004) Transcript profiling in the chl1-5 mutant of *Arabidopsis* reveals a role of the nitrate transporter NRT1.1 in the regulation of another nitrate transporter, NRT2.1. *Plant Cell* 16:2433-2447

Nacry, P., Canivenc, G., and Muller, B. (2005) A role for auxin redistribution in the responses of the root system architecture to phosphate starvation in Arabidopsis. *Plant Physiology* 138:2061–2074

Nap, J. and Bisseling, T. (1990) The roots of nodulins. Physiologia Plantarum 79:407-414

Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., Hoefgen, R. (2003) Transcriptome analysis of sulfur depletion in *Arabidopsis thaliana*: interlacing of biosynthetic pathways provides response specificity. *The Plant Journal* 33:633-650

Nishimura, T., Nakano, H., Hayashi, K-I., Niwa, C., Koshiba, T. (2009) Differential downward stream of auxin synthesized at the tip has a key role in gravitropic curvature via TIR1/AFBs-mediated auxin signaling pathways. *Plant Cell Physiol*. 50:1874-1885

Noh, B., Bandyopadhyay, A., Peer, W. A., Spalding, E. P., and Murphy, A. S. (2003) Enhanced gravi- and phototropism in plant mdr mutants mislocalizing the auxin efflux protein PIN1. *Nature* 423:999–1002

Noh, B., Murphy, A. S., and Spalding, E. P. (2001) Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development. *The Plant cell* 13:2441–2454

Nutman, P.S. (1948) Physiological studies on nodule formation: I. The relation between nodulation and lateral root formation in red clover. *Annals of Botany* 12:81-96

Ogawa, M., Kay, P., Wilson, S., Swain, S.M. (2009) *ARABIDOPSIS* DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are Polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant Cell* 21:216-233

Ohkama, N., Takei, K., Sakakibara, H., Hayashi, H., Yoneyama, T., Fujiwara, T. (2002) Regulation of sulfur-responsive gene expression by exogenously applied cytokinins in *Arabidopsis thaliana*. *Plant Cell Physiol*. 43:1493-1501

Okushima, Y., Overvoorde, P.J., Arima, K., Alonso, J.M., Chan, A., Chang, C., Ecker, J.R., Hughes, B., Lui, A., Nguyen, D., Onodera, C., Quach, H., Smith, A., Yu, G., Theologis, A. (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17:444-463

Olah, B., Briere, C., Becard, G., Denarie, J., Gough, C. (2005) Nod factors and a diffusible factor

from arbuscular mycorrhizal fungi stimulate lateral root formation in *Medicago truncatula* via the DMI1/DMI2 signaling pathway. *Plant J*. 44:195-207

O'Malley, R.C., Rodriguez, F.I., Esch, J.J., Binder, B.M., O'Donnell, P., Klee, H.J., Bleecker, A.B. (2005) Ethylene-binding activity, gene expression levels, and receptor system output for ethylene receptor family members from *Arabidopsis* and tomato. *Plant J*. 41:651-659

Osmont, K.S., Sibout, R., Hardtke, C.S. (2007) Hidden branches: developments in root system architecture. *Annu. Rev. Plant. Biol*. 58:93-113

Overvoorde, P., Fukaki, H., and Beeckman, T. (2010) Auxin control of root development. *Cold Spring Harbor Perspectives in Biology* 2/6/ a001537

Pacheco-Villalobos, D., and Hardtke, C. S. (2012) Natural genetic variation of root system architecture from *Arabidopsis* to *Brachypodium*: towards adaptive value. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 367:1552–1558

Paquette, A.J. and Benfey, P.N. (2005) Maturation of the ground tissue of the root is regulated by gibberellin and SCARECROW and requires SHORT-ROOT. *Plant Physiol*. 138:636-640

Pareek, A., Singla-Pareek, S.L., Sopory, S.K., Grover, A. (2007) Analysis of salt stress-related transcriptome fingerprints from diverse plant species. *Genomics-Assisted Crop Improvement* 1:267-287

Parniske, M. (2008) Arbuscular mycorrhiza: the mother of plant root endosymbiosis. *Nat. Rev. Microbiol*. 6:763-775

Parry, G., Marchant, A., May, S., Swarup, R., Swarup, K., James, N., Graham, N., Allen, T., Martucci, T., Yemm, A., Napier, R., Manning, K., King, G., Bennett, M. (2001) Quick on the uptake: Characterization of a family of plant auxin influx carriers. *J. Plant Growth Regul*. 20:217-225

Paszkowski, U. and Boller, T. (2002) The growth defect of lrt1, a maize mutant lacking lateral roots, can be complemented by symbiotic fungi or high phosphate nutrition. *Planta* 214:584-590

Péret, B., De Rybel, B., Casimiro, I., Benková, E., Swarup, R., Laplaze, L., Beeckman, T., and Bennett, M. J. (2009a) Arabidopsis lateral root development: an emerging story. *Trends in Plant Science* 14:399–408

Péret, B., Larrieu, A., Bennett, M.J. (2009b) Lateral root emergence: a difficult birth. *J. Exp. Bot*. 60:3637-3643

Perez-Perez, J.M. (2007) Hormone signaling and root development: an update on the latest *Arabidopsis thaliana* research. *Functional Plant Biology* 34:163-171

Perez-Torres, C-A., Lopez-Bucio, J., Cruz-Ramirez, A., Ibarra-Laclette, E., Dharmasiri, W., Estelle, M., Herrera-Estrella, L. (2008) Phosphate availability alters lateral root development in *Arabidopsis* by modulating auxin sensitivity via a mechanism involving the TIR1 auxin receptor. *Plant Cell* 20:3258-3272

Pernisova, M., Klima, P., Horak, J., Valkova, M., Malbeck, J., Soucek, P., Reichman, P., Hoyerova, K., Dubova, J., Friml, J., Zazimalova , E., Hejatko, J. (2009) Cytokinins modulate auxin-induced organogenesis in plants via regulation of the auxin efflux. *Proc. Natl. Acad. Sci. U.S.A.*106:3609-3614

Perret, J.S., Al-Belushi, M.E., Deadman, M. (2007) Non-destructive visualization and quantification of roots using computed tomography. *Soil Biol. Biochem*. 39:391-399

Perrin, R.M., Young, L-S., Murthy, U.M.N., Harrison, B.R., Wang, Y., Will, J.L., Masson, P.H. (2005) Gravity signal transduction in primary roots. *Ann. Bot*. 96:737-743

Petrásek, J., and Friml, J. (2009) Auxin transport routes in plant development. *Development (Cambridge, England)* 136:2675–2688

Pirozynski, K.A. and Malloch, D.W. (1975) The origin of land plants: A matter of mycotrophism. *BioSystems* 6:153

Provorov, N.A. (2000) The population genetics of nodule bacteria. Zh. Obshch. Biol. 61:229-257

Puig, J., Pauluzzi, G., Guiderdoni, E., and Gantet, P. (2012) Regulation of shoot and root development through mutual signaling. *Molecular Plant* 5:974-983

Quintero, F.J., Ohta, M., Shi, H., Zhu, J-K., Pardo, J.M. (2002) Reconstitution in yeast of the *Arabidopsis* SOS signaling pathway for Na+ homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 99:9061-9066

Rani Debi, B., Chhun, T., Taketa, S., Tsurumi, S., Xia, K., Miyao, A., Hirochika, H., Ichii, M. (2005) Defects in root development and gravity response in the aem1 mutant of rice are associated with reduced auxin efflux. J. *Plant Physiol*. 162:678-685

Rebouillat, J., Dievart, A., Verdeil, J.L., Escoute, J., Giese, G., Breitler, J.C., Gantet, P., Espeout, S., Guiderdoni, C., Perin, C. (2009) Molecular genetics of rice root development. *Rice* 2:15-34

Reed, J.W. (2001) Roles and activities of Aux/IAA proteins in *Arabidopsis*. *Trends Plant Sci.* 6:420-425

Relic, B., Talmont, F., Kopcinska, J., Golinowski, W., Prome, J.C., Broughton, W.J. (1993) Biological activity of *Rhizobium sp*. NGR234 nod-factors on *Macroptilium atropurpureum*. *Mol. Plant Microbe Interact*. 6:764-774

Remans, T., Nacry, P., Pervent, M., Filleur, S., Diatloff, E., Mounier, E., Tillard, P., Forde, B.G., Gojon, A. (2006a) The *Arabidopsis* NRT1.1 transporter participates in the signaling pathway triggering root colonization of nitrate-rich patches. *Proc. Natl. Acad. Sci. U. S. A.* 103:19206-19211

Remans, T., Nacry, P., Prevent, M., Girin, T., Tillard, P., Lepetit, M., Gojon, A. (2006b) A central role for the nitrate transporter NRT2.1 in the integrated morphological and physiological responses of the root system to nitrogen limitation in *Arabidopsis*. *Plant Physiol*. 140:909-921

Reymond, M., Svistoonoff, S., Loudet, O., Nussaume, L., Desnos, T. (2006) Identification of QTL controlling root growth response to phosphate starvation in *Arabidopsis thaliana*. *Plant Cell Environ*. 29:115-125

Richter, S., Anders, N., Wolters, H., Beckmann, H., Thomann, A., Heinrich, R., Schrader, J., Singh, M.K., Geldner, N., Mayer, U., Jurgens, G. (2010) Role of the GNOM gene in *Arabidopsis* apical-basal patterning - From mutant phenotype to cellular mechanism of protein action. *Eur. J. Cell Biol*. 89:138-144

Rodriguez-Barrueco, C. and De Castro, F.B. (1973) Cytokinin-induced pseudonodules on Alnus glutinosa. *Physiologia Plantarum* 29:277-280

Ruppel, N.J., Hangarter, R.P., Kiss, J.Z. (2001) Red-light-induced positive phototropism in *Arabidopsis* roots. *Planta* 212:424-430

Rus, A., Yokoi, S., Sharkhuu, A., Reddy, M., Lee, B-H., Matsumoto, T.K., Koiwa, H., Zhu, J.K., Bressan, R.A., Hasegawa, P.M. (2001) AtHKT1 is a salt tolerance determinant that controls Na(+) entry into plant roots. *Proc. Natl. Acad. Sci. U. S. A.* 98:14150-14155

Ruyter-Spira, C., Kohlen, W., Charnikhova, T., Van Zeijl, A., Van Bezouwen, L., De Ruijter, N., Cardoso, C., Lopez-Raez, J. A., Matusova, R., Bours, R., Verstappen, F., Bouwmeester, H.(2011) Physiological effects of the synthetic strigolactone analog GR24 on root system architecture in Arabidopsis: another belowground role for strigolactones? *Plant Physiology* 155:721–734

Ruzicka, K., Simaskova, M., Duclercq, J., Petrasek, J., Zazimalova, E., Simon, S., Friml, J., Van Montagu, M,C.E., Benkova, E. (2009) Cytokinin regulates root meristem activity via modulation of the polar auxin transport. *Proc. Natl. Acad. Sci. U.S.A.* 106:4284-4289

Ryan, P.R., Ditomaso, J.M., Kochian, L.V. (1993) Aluminium toxicity in roots: an investigation of spatial sensitivity and the role of the root cap. *J. Exp. Bot*. 44:437-446

Saab, I. N., Sharp, R. E., Pritchard, J., and Voetberg, G. S. (1990) Increased endogenous abscisic acid maintains primary root growth and inhibits shoot growth of maize seedlings at low water potentials. *Plant Physiology* 93:1329–1336

Sabatini, S., Beis, D., Wolkenfelt, H., Murfett, J., Guilfoyle, T., Malamy, J., Benfey, P., Leyser, O., Bechtold, N., Weisbeek, P., Scheres, B. (1999) An auxin-dependent distal organizer of pattern and polarity in the *Arabidopsis* root. *Cell* 99:463-472

Sabatini, S., Heidstra, R., Wildwater, M., Scheres, B. (2003) SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes Dev*. 17:354

Saito, K. (2000) Regulation of sulfate transport and synthesis of sulfur-containing amino acids. *Curr. Opin. Plant Biol*. 3:188-195

Saito, K., Yoshikawa, M., Yano, K., Miwa, H., Uchida, H., Asamizu, E., Sato, S., Tabata, S., Imaizumi-Anraku, H., Umehara, Y., Kouchi, H., Murooka, Y., Szczyglowski, K., Downie, J.A., Parniske, M., Hayashi, M., Kawaguchi, M. (2007) NUCLEOPORIN85 is required for calcium spiking, fungal and bacterial symbioses, and seed production in *Lotus japonicus*. *The Plant Cell* 19:610–624

Sanchez-Calderon, L., Lopez-Bucio, J., Chacon-Lopez, A., Cruz-Ramirez, A., Nieto-Jacobo, F., Dubrovsky, J.G., Herrera-Estrella, L. (2005) Phosphate starvation induces a determinate developmental program in the roots of *Arabidopsis thaliana*. *Plant Cell Physiol*. 46:174-184

Sano, T. and Nagata, T. (2002) The possible involvement of a phosphate-induced transcription factor encoded by Phi-2 gene from Tobacco in ABA-signaling pathways. *Plant Cell Physiol*. 43:12-20

Scheres, B., McKhann, H. I., and Van Den Berg, C. (1996) Roots redefined: Anatomical and genetic analysis of root development. *Plant Physiology* 111: 959–964

Schmidt, W. and Schikora, A. (2001) Different pathways are involved in phosphate and iron stress-induced alterations of root epidermal cell development. *Plant Physiol*. 125:2078-2084

Schmohl, N., Pilling, J., Fisahn, J., Horst, W.J. (2000) Pectin methylesterase modulates aluminium sensitivity in *Zea mays* and *Solanum tuberosum*. Physiol. Plantarum 109:419-427

Schulze, J., Temple, G., Temple, S.J., Beschow, H., Vance, C.P. (2006) Nitrogen fixation by white lupin under phosphorus deficiency. *Ann. Bot.* 98:731-740

Schüßler, A., Schwarzott, D., Walker, C. (2001) A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol. Res.* 105:1413

Shane, M. and Lambers, H. (2005) Cluster roots: a curiosity in context. *Plant and Soil* 274:101-125

Sharp, R. E., Silk, W. K., and Hsiao, T. C. (1988) Growth of the maize primary root at low water potentials : I. Spatial distribution of expansive growth. *Plant Physiology* 87:50–57

Shashidhar, H. E., Henry, A., and Hardy, B. (2012) Methodologies for Root Drought Studies in Rice. Los Banos (Philippines) *Int. Rice Res. Inst.* 65p.

Sheng, M., Tang, M., Chen, H., Yang, B., Zhang, F., Huang, Y. (2008) Influence of arbuscular mycorrhizae on photosynthesis and water status of maize plants under salt stress. *Mycorrhiza* 18:287-296

Shilev, S., Sancho, E.D., Benlloch-Gonzalez, M. (2010) Rhizospheric bacteria alleviate salt-produced stress in sunflower. *J. Environ. Manage*.95:S37-S41

Shimizu, A., Yanagihara, S., Kawasaki, S., Ikehashi, H. (2004) Phosphorus deficiency-induced root elongation and its QTL in rice (*Oryza sativa* L.). *Theor. Appl. Genet*. 109:1361-1368

Signora, L., De Smet, I., Foyer, C.H., Zhang, H. (2001) ABA plays a central role in mediating the regulatory effects of nitrate on root branching in *Arabidopsis*. *Plant J*. 28:655-662

Simon, L., Bousquet, J., Levesque, R.C., Lalonde, M. (1993) Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature* 363:67-69

Sivaguru, M. and Horst, W.J. (1998) The distal part of the transition zone is the most Aluminum-sensitive apical root zone of Maize. *Plant Physiol*. 116:155-163

Sivaguru, M., Ezaki, B., He, Z-H., Tong, H., Osawa, H., Baluska, F., Volkmann, D., Matsumoto, H. (2000) Aluminum-induced 1-->3-beta-D-glucan inhibits cell-to-cell trafficking of molecules through plasmodesmata. A new mechanism of aluminum toxicity in plants. *Plant Physiol*. 124:991-1006

Sivaguru, M., Horst, W.J., Eticha, D., Matsumoto, H. (2006) Aluminum inhibits apoplastic flow of high-molecular weight solutes in root apices of *Zea mays* L. *Journal of Plant Nutrition and Soil Science* 169:679-690

Smit, P., Raedts, J., Portyanko, V., Debelle, F., Gough, C., Bisseling, T., Geurts, R. (2005) NSP1 of the GRAS protein family is essential for rhizobial Nod factor-induced transcription. *Science* 308:1789-1791

Snapp, S.S. and Shennan, C. (1992) Effects of salinity on root growth and death dynamics of tomato, *Lycopersicon esculentum*. *Mill. New Phytol*. 121:71-79

Soni, R. (1995) A family of Cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. *The Plant Cell* 7:85–103

Sreevidya, V. S., Hernandez-Oane, R. J., Gyaneshwar, P., Lara-Flores, M., Ladha, J. K., and Reddy, P. M. (2010) Changes in auxin distribution patterns during lateral root development in rice. *Plant Science* 178:531–538

Steffens, B. and Sauter, M. (2005) Epidermal cell death in rice is regulated by ethylene, gibberellin, and abscisic acid. *Plant Physiol*. 139:713-721

Steffens, B. and Sauter, M. (2009) Epidermal cell death in rice is confined to cells with a distinct molecular identity and is mediated by ethylene and H2O2 through an autoamplified signal pathway. *Plant Cell* 21:184-196

Steffens, B., Wang, J., Sauter, M. (2006) Interactions between ethylene, gibberellin and abscisic acid regulate emergence and growth rate of adventitious roots in deepwater rice. *Planta* 223:604-612

Steinmann, T., Geldner, N., Grebe, M., Mangold, S., Jackson, C.L., Paris, S., Galweiler, L., Palme, K., Jurgens, G. (1999) Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science* 286:316-318

Strabala, T., and O'Donnell, P. (2006) Gain-of-function phenotypes of many *CLAVATA3/ESR* genes, including four new family members, correlate with tandem variations in the conserved *CLAVATA3/ESR*. *Plant Physiology* 140:1331-1344….

Strack, D., Fester, T., Hause, B., Schliemann, W., Walter, M.H. (2003) Arbuscular mycorrhiza: biological, chemical, and molecular aspects. *J. Chem. Ecol*. 29:1955-1979

Stracke, S., Kistner, C., Yoshida, S., Mulder, L., Sato, S., Kaneko, T., Tabata, S., Sandal, N., Stougaard, J., Szczyglowski, K., Parniske, M. (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* 417:959–962

Sun, P., Tian, Q-Y., Chen, J., Zhang, W-H. (2010) Aluminium-induced inhibition of root elongation in *Arabidopsis* is mediated by ethylene and auxin. *J. Exp. Bot*. 61:347-356

Sun, J., Xu, Y., Ye, S., Jiang, H., Chen, Q., Liu, F., Zhou, W., Chen, R., Li, X., Tietz, O., et al. (2009) Arabidopsis ASA1 is important for jasmonate-mediated regulation of auxin biosynthesis and transport during lateral root formation. *The Plant Cell* 21:1495–1511

Suralta, R. R., Inukai, Y., and Yamauchi, A. (2008) Genotypic variations in responses of lateral root development to transient moisture stresses in rice cultivars. *Plant Production Science* 11:324–335

Svistoonoff, S., Sy, M-O., Diagne, N., Barker, D.G., Bogusz, D., Franche, C. (2010) Infection-specific activation of the *Medicago truncatula* Enod11 early nodulin gene promoter during actinorhizal root nodulation. *Mol. Plant Microbe Interact*. 23:740-747

Swarup, R., Kramer, E.M., Perry, P., Knox, K., Ottoline Leyser, H.M., Haseloff, J., Beemster, G.T.S., Bhalerao, R., Bennett, M.J. (2005) Root gravitropism requires lateral root cap and epidermal cells for transport and response to a mobile auxin signal. *Nat. Cell Biol*. 7:1057-1065

Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., Parry, G., Nielsen, E., De Smet, I., Vanneste, S., Levesque, M.P., Carrier, D., James, N., Calvo, V., Ljung, K., Kramer, E., Roberts, R., Graham, N., Marillonnet, S., Patel, K., Jones, J.D.G., Taylor, C.G., Schachtman, D.P., May, S., Sandberg, G., Benfey, P., Friml, J., Kerr, I., Beeckman, T., Laplaze, L., Bennett, M.J. (2008) The auxin influx carrier LAX3 promotes lateral root emergence. *Nat. Cell Biol*. 10: 946-954

Swensen, S.M. (1996) The evolution of actinorhizal symbioses: Evidence for multiple origins of the symbiotic association. *Am. J. Bot.* 83, pp. 1503-1512

Takahashi, H. (2010) Regulation of sulfate transport and assimilation in plants. In International Review of Cell and Molecular Biology (Kwang W. Jeon, ed), pp. 129-159, Academic Press

Takahashi, H., Noji, M., Saito, K. (1999) Molecular regulation and engineering of sulfur transport and assimilation. *Tanpakushitsu Kakusan Koso* 44:2291-2298

Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J., Saito, K. (2000) The roles of three functional sulfate transporters involved in uptake and translocation of sulfate in *Arabidopsis thaliana*. *Plant J.* 23:171-182

Taramino, G., Sauer, M., Stauffer, J.L., Multani, D., Niu, X., Sakai, H., Hochholdinger, F. (2007) The maize (*Zea mays* L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *Plant J.* 50:649-659

Tester, M. and Davenport, R. (2003) Na+ tolerance and Na+ transport in higher plants. *Ann. Bot.* 91:503-527

Ticconi, C.A., Delatorre, C.A., Lahner, B., Salt, D.E., Abel, S. (2004) *Arabidopsis* pdr2 reveals a phosphate-sensitive checkpoint in root development. *Plant J.* 37:801-814

Ticconi, C.A., Lucero, R.D., Sakhonwasee, S., Adamson, A.W., Creff, A., Nussaume, L., Desnos, T., Abel, S. (2009) ER-resident proteins PDR2 and LPR1 mediate the developmental response of root meristems to phosphate availability. *Proc. Natl. Acad. Sci. U. S.* A. 106:14174-14179

Tillich, H. (1977) Vergleichend morphologische Untersuchungen zur Identität der Gramineen-Primärwurzel. *Flora* 166:415-421

Tirichine, L., Sandal, N., Madsen, L.H., Radutoiu, S., Albrektsen, A.S., Sato, S., Asamizu, E., Tabata, S., Stougaard, J. (2007) A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* 315:104-107

Tirlapur, U.K. and Konig, K. (1999) Technical advance: near-infrared femtosecond laser pulses as a novel non-invasive means for dye-permeation and 3D imaging of localised dye-coupling in

the *Arabidopsis* root meristem. *Plant J*. 20:363-370

Tracy, S.R., Roberts, J.A., Black, C.R., McNeill, A., Davidson, R., Mooney, S.J. (2010) The X-factor: visualizing undisturbed root architecture in soils using X-ray computed tomography. *Journal of Experimental Botany* 61:311-313

Tsuchisaka, A. and Theologis, A. (2004) Unique and overlapping expression patterns among the *Arabidopsis* 1-amino-cyclopropane-1-carboxylate synthase gene family members. *Plant Physiol*. 136:2982-3000

Tung, C-W., Zhao, K., Wright, M.H., Ali, M.L., Jung, J., Kimball, J., Tyagi, W., Thomson, M.J., McNally, K., Leung, H. (2010) Development of a research platform for dissecting phenotype-genotype associations in rice (*Oryza* spp.) *Rice* 3:205-217 DOI:10.1007/s12284-010-9056-5

Ueda, M., Koshino-Kimura, Y., and Okada, K. (2005) Stepwise understanding of root development. *Current Opinion in Plant Biology* 8:71–76

Uexküll, H.R. and Mutert, E. (1995) Global extent, development and economic impact of acid soils. *Plant and Soil* 171:1-15

Umehara, M., Hanada, A., Magome, H., Takeda-Kamiya, N., and Yamaguchi, S. (2010) Contribution of strigolactones to the inhibition of tiller bud outgrowth under phosphate deficiency in rice. *Plant & Cell Physiology* **51**:1118–1126

Uozumi, N., Kim, E.J., Rubio, G., Yamaguchi, T., Muto, S., Tsuboi, A., Bakker, E.P., Nakamura, T., Schroeder, J.L. (2000) The *Arabidopsis* HKT1 gene homolog mediates inward Na(+) currents in xenopus laevis oocytes and Na(+) uptake in *Saccharomyces cerevisiae*. *Plant Physiol*. **122**:1249-1259

Verma, D. P. S., Hu, C.-A., and Zhang, M. (1992) Root nodule development: origin, function and regulation of nodulin genes. *Physiologia Plantarum* **85**:253–265

Vernie, T., Moreau, S., de Billy, F., Plet, J., Combier, J-P., Rogers, C., Oldroyd, G., Frugier, F., Niebel, A., Gamas, P. (2008) EFD Is an ERF transcription factor involved in the control of nodule number and differentiation in *Medicago truncatula*. Plant Cell **20**:2696-2713

Vierheilig, H., Bago, B., Lerat, S., and Piche, Y. (2002) Shoot-produced, light-dependent factors are partially involved in the expression of the arbuscular mycorrhizal (AM) status of AM host and non-host plants. *Journal of Plant Nutrition* **165**:21–25

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009) EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Research* **19**:327–335

Walch-Liu, P., Ivanov, I.I., Filleur, S., Gan, Y., Remans, T., Forde, B.G. (2006) Nitrogen

regulation of root branching. Ann. Bot. **97**:875-881

Wang, J.R., Hu, H., Wand, G-H., Li, J., Chen, J-Y., Wu, P. (2009) Expression of PIN genes in rice (*Oryza sativa* L.): tissue specificity and regulation by hormones. Mol. Plant. **2**:823-831

Wang, J.-W., Wang, L.-J., Mao, Y.-B., Cai, W.-J., Xue, H.-W., and Chen, X.-Y. (2005) Control of root cap formation by MicroRNA-targeted auxin response factors in Arabidopsis. *The Plant Cell* **17**:2204–2216

Williamson, L.C., Ribrioux, S.P.C.P., Fitter, A.H., Leyser, H.M.O. (2001) Phosphate availability regulates root system architecture in *Arabidopsis*. *Plant Physiol*. **126**:875-882

Wiegers, B. S., Cheer, A. Y., and Silk, W. K. (2009) Modeling the hydraulics of root growth in three dimensions with phloem water sources. *Plant Physiology* **150**:2092–103

Woll, K., Borsuk, L.A., Stransky, H., Nettleton, D., Schnable, P.S., Hochholdinger, F. (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant rum1. *Plant Physiol*. **139**: 1255-1267

Wu, G., Lewis, D. R., and Spalding, E. P. (2007) Mutations in Arabidopsis multidrug resistance-like ABC transporters separate the roles of acropetal and basipetal auxin transport in lateral root development. *The Plant Cell* **19**:1826–1837

Xie, Q. (2000) *Arabidopsis* NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes & Development* **14**:3024–3036

Xu, M., Zhu, L., Shou, H., Wu, P. (2005) A PIN1 family gene, *OsPIN1*, involved in auxin-dependent adventitious root emergence and tillering in rice. *Plant Cell Physiol*. **46**:1674-1681

Xu, K., Xu, X., Fukao, T., Canlas, P., Maghirang-Rodriguez, R., Heuer, S., Ismail, A. M., Bailey-Serres, J., Ronald, P. C., and Mackill, D. J. (2006) *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**:705–708

Xue, R., and Zhang, B. (2007) Increased endogenous methyl jasmonate altered leaf and root development in transgenic soybean plants. *Journal of Genetics and Genomics* 34**:**339–346

Yamada, H., Suzuki, T., Terada, K., Takei, K., Ishikawa, K., Miwa, K., Yamashino, T., and Mizuno, T. (2001) The *Arabidopsis* AHK4 histidine kinase is a cytokinin-binding receptor that transduces cytokinin signals across the membrane. *Plant and Cell Physiology* **42**:1017–1023

Yamaji, N., Huang, C.F., Nagao, S., Yano, M., Sato, Y., Nagamura, Y., Ma, J.F. (2009) A zinc finger transcription factor ART1 regulates multiple genes implicated in aluminum tolerance in rice. *Plant Cell* **21**:3339-3349

Yan, J., Shah, T., Warburton, M., Buckler, E.S., McMullen, M.D., Crouch, J. (2009) Genetic

characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4:e8451

Yano, K., Yoshida, S., Müller, J., Singh, S., Banba, M., Vickers, K., Markmann, K., White, C., Schuller, B., Sato, S., Asamizu, E., Tabata, S., Murooka, Y., Perry, J., Wang, T.L., Kawaguchi, M., Imaizumi-Anraku, H., Hayashi, M., Parniske, M. (2008) CYCLOPS, a mediator of symbiotic intracellular accommodation. *Proceedings of the National Academy of Sciences* **105**:20540–20545

Yi, K., Wu, Z., Zhou, J., Du, L., Guo, L., Wu, Y., Wu, P. (2005) OsPTF1, a novel transcription factor involved in tolerance to phosphate starvation in rice. *Plant Physiol*. **138**:2087-2096

Yokoyama, R., and Nishitani, K. (2004) Genomic basis for cell-wall diversity in plants. A comparative approach to gene families in rice and *Arabidopsis*. *Plant and Cell Physiology* **45**:1111–1121

Yoneyama, K., Takeuchi, Y., Sekimoto, H. (2007a) Phosphorus deficiency in red clover promotes exudation of orobanchol, the signal for mycorrhizal symbionts and germination stimulant for root parasites. *Planta* **225**:1031-1038

Yoneyama, K., Xie, X., Kusumoto, D. (2007b) Nitrogen deficiency as well as phosphorus deficiency in sorghum promotes the production and exudation of 5-deoxystrigol, the host recognition signal for arbuscular. *Planta* **227**:125-132

Yoneyama, K., Xie, X., Sekimoto, H. (2008) Strigolactones, host recognition signals for root parasitic plants and arbuscular mycorrhizal fungi, from Fabaceae plants. *New Phytologist* **179**:(2) 484-494

Yoshida, S., and Parniske, M. (2005) Regulation of plant symbiosis receptor kinase through serine and threonine phosphorylation. *Journal of Biological Chemistry* **280**:9203-9209

Yoshimoto, N., Takahashi, H., Smith, F.W., Yamaya, T., Saito, K. (2002) Two distinct high-affinity sulfate transporters with different inducibilities mediate uptake of sulfate in *Arabidopsis* roots. *Plant J*. **29**:465-473

Young, L.M. and Evans, M.L. (1996) Patterns of auxin and abscisic acid movement in the tips of gravistimulated primary roots of maize. *Plant. Growth Regul*. **20**:253-258

Yun, H., Joo, S., Park, C., Kim, S. (2009) Effects of brassinolide and IAA on ethylene production and elongation in maize primary roots. *Journal of Plant Biology* **52**:268-274

Zeng, H.-Q., Zhu, Y.-Y., Bao, Y., Shen, Q.-R., Guo, K., Huang, S.-Q., and Yang, Z.-M. (2010) Relationship between the development of tomato lateral roots and expression of miR164, NAC1 under P deficiency. *Acta Metallurgica Sinica* **16**:166–171

Zhang, H. and Forde, B.G. (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. *Science* **279**:407-409

Zhang, H., Jennings, A., Barlow, P.W., Forde, B. (1999) Dual pathways for regulation of root branching by nitrate. *Proc. Natl. Acad. Sci. U. S. A.* **96**:6529-6534

Zhang, Z., Ersoz, E., Lai, C-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M., Buckler, E.S. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet*. **42**:355-360

Zhao, Y., Hu, Y., Dai, M., Huang, L., and Zhou, D.-X. (2009) The WUSCHEL-related homeobox gene *WOX11* is required to activate shoot-borne crown root development in rice. *The Plant Cell* **21**:736–748

Zhu, J.K. (2002) Salt and drought stress signal transduction in plants. *Annu. Rev. Plant. Biol*. **53**:247-273

Zhu, J.K., Liu, J., Xiong, L. (1998) Genetic analysis of salt tolerance in *Arabidopsis*. Evidence for a critical role of potassium nutrition. *Plant Cell* **10**:1181-1191

Zidan, I., Azaizeh, H., Neumann, P.M. (1990) Does salinity reduce growth in maize root epidermal cells by inhibiting their capacity for cell wall acidification? *Plant Physiol*. **93**:7-11

Zimmermann, R. and Werr, W. (2005) Pattern formation in the monocot embryo as revealed by NAM and CUC3 orthologues from *Zea mays L. Plant Mol. Biol*. **58**:669-685

CHAPTER 4 - RICE: RESEARCH TO PRODUCTION - COURSE HISTORY, BACKGROUND, OVERVIEW, AND FORMAL EVALUATION[2]

Introduction

The Rice: Research to Production course is an intense three-week field and lecture course aimed specifically at introducing junior researchers and graduate students to the global scientific, agronomic, and socioeconomic aspects of rice research. Initiated in 2007 as a collaboration between Cornell University and the International Rice Research Institute, the course endeavors to educate and inspire a new generation of rice scientists" by bringing rice research to life and forming networks for international and interdisciplinary groups of young scientist participants. This process and impact evaluation was designed to access the course outcomes over its first five years, specifically focusing on areas related to cultural learning and understanding, post-course participant networking, and course and participant goal fulfillment. In addition, this evaluation offers three main deliverables in the form of a set of formal networking tools for continued participant and staff contact, suggested short, mid, and long-term changes for further course refinement, and a set of guidelines to keep under consideration when effecting future changes in course content and methodology.

Course History

[2] Published as: Jung, J., Caffarella, R., & Caflarella, R. S. (2010). Advancing cultural knowledge: Experiential learning international graduate study training programs in the health and STEM disciplines. Proceedings of the Adult Education Research

The intensive three-week course entitled 'Rice: Research to Production' was the brainchild of Susan McCouch, Hei Leung and Bob Zeigler during conversations in 2005. The idea was a response to the challenge of how to bring the activities and mission of the International Rice Research Institute (IRRI) to the attention of a wider spectrum of young research scientists around the world. Contemplating the fact that not even her own graduate students studying rice genetics at Cornell University had much chance to visit IRRI or interact with the international rice community, McCouch, Leung (a long-time scientist at IRRI) and Zeigler, the Director General of IRRI, set out to develop a short course, to be hosted by IRRI, that would introduce an international group of young rice researchers to each other, to IRRI, to the international rice community, and to current topics in rice research.

The course idea and design was based on a very successful two-week 'Rice Production' course that ran at IRRI from 1970-1990. The original 'Rice Production' course was required for all incoming IRRI researchers, serving to familiarize newly recruited scientists with the field operations at IRRI and with the basics of rice crop production, including modern and traditional methods of land preparation, planting, and harvest, as well as insect and weed recognition. Over time, interest in the course waned due to a shift in research focus at IRRI toward more upstream research and the lengthy course duration, thus the decision was made to terminate the course in 1990.

Drawing from her own personal experience as a former IRRI scientist, Dr. McCouch suggested that the new short course be designed as a reinvigorated, modernized version of the original, which would serve to introduce graduate students and young scientists to the world of field-based tropical rice agriculture, and the research, resources, and opportunities for rice science

available at IRRI and in collaborating institutions around the world. The course concept, along with proposed internship opportunities for participants, was pitched to several international and US funding agencies, and a four-year grant from the National Science Foundation's Developing Country Collaboration program (NSF-DCC) was awarded to Cornell University in collaboration with IRRI in 2006. The NSF agreed to fund ten U.S. participants each year to attend the newly-christened 'Rice: Research to Production' (R2P) course, and it also provided funds to cover the costs developing and hosting the course at IRRI each year. IRRI committed to selecting and funding a complement of around ten Asian participants--graduate students and young scientists interested in rice science and agriculture--, and additional funds from other sources have been procured for up to four European and four African participants each year.

Dr. Noel Magor, IRRI's newly-hired Head of Training Center, took on the position as the course director in charge of education and coordination, IRRI scientist Dr. Hei Leung became the scientific director, David Shires, an IRRI training center consultant was asked to help with the design and organization of the course. and Dr. McCouch rounded out the course directorship in charge of the recruitment of US participants and the management of the NSF-DCC grant. An outstanding team of several IRRI training center staff is also in charge of much of the annual coordination, management, and running of the course. Since the inaugural R2P course in May 2007, the course has run annually for the past 5 years, graduating a total of 73 graduate and undergraduate students, 40 junior scientists, 8 senior scientists, 10 postdoctoral scholars, 4 administrators, and 2 high school teachers participants, from 34 different countries.

Course Goals and Objectives

The overarching course goal, as initially expressed by the directors, was "to create a new

generation of plant scientists that are both well-networked in the international community and understand the importance of innovative plant science in addressing global problems" (McCouch, 2008). Course directors aimed to bring rice research to life for the participants, making the course an exciting and engaging experience and an introduction to the global reality of the crop, such that "getting their feet wet in the rice paddies of Asia" will help participants "develop a feeling for the rice plant in its native environment" and "appreciate many facets of the crop's importance throughout the world" (McCouch, 2008).

The first secondary goal was to bring together two diverse communities of graduate students and young scientists: those from America and Europe, who work on rice as a model organism, but are largely ignorant of rice within a global agricultural context, and those from rice growing nations of Asia and Africa who are deeply familiar with rice from a historical, cultural, economic, and biological perspective. The second secondary goal was to highlight the scientists, ongoing research and institutional resources available for rice science at IRRI and other partner institutes. To achieve these goals, two main objectives were identified: 1) encourage plant science graduate students from developed nations to consider careers in international agriculture or rice-related research, and 2) create an international network of young scientists to help solve food security issues.

IRRI R2P course directors and coordinators also identified six specific participant outcome objectives in their 2008 course outline and summary report:

- An understanding of the basics of rice production, with on field-based experience at IRRI, and field visits to the main rice production region in the Philippines

- Familiarity with the germplasm collection at IRRI and current issues related to germplasm exchange and property rights
- An appreciation of the research issues of IRRI and its development partners
- Hands-on skills relating to rice breeding, molecular genetics, and genomics
- An understanding of how to structure effective international collaborations
- A plan and personal contacts to work effectively as part of the international research community in the future (IRRI, 2008)

Participant-identified Goals and Objectives

Participants also brought their own personal goals for the course, as subconscious or non-expressed expectations, and as major goals articulated in their applications as the expected benefits they hoped to derive. Participant goals and objectives, specific to the course year, and to each individual, did not seem to be taken under very serious consideration when planning each year's course content, or named by course directors as an additional aggregate goal which they aimed to meet. Participant goals and objectives were identified through this evaluation and will be discussed in the Findings and Recommendation section.

Course Content and Structure

As suggested by the course name, the course content is selected to introduce and educate participants on a wide range of techniques and issues pertaining to rice farming, cultivation methods, crop breeding and improvement, and basic and applied rice research. IRRI scientists and staff present their current areas of research on the global rice economy, germplasm collection and evaluation, traditional and marker-based breeding programs, and engineering

improvements in cultivation and processing technologies. The course content was originally designed for participants with interests and expertise in the plant sciences, particularly in genetics and molecular biology. However, in response to a concerted demand from the 2007 inaugural year participants, for more social science content, as well as a growing number of participants from agricultural education and extension-related fields, social science issues relevant to rice agriculture, rural development, and poverty alleviation have continued to be a growing component of the course.

Participants not only view demonstrations, but enter the rice paddies to take part in the entire rice growing process, using both traditional and modern methods of land preparation, seed nursery setup, transplanting and sowing, harvesting, and post harvest processing. Group visits are made to IRRI research programs, experimental plots, greenhouses, and the seedbank, with lab activities ranging from DNA isolation, seed sorting and pathogen identification, to panicle emasculation and crossing. Major off-site trips are taken to a beach for recreational time, a tour of Phil-Rice, the Philippine National Rice Research Institute, and farmer interviews at the 2,000 year-old Ifugao rice terraces in Banaue. In addition, participants are required to develop and present a final group project on a national or regional agricultural development issue requiring social and/or scientific interventions.

The course has three main components: seminars and classroom-based activities, field and lab practical exercises, and visits to IRRI labs and facilities, other institutes, farmer's fields, and rice agriculture-related sites. The daily program generally runs from 8:30AM to 5:30PM daily, with mid-morning, lunch, and mid-afternoon breaks. In each of the course schedules for the past 5 years, there is usually a daily mix of seminars and either practical exercises or program tours,

interspersed with a few days solely devoted to seminars and classroom-centered discussions and group work. Aside from a few scheduled formal dinner events, participants are free to relax, mingle, and explore the IRRI campus or surrounding town of Los Banõs in the evenings.

Since 2008, when a more significant off-institute travel component was added, specifically the trips to PhilRice and Banaue, off-campus travel to visit tourist or agriculture-related sites has been scheduled during overnight or multi-day trips on the first and second weekends of the course.

Need for Evaluative Study

The R2P goals and objectives appear to be relatively well-defined within the planning stages and introductory years of the course initiation. However, beyond coordinator design and review of immediate post-course program evaluations completed by the participants, no internal training center structures were put into place to assess immediate or longer term course or individual participant impact or goal fulfillment. Participant suggestions for changes and improvements were seriously reviewed by the course coordinators and often implemented when planning the following year's course. In addition, there had been no move to establish a formal communication network among past participants, follow-up on their academic or professional development, or develop plans for future course innovation and expansion.

This evaluation was conceived and undertaken by a Cornell graduate student minoring in Education who was a participant in the 2008 R2P course. The particular intent of this study was to evaluate cultural learning in the course, but also to fulfill the need for a broad evaluation of the course and the institutional and individual impact over a five-year span, as well as to develop

networks, guidelines, and evaluation methods for potential implementation.

Purpose of Evaluation

This study was designed as both a process and impact evaluation to consider elements of the course progress, and assess impact and goal fulfillment of the R2P course throughout the first five years of its existence. The primary areas of focus are aspects related to building professional, intercultural working relationships between and among the participants. The data drawn from this study was also used to determine what participants learned, the value of the instructional techniques used, and the thoughts of staff, scientists, and participants on the general impact and future directions of the course.

Evaluation Goals

The initial goals outlined for this course evaluation were primarily focused on evaluating the cultural learning of participants, their formation of personal or professional relationships during the course and an assessment of ongoing interaction thereafter.

Primary Focus

1. To determine the participants' perceptions of whether, and if so, how cultural understanding was enhanced as a result of the program.
2. To ascertain from staff and participants a set of possible networking tools, which could assist in developing and maintaining a sense of professional community and building international collaborative inquiry after the course is completed.

3. To explore whether participants continue to interact after the course, either between individuals, or in groups related to their research, and if so what mechanisms have they used to build and sustain these relationships.

Secondary Focus

1. To assess from the perceptions of the participants what they have learned from the course related to their work as researchers immediately after the course and nine months after the course.
2. To determine from the evaluation of participants', scientists', and staffs' perceptions which instructional techniques were the most effective in aiding participant learning.

Evaluation Objectives and Deliverables

Objectives

1. An evaluation of the fulfillment of participant, staff, and scientist goals and expectations
2. An indication of whether and if so, why and how, cultural understanding developed amongst participants
3. An initial plan for concrete networking tools with which participants can easily keep in contact with each other and with staff

Deliverables

1. A set of appropriate networking tools to encourage the maintenance and strengthening of professional relationships after the course

2. Proposed changes to assist in the further tailoring of future R2P courses or other interdisciplinary IRRI training courses
3. Suggested content and methodological considerations and guidelines to direct the refinement of future versions of the course

Evaluation Methodology

Qualitative design

This study was designed as a qualitative study involving styles of both process and impact evaluation to determine the status and evolution of course content and methodology, and examine the definition and fulfillment of different goals, over the five-year history of the R2P course. Process evaluations use a wide range of methods including interviews, focus group sessions, and written document review to record discussion and reflection on program status and suggestions for improvement (Rogers and Goodrick, 432). Impact evaluation provides a way to explore a range of program impacts—from the intended to the unanticipated—and what they mean to the participants, staff, and other stakeholders involved in the program (Rogers and Goodrick, 432).

Study participants

Participants involved in the study include all 137 participants of the R2P course from 2007 to 2011, as well as three IRRI center staff, three IRRI scientists, and three course directors. Informed consent was gathered from all participants prior to the use of any primary documents they had generated either in written form or as interviews before being included in this study.

Data collection methods

Data was gathered from existing and self-collected primary document sources, including participant applications, mid and post-course evaluations, participant and staff interviews, and short-term follow-up surveys. Pre-existing primary document sources included 2007 to 2011 participant applications and 2007 to 2011 IRRI designed participant evaluations, made available by Cornell and IRRI staff involved with the course (Table 4.1).

Additional data collected during the 2010 R2P course include midcourse daily question sets answered by the participants, midcourse interviews by key informants, including staff, scientists, directors, and participants, independent observations by one evaluator, and self-designed immediate post-course evaluation also answered by the participants. Two short to midterm follow-up surveys were also designed by the evaluator to investigate participant career and personal development, network and relationship retention, and suggestions for course improvement. The first of these were offered in late 2008 to participants from the 2007 and 2008 courses, 1.5 years and six months after their participation in the course, respectively. The second follow-up survey was offered in late 2011 to participants from the 2007 to 2011 RTP courses, 4.5, 3.5, 2.5, 1.5 years, and six months, after their participation in the course, respectively (Table 4.1).

Table 4.1. Primary document sets used in this evaluation, course years of document type, and number of documents (*N*) analyzed from each course year.

Primary document type	Course years from which documents were available	*N*[1]
Application Information (US participants only)	2007	2
	2008	8
	2009	11
	2010	13
	2011	11
Midcourse interviews - staff and directors	2010	6
Midcourse interviews - scientist	2010	3
Midcourse group interviews - participants	2010	3
Midcourse interviews - individual participant	2010	1
Midcourse daily question sets	2010	172
Course independent observations	2010	1
IRRI immediate post-course evaluations	2007	24
	2008	29
	2009	28
	2010	24
	2011	26
Self-designed immediate post course evaluations	2010	24
Participant Followup - Informal survey 2008	2007	11
	2008	20
Participant Followup - Informal survey 2011	2007	5
	2008	10
	2009	12
	2010	16
	2011	4

[1] For all document types, *N* is also equivalent to the number of individual respondents, with the exception of the midcourse daily question sets, in which N is the sum of individual responses to six subsets of questions, each answered by 23-28 participants.

Use of secondary sources

Two secondary sources were used in this study as reference sources to identify R2P course goals and objectives as initially outlined by the course directors. The secondary sources include: 1) the successful National Science Foundation (NSF) 2008 Developing Country Collaboration (DCC) Request for Supplemental Extension authored by McCouch, and 2) the IRRI training center-authored unpublished 2008 course outline and summary.

Data analysis procedures

Pre-existing primary document sources consisting of standardized questions from multiple course years were compiled and converted into Excel spreadsheets. These included participant applications, self-designed and IRRI immediate post-course evaluations, daily question sets, and both follow-up surveys. All interview recordings were transcribed by a professional transcriptionist.

All primary document data sets were organized and coded by the evaluator using the qualitative data analysis software Atlas.TI. Quotes from the data were coded according to subject and significance. Codes were then organized into themes relevant to the fulfillment of one or more of the evaluation goals, objectives, and deliverables. When possible, themes, codes, and their corresponding quotes were linked together into networks according to integral, correlational, causal, or topical relationships. Networks, as well as individual themes, codes, and quotes, were analyzed and used to derive findings and responses for the evaluation objectives. The number of quotes associated with each code was also taken into consideration as a representative statistic of the relative significance of the code as pertaining to the theme(s) it was subclassified under.

Findings and Recommendations

This section details the findings derived from primary document analysis and associated recommendations to modify the course approach, content or methodology based on those findings. Each year, the course coordinators have taken into consideration the previous year's participant evaluations while planning the course for each successive year, incorporating recommended changes or not according to perceived importance, and logistical ability. The comprehensive recommendations detailed here are derived from and incorporate a much wider variety of sources, informants, themes, course cohorts and timescale than have been previously analyzed together. Below each set of findings are bulleted lists of recommended actions corresponding to that topic.

Participant-identified Goals and Objectives

Participant goals for the course, pertaining to their individual interests and their expected derived benefits from the course were determined through compilation and comparison of responses to application questions on why they wish to participate in the course, the skills or knowledge they want to acquire, and the course's relevance to their work. Only the applications of successful applicants who were NSF-funded US citizens and a few non-US citizen participants from US institutions were reviewed in this study; applications of participants from outside the US were unavailable. The majority of these participant applications reviewed were from graduate students, as well as a few undergraduates, postdocs, and teachers. Their responses on expected course benefits indicate a range of educational, professional, and personal goals, many of which were unique to the individual, but several main goals were often shared amongst most participants.

As might be expected, the majority of respondents indicated “rice production /cultivation," “networking,” “connecting rice as a model organism to [rice as] a crop plant,” or an interest in the “mission of IRRI” or other CG centers as reasons for wanting to join the course, all of which dovetail nicely with the course directors initial reasons for creating the course. Participants also prominently cited educational and career goals, as reasons for coming, such as: "find future employment,” “future or current research application” and “professional development” as specific interests prior to the course.

Some expectations or goals expressed by many participants were not met or not meant to be addressed during the course e.g., those having to do with gaining technical skills and expertise in particular disciplines. Participants also expressed a desire to expand and apply their knowledge and understanding beyond their specific areas of expertise, wanting to “understand stakeholder opinions,” learn about “research/knowledge application” and “complement research experience with production experience” and “gain global perspective,” and “solve global problems”.

Retrospectively, the main, presumptively-fulfilled participant goals and objectives, as derived from immediate post-course evaluation responses on reasons for coming, include: “a broad comprehensive overview of rice,” “interest in rice production, cultivation, and/or research,” “networking,” “solving production/social problems.” The “broadening of specific understanding/knowledge,” “career development/discernment,” “interaction with different people/ideas/cultures,” “job application/effectiveness,” “learning about IRRI/CG/IRRI scientists” also seem to be major objectives for many participants which were likely fulfilled by the course.

- Advertise true course content -- the course is not intended to help participants gain mastery of technical or research-based skills, but to provide a broad knowledge base, different perspectives, and the opportunity to develop interpersonal skills
    - Remove or reword application questions on skills participants hope to gain in order to relieve participant expectations in that area
- Take note of participant goals and objectives as well as their post course fulfillment
- Modify course content and structure to more closely address a wider range of main participant goals such as those regarding career development and discernment, and networking with IRRI staff and scientists

Cultural Understanding of Participants:

Many participants reported that their cultural understanding was enhanced, but often this appeared to be more in relation to social-economic, institutional, or departmental/field-related cultures. Socio-ethnic cultural understanding was often realized in either high-stress situations, e.g. during group project work, when participants from different countries and backgrounds had to work intensely together on a common project, or conversely in relaxed settings during informal conversation. In responses to evaluation questions about course-derived benefits and the most valuable aspects of the course, participants often said they valued the "diversity" and range of people from "different countries", "fields" and "backgrounds," adding that "meeting people with shared common interests and goals was very helpful and stimulating". In post-course evaluations, eight respondents found cultural experiences to be a highlight of the course, while 34 cited culture as a specific benefit. The development of cultural understanding in the

course has had long-term effects on participants' psyche and careers: in follow-up surveys, six respondents mentioned cultural components as leaving the strongest impression. Institutional and socio-ethnic cultural exposure may be seen to affect several participants in career decisions concerning work at IRRI or other CG centers, and an increased capability or desire to work internationally or with diverse groups of people.

- Support and increase informal and formal opportunities to further participants' cultural understanding from multiple perspectives: local Filipino/tribal; ethnic and field/discipline-based; institutional
- Encourage participants to actively take advantage of the opportunities to enhance their cultural understanding in these areas

Networking Tools

Participants recommended existing online social networking groups, primarily Facebook, and LinkedIn, as well as email groups or listserves as ways to keep in contact. Other options mentioned were the development of a non-institute affiliated, participant-developed course webpage, and also the possibility of in-person reunions and sharing of research and career development at a R2P post-course symposium.

- Develop R2P networking group(s) based on existing social or professional networking sites for past, current and future course participants, as well as staff and scientists

Post-course Participant Interaction and Networking

Follow-up study responses suggest that there does not seem to be a constant, continuing

interaction among most participants or participants and staff, unless they became close friends or have had or made opportunities to do collaborative research with each other. Network and relationship development was mentioned as being a very important result of the course in immediate post-course evaluations by participants, and several had called on IRRI and the training center to provide a way in which these networks could continue to exist and be strengthened. Network formation was mentioned by course coordinators as one of the two primary goals of the course, however, there has been little action on formal network set-up or maintenance, as a responsibility of the course directors/training center.

Participants initially saw a need for social networking platforms, developing an e-mail listserv for the 2007 course members and a Facebook group after the 2010 course, with both efforts initiated shortly after the course had been concluded in each of these years. However, there was found to be a drop-off in communication as people returned to their home institutes and work routines, such that most participants eventually lost contact with most members of their cohort. As the primary means of communication is electronic, through e-mails and internet networking sites, another problem identified was that of people who couldn't be contacted because of email address changes.

- Develop, manage, and facilitate an IRRI training center-based networking platform for past, current and future R2P participants, as well as course staff and scientists
- Develop and keep updated a database of contact info for all R2p participants, either internally or through the networking platform to facilitate communication

- Design and offer group digital or in-person opportunities, such as educational resources, courses, newsletters, or conferences to keep participant networks active

Course-derived Benefits to Participants

Participant goals and expectations vs. their perceived post-course derived benefits are not in line with each other. Although some major content-related goals are the same: many cited the 'broad understanding of rice' or an 'understanding of global agricultural issues' and 'networks' in both categories, responses indicate that participants perceive more benefits regarding their professional or psycho-social development as opposed to information and skill acquisition. Participants have said they were 'challenged,' and gained 'clarity,' 'confidence,' 'exposure,' 'hope,' 'insight,' and 'inspiration,' as well as a 'new approach to problems/different disciplines,' a better appreciation of farm labor or of rice research, and a 'greater awareness of social/cultural/economic issues'. While over 80 responses mentioned 'field practicals and hands-on exercises,' and over 20 'staff and scientist engagement and accessibility' as highlights of the course, the ability to 'better plan research goals' and 'interact with different people' as well as having their 'assumptions,' 'strategic thinking,' and 'comfort zones' challenged were concrete benefits.

Immediate benefits from the course, vs. long-term course-related career effects and perceptions of course components with the greatest impact are surprisingly also correlated. In both long and short-term surveys, intangible elements involving relationships and the development of professional or personal thinking feature far more prominently than the gain or utility of information or techniques. Only five and seven responses, respectively, detailed knowledge/information and IRRI facilities, resources, and scientists as having made a strong, and lasting impression, yet over 55 responses featured culture, diversity,

relationships/interactions with other participants as having the greatest impact. The trip to Banaue and field or hands-on activities were also mentioned in many responses as having made a big impression. At the times the follow-up surveys were completed, a high number of participants report that the course was inspirational or motivational in helping them consider, explore, continue, or renew career interests, pathways, or possibilities.

The relative youth of the participants and the fact that many of them report starting, continuing, or finishing advanced graduate degrees, and beginning professional careers in the years after the course heighten the positive impact potential the R2P course can have on these young researchers. As a result of the course, at least three former participants have gone on to not only take internships with IRRI scientists, but also initiate collaborations with IRRI that have formed a significant part of their graduate research. While several former participants have said that they are continuing or planning to work on plant research or agriculture, if not specifically on rice research, it will be interesting to follow up with all the participants some ten years after the course after they are more settled in their careers to see where their paths have led and how the course may have played a role in shaping those trajectories.

- Take past participant long and short-term benefits into consideration when planning future courses
- Consider developing and funding IRRI-administered competitive internship opportunities, or other such opportunities to strengthen participant collaborative relationships with IRRI staff and scientists

Effective Instructional Techniques and Approaches

Hands-on/practicum activities, group work and discussion, and interaction with staff and scientists, as well as farmer interaction, and trips to other institutions and rice-growing regions were cited by participants as the factors most important or impactful to learning. All of these involved components of experiential, action-based learning and interactive, discussion-fueled exchange of ideas and information. This stands in stark contrast to the lengthy, presenter-centric, often highly technical seminars which made up the majority of the classroom activities, but were rarely mentioned by participants as being useful to learning.

One prominent exception is the comprehensive and applicatory approach to research exemplified by Dr. K.L. Heong's talk on his rice insect pest research and the social adoption of participatory management strategies derived from his studies. Multiple participants from multiple course years highlighted Dr. Heong's seminar as a widely accessible, engaging, and inspirational talk, due to his thorough introduction to the topic and problems, simple description of his team's research strategies, results, and the dissemination issues of his management techniques to Southeast Asian rice farming communities. Dr. Heong's combination of good presentation technique and a compelling problem-centric story, detailing the research process from agricultural issue to experimental design and analysis to result-based recommendations and social application, using simple, yet thorough explanations seems to be a successful presentation model which could be recommended to future course presenters.

One notable example of a subpar presentation was the 2010 seminar on communications by participant Jill Kuehnert. Jill's seminar represents an excellent attempt by IRRI to utilize and incorporate participant expertise as part of the course. However, several participants with backgrounds in the social sciences noted that while the constructing messages specific to each

section of one's audience demographic may be useful for one-way information transfer, the development of one-sided, opinionated messages, i.e. telling farmers to "do this" and detractors to "shove it" or "stop it," is a highly presumptive and dictatorial approach, instead of an attempt at genuine communication and open dialogue. More disturbing is the fact that this negative perception of the communications seminar was largely overlooked by participants coming from a biological research background who may lack the perception to see that the discussion was based on the biased presumption that everyone naturally supported transgenic crops. One participant with a molecular biological background, specifically noted that he would have been unaware of the shortcomings of this session, had he not talked with two other participants with expertise in agricultural development, outreach, and extension.

- Shorten and enforce lecture periods to one hour at most, leaving time for discussion with the presenter
- Integrate more practicum, group work, and problem-solving discussion-based activities into the course schedule, especially to break up multiple lecture series
- Continuously ground all course content in an agricultural issue or application- specific context, so all participants—regardless of their differing backgrounds-- can understand the greater agronomic utility of the research, technique, or resources being discussed
- Develop and recommend a series of presentation models, audience background summaries, and/or structural guidelines to IRRI scientist presenters for use as guidelines to tailor their presentations for maximal comprehension
- Avoid presenting grossly outdated or opinionated content unless encouraging and exploring a wider range of opinions views through discussion on the topic

Cultural Suitability of Teaching and Communication Methods

This issue was not sufficiently addressed within the primary documents that were analyzed.

Other Findings

Informal learning: Much of the learning taking place in the course is informal learning in non-structured environments outside the classroom or daily course schedule. In their immediate post-course evaluations and follow-up surveys, participants relate many interactions with each other or with staff over meals or recreation from which they garnered interesting and insightful knowledge that they would not have had the chance to do within the classroom context.

Individual participants are often driven to pursue these areas of discussion because they find these personally interesting and meaningful, however, many of these personal insights could also be useful to all participants. It is thus worth considering how IRRI might consider incorporating some of these issues into the formal context of the course, and allowing for more free time and comfortable environments to facilitate such interactions.

- Recommendations: similar to those under 'Cultural Understanding of Participants'

Scientist engagement: The three IRRI scientists who were interviewed all professed a deep interest in how they might best support the learning and engagement of the participants in regard to their respective contributions to the course curriculum, as well as the future of the course and its continuation. Dr. Michael Thomson specifically asked that any participant feedback on seminars and exercises be made available to the staff or researchers involved so that they can continue to make improvements in their presentation style and content. He also asked that lecturers be more well-informed by training center staff of the range of backgrounds and

experiences of the participants in order to tailor presentations for better engagement and comprehension. Dr. Endang Septiningsih suggested that there be a couple of research teams and projects the participants could focus on exploring in depth. By exposing participants to the research from initial problem and research conception, to end-product dissemination and adoption for 1-2 full days through exercises, lab/field tours, discussions, and demonstrations, participant comprehension of research significance and effect could be improved greatly.

- Ask for and give scientists participant and/or training center feedback and suggestions on their involvement with their course, whether with presentations or practicals
- Work with 1-2 willing scientists, such as Dr. Septiningsih, to develop out their course involvement to cover one of their research projects and take participants through the process of issue, goal, design, set-up, data collection, analysis, and end product development and adoption
- Provide, to all engaged speakers well in advance of the course, a brief list of the range of participant disciplines, educational background, as well a what course material will have already been covered, and suggestions on what the training staff would like them to present
- Have scientists give a short, one minute introduction on their background and how they became interested in their area of research

Objective-based Deliverables

Evaluation of Goal and Expectation Fulfillment

The main course director goals involving networking and the development of a new generation

of rice scientists are mid to long term goals needing further future evaluation. However, it that networks and network tools have been formed, but aren't necessarily being managed or used to their full potential by former participants and staff. To date, from 6 months to four years after their participation, most participants have continued in rice, plant, or science and agriculture-related careers or higher education—whether they do indeed become a new, innovative, and networked generation of rice scientists remains to be seen and bears future follow-up.

The first three participant outcome objectives: an understanding of the basics of rice field production, IRRI's germplasm resources, and the research done by IRRI and collaborators, respectively, appear to have been met, according to participant responses on course benefits and items learned. However, the second three objectives: acquisition of hands-on skills in plant breeding, genetics, and genomics, an understanding of how to structure effective international collaborations, and a plan and contacts for international research community networking, appear to have only been partially incorporated into the course and partially fulfilled. Participants participated in demonstrations of techniques used in plant breeding and genetics, such as spikelet emasculation and crossing, and DNA isolation, but did not and could not acquire or master such skills with a single attempt. Similarly, participants may have gained a list of contacts and some experience working with an international group of people, but were certainly not provided with or asked to generate a networking plan or precise insight into structuring collaborations.

Participant learning and benefits seem to far outstrip their initial course goals and expectations with respect to professional and personal psycho-social development. The depth and value of the relationships, perspectives, insight, and understanding participants gained on institutions, disciplines, peoples, and agronomic problems were largely unlooked for and unanticipated prior

to the course. With regard to course content and outcomes, most participants reported that they had gained the 'broad, comprehensive overview of rice,' and the 'networking' opportunities that they had attended the course in hopes of attaining. In spite of these successful examples of goal fulfillment, participants had many suggestions for course improvement to better serve future attendees. The most important of these were the demands for increased interaction, discussion, and inquiry-based activities, the more logical organization and explanation of technical content, and the mostly-unspoken needs for a greater clarity and parity of opinions and insights to be expressed, with emphasis on the relevance of concepts and techniques toward "big picture" problem-solving applications.

Networking Tools

One of the original objectives related in the design of this course evaluation was to determine from participant and staff feedback whether there was a demand for a formal, professional, R2P course network, and if so, what their recommendations were for the ideal means of communication, and the use of existing or new networking systems. We found, however, that not only was there a demand for course networking, but that it had already driven participants to develop their own informal or semi-formal networking tools, specifically a year-specific listserv generated by the 2007 group, and a Facebook group page created by a 2010 participant shortly after the end of that year's course.

As such, the development of additional networking tools became an imperative objective of this course evaluation. Queries on participant communication and networking suggestions were part of both follow-up studies, and the 2010 immediate post-course evaluation. From these responses, it was clear that most people preferred an e-mail or internet-based means of

networking, particularly one based on pre-existing social or professional networking sites. Based on these suggestions, it was decided to generate a new LinkedIn R2P course group as a professional networking site, which several participants suggested as they already had open accounts on the site, as well as a GoogleGroups R2P page that functions as both a listserv and forum for discussion and posts. The evaluator also became a co-moderator of the Facebook group site. Invitations to join these groups were sent to all former R2P, as well as IRRI and Cornell staff and scientists allowing the formation of larger multi-year course networks. An updated (as of late 2011) database of course participant contact information, collected from the 2011 follow-up surveys, will also be made available to the course directors and training center staff. Participants have consented for this personal information to be made available for the private, internal use of Cornell and IRRI training center staff.

Content and Methodological Guidelines

Content Guidelines

The points listed below refer to major items/aspects which should be considered when reviewing or adding content to the course:

Connect the dots. Participants want and need to know how each aspect of the course content is related to understanding or solving major problems, such as food security, climate change, and community development; either incorporate these as parts of seminars or have participants discuss them.

Be more social. Participants want to see a greater social science component in the class to help them understand the views, research techniques, and outreach and extension components critical

to the adoption of agricultural technology products and varieties.

Utilize all resources. Include and encourage group opportunities in which participants as well as training center staff can share their expertise with each other. Participants are often unaware that their fellows can hold the keys to perspective and understanding which they themselves might lack.

Aim for parity. Avoid the inclusion of outdated or subjective material, research, and techniques, and viewpoints, unless it is for historical perspective or to stimulate discussion. Explain benefits and drawbacks, strengths and weaknesses. Given the wide range of participant expertise, at least some of them will be very aware if a skewed perspective is being presented.

Methodological Guidelines

The points listed below refer to major items/aspects which should be kept in mind when considering changes in course methodology:

Engagement is critical. Participants learn best when given opportunities to actually participate—in discussions, hands-on problem solving activities, and projects—as opposed to being passive listeners.

Less is more. Lecture times should be shortened and enforced at a maximum of one hour, followed by discussion to help participants process and internalize the content. Most scientists are used to tailoring their presentations for a 45 minute to 1 hour period.

Proceed in an orderly fashion. Lectures may be difficult to schedule due to limited scientist availability, but try to ensure that proper foundations are built and content presented in a logical

order, so that participants know, for instance, basic rice plant anatomy prior to doing emasculations and crossing.

Perform introductions. Of participants, staff, scientists, and material. This need not be lengthy or comprehensive, but would help the participants get to know each other, the scientists and staff, people's motivations and backgrounds, and technical material much better.

Overall Recommendations

The following are tables of concrete, suggested modifications to the R2P course which are organized according to a short, mid, or long-term timescale for implementation.

Table 4.2. Short-term modifications which may be put into place for the next course year, organized by category.

Category	Suggested change
Applications	Make sure the goals of the selected participants are in line with the course content
Pre-course	Pre-course survey to ensure participant goals are met
Pre-course	Give background reading material before start of course
Pre-course	Better communication/info before course
Content	Give overview of IRRI divisions, research fields before/at start
Content	Have participants present country/background report in beginning
Content	Discuss IRRI history/mission/approach
Content	More on the social implications of ag. development
Content	Subjectively present a wide range of opinions and perspectives on social development and technological implementation
Interaction/ Relationships	Integrate participant experience/knowledge-sharing as part of discussions/activities
Interaction/ Relationships	Inform speakers of audience backgrounds so they can tailor presentations to a diverse audience
Interaction/ Relationships	Increase and encourage formal and informal opportunities for participants to share knowledge
Structure	Shorten/limit/break-up lecture times
Structure	Give handouts before seminars
Structure	Mix practical sessions with lectures each day
Structure	Schedule activities after lunch/between lectures
Structure	Rearrange topic sequence
Teaching methods	Restructure/reorganize group project components to be more goal driven, perhaps developing real proposals for submission
Teaching methods	Give more problem-solving type activities

Table 4.3. Mid-term modifications which may be put into place over the next 1-2 course years, organized by category

Content	Integrate a holistic, "big picture" perspective throughout course seminars and activities. Focus on application and synthesis of material--
Content	Make presentations progressive and globally relevant
Content	Intro material/sessions needed on genetics/genomics/mol. biology, plant breeding, impact assessment and project design
Content	More social science content
Content	More interaction with different scientists/farmers/inst.
Field trips	Increased stay in Banaue
Interaction/ Relationships	Increase group interaction/discussion
Interaction/ Relationships	More emphasis on building lasting relationships
Participant makeup	Increase global diversity of participants
Teaching methods	More hands-on activities and more time for hands-on activities
Teaching methods	Have daily exercises to synthesize/eval material
Time management	More time with farmers
Time management	More time to digest/synthesize/evaluate info
Time management	More free time to meet with scientists/staff individually
Post-course	Develop an IRRI-based internal networking platform to keep participants in contact with each other and with IRRI
Post-course	Follow-up on the educational and professional progress of participants
Post-course	Develop methods to keep participant learning active after course

Table 4.4. Long-term modifications which may be put into place over the next 2-5 course years, organized by category

Future	Offer participant internship/extended research opportunities
Future	Find specific pools of funding to support all course participants and a wider pool of diverse participants, not just internal candidates

Conclusions

In the first five years of its operation, the R2P course has been largely successful at fulfilling the goals and objectives set by the course directors, and even exceeded some of the goals and expectations of the participants. Participants have themselves generated networking tools to keep in contact with the diverse group of people they have met through the course, and are embarking or continuing on with their careers, most of which are still in plant-related research or agricultural education and extension. In follow-up surveys, participants have reported that the insight, knowledge, and relationships gained in the course have certainly had a profound effect on many of their career paths and personal or professional development. Nevertheless, there are many modifications and improvements that can be made to the course content and structure, as suggested by participants, staff, and scientists, particularly involving better communication and educational methods, and opportunities to continue learning and relationships after the course, such as through internships and an IRRI-centered networking platform.

Whether or not the new relationships, perspectives, and understanding gained from the course help shape the some of the former R2P participants into "a new generation of plant scientists [and agricultural development and extension specialists]" remains to be seen. They will need to not only "understand the importance of innovative plant science in addressing global problems" but become the innovators in charge of addressing those global problems. The improvements suggested in this evaluation can hopefully help ensure that the Rice: Research to Production course, and other courses like it, will continue to be an impactful experience to educate and inspire young scientists and agricultural professionals in the years to come.

REFERENCES

ATLAS.ti. Version 6.2. [Computer software] (2010) Berlin, Scientific Software Development GmbH.

McCouch, S. (2008). Developing Country Collaboration (DCC) Request for Supplemental Extension submitted to the NSF. Unpublished grant proposal, Cornell University.

Rogers, P.J., & Goodrick, D. (2011). Qualitative data Analysis. In J.S. Wholey, H.P. Hatry, and K. E. Newcomer (Eds.). Handbook of practical program evaluation (3rd Ed.), pp. 429-453). San Francisco: Jossey-Bass.

2008 Course Outline. (2008). IRRI Training Center. Unpublished report. International Rice Research Institute.

CHAPTER 5- ADVANCING CULTURAL KNOWLEDGE: EXPERIENTIAL LEARNING IN INTERNATIONAL GRADUATE STUDY TRAINING PROGRAMS FOR THE HEALTH AND STEM DISCIPLINES[1]

Introduction

In today's global society, acquiring cultural knowledge is an essential part of learning in adulthood. Although there is a need to incorporate this understanding in the health and STEM (Science, Technology, Math and Engineering) fields, many such departments do not address this content as part of graduate programs in their fields. One way to rectify this need is through international experiential learning opportunities in a professional, research-centric context. Experiential learning can be one of the most influential ways for adults to embrace this concept as "real" by encountering and learning how to live and work with people whose cultures differ from their own. We define cultural knowledge as ways of knowing and learning that emanate from cultural differences, such as ethnicity, geographic location, religion, political climate, and economic conditions (Regan, 2005; Merriam & Associates, 2007). Experiential learning happens in a number of ways, including learning in informal and formal settings, through conversations among individuals and groups, to studying and working with others who represent cultural groups different than one's own (Merriam, Caffarella, & Baumgartner, 2007; Fenwick, 2008).

The purposes of this paper are twofold: to share a knowledge generated primarily through case studies of health science and STEM graduate programs that have developed international experiential learning opportunities, and to suggests ways in which these programs could be strengthened to encourage the development of cultural understanding in their participants. We have found the most prevalent program models for these training opportunities have been internships, research collaborations, and short courses. We have chosen to study science-related

programs for two reasons: graduates in these fields are moving into a work force that is increasingly international in nature, and there is a great need to train globally and culturally proficient professionals to develop, carry out, and communicate research and practical applications to meet public needs in developing nations. Described first are the methods used to collect data for this study, followed by three case study examples that illustrate these data. Next we address the major findings, and discuss the implications for research and practice.

Methods

For this study, we conducted an in-depth review of the literature on graduate level international training programs in the STEM disciplines and health sciences. Although there is a vast body of literature on an array of undergraduate study abroad programs and their long and short-term impacts, there are few formal publications focused solely on graduate (post-baccalaureate) programs, or even mixed undergraduate-graduate programs.

Of the twenty-odd articles reviewed on STEM and health-related graduate international programs, fully three-fourths of these were either from health or engineering-related disciplines. The overrepresentation of these fields in the literature is likely due to two interrelated reasons. First, engineering and health and medicine are two of the scientific fields with the most recognizable real-world applicability and impact, so they naturally lend themselves to educational, international development projects especially within developing countries lagging in these areas. Second, these fields have a historical commitment to high-quality undergraduate and graduate-level education and a concentration of dedicated teaching faculty who have developed a variety of international experiential learning programs. Furthermore, the faculty members in nursing and engineering are encouraged to publish the results of their programs for educational

and scientific advancement in a number of field-specific international education journals created solely for that purpose.

Conversely, most of the other STEM disciplines, such as biology, chemistry, physics, informatics, and even the applied fields such as ecology and environmental science, and agriculture are less well-represented in graduate educational program literature. This lack of literature is primarily due to complex institutional, funding, and field-specific demands on these faculty members to generate research-based publications for career advancement--a process which is largely reiterated in their education and training of graduate students.

Findings

In reviewing these studies, we discovered that detailed programmatic and/or evaluative publications on international graduate programs are scare and largely limited to those in the health or engineering fields, due to the factors noted above. From those materials we reviewed four program models are prevalent: the sandwich degree, short-course, internship, and individual research. The traditional form of international research experience in the 1970's and beyond were so-called "sandwich degree programs," which involved spending 2-3 years of one's degree doing research or fieldwork outside of the United States (U.S.). New, shorter programs are deliberately crafted to allow graduate students to spend shorter periods of time abroad, usually a few weeks to a few months. This allows for immersive cultural and interdisciplinary learning, and continued progression through a degree program with incorporated global opportunities that are supportive of a student's primary research and/or career interests and graduate requirements.

The most common graduate short-term program models found were short-courses, internships,

and research-abroad programs (Parkinson, 2007; Spencer & Tuma, 2002). The short course is usually a two-week to two-month course developed and facilitated through a partnership between the home university and a foreign university or institution which hosts the course, and often includes participants from outside of the founding institutions. In contrast, internships and research abroad are primarily focused on individual participants and rely on student application, company recruiting, and institutional or lab-specific relationships to facilitate a student doing part of his or her graduate research abroad. Service learning projects are an alternate short-term model common in and largely specific to the health and engineering disciplines, which are best-suited to programs with an immersive, team-oriented framework and concrete, specific goals, such as the construction of houses, or running of temporary clinics. As programs models can be both indicative of a program's rationale, objectives, and goals, as well as key variables, including program duration, size, and partner institutions, the model used can have a huge influence on program outcomes (Spencer & Tuma, 2002), a factor that is discussed in more detail later.

Our most unexpected and significant finding from our literature review was that many of these programs place relatively little emphasis on advancing or evaluating their participants' cultural knowledge within their publications. This lack of acknowledgement does not necessarily imply that graduate participants did not have meaningful cultural experiences or come away with a deeper cultural understanding. Rather, the specific advancement of cultural knowledge may not have been identified as a primary goal of either the program or the publication it which it was featured. Repeated emphasis of case studies on, for example, "international agricultural research" (Phillips et al., 2008), or "a global living laboratory" (Sadjadi et al, 2009) imply that the directors and developers of such programs do intend that their students attain an international research-based cultural understanding, but may be unable or unwilling to use the term "culture,"

perhaps for fear of it sounding too vague or subjective.

The Case Studies

We present three case studies of international STEM and health based educational programs: a short course on agriculture and rice research, a research abroad program in nursing science, and an internship for research experiences in engineering and computer science.

The Rice: Research to Production Short-course.

The "Rice: Research to Production" short-course started in 2007 as a collaborative project between Cornell University and the International Rice Research Institute (IRRI). The three week course, hosted at IRRI in the Philippines, brings together graduate students, junior scientists, and undergraduates from around the world to more fully consider global agricultural and food security issues by engaging in and learning about rice cultivation, agricultural research, and farmer extension (Phillips et al., 2008). Since its initiation in 2007, over 80 participants from 28 countries, including over 30 graduate students, have taken part in the course.

The article featuring this course highlighted two main program objectives: encouraging plant science graduate students from developed nations to consider a career in international agricultural research, and creating an international network of young scientists to help solve food security issues. In addition to fieldwork involving students in traditional and modern rice cultivation, the interaction and relationships built between the diverse group of participants from a range of countries, disciplines, professions, and experiences added an additional, meaningful international dimension.

Field trips to the local rice growing communities and farmer interviews, not mentioned in the article, also were significant cultural experiences (unpublished data). One graduate student indicated that the course "reminded me that I began studying agriculture because of its essential place in supporting lives and societies," while another observed: "it has been inspiring to join the group of scientific contemporaries...who, despite originating from a hugely diverse range of backgrounds have so much in common" (Phillips et al. 2008, p.14-15). Even though gaining cultural understanding was not explicitly defined as a course objective, it is evident that the program still impacted the cultural understanding of these graduate participants.

The Minority International Research Training Program.

The Minority International Research Training Program (MIRT) was initiated in 1994 by the College of Nursing at the University of Illinois at Chicago to give qualified undergraduate and graduate minority nursing students an immersive experience in global nursing research. Each student was paired with a faculty mentor to carry out a 10-14 week research project on a local biological or social health-related issue at a host center in one of nine countries across Asia, Africa and South America, including Malawi, Chile, and Thailand (McElmurry et al., 2003).

MIRT program developers specifically cited cultural sensitivity and the ability to work with people from diverse fields and backgrounds as a requirement for success in international health workers. Increasing the cultural knowledge of the student participants was therefore not only a goal of the program, but a resource in and of itself; program developers placed great value on the broader cultural perspectives gained by both students and faculty for their potential impact on health policy and research. An intimate student/faculty mentor relationship was cited as one of the most influential components of the MIRT program, as faculty mentors were essential in

strengthening their students' skills in cultural assessment and problem solving. The program structure was clearly laid out to facilitate cultural knowledge as an aspect of personal and professional development. Students are expected to show predeparture preparation in cultural diversity and international research ethics, participate in cultural activities during programs, and evaluate their own cultural experiences as part of a post-program report. For their part, faculty and directors are expected to assist in students' cultural development by conducting a predeparture cultural orientation, counseling students dealing with cultural shock, and are involved in pre-and post program assessments of each student. At the time of the 2003 publication, the program was also planning annual follow-up surveys to track, among other aspects, the global and cultural development of past MIRT scholars.

Clearly, this experiential learning nursing program has made cultural knowledge a priority for their graduate participants by embedding it in their planning, implementation, and evaluation stages. While the article did not present any specific results or evaluatory data, if the comment of a student who noted: "My future career in research depends heavily on my objectivity…we must not judge people using our culture as the rule, but accept the various cultural norms and differences," may be taken as evidence that the program has been successful in raising students' cultural knowledge in a professional context (McElmurry et al., 2003, pg 27).

A Global Living Laboratory for Cyberinfrastructure Application Enablement

This internship and collaborative research program for graduate and undergraduate students was initiated by Florida International University (FIU) and Florida Atlantic University (FAU). It is a part of a multinational collaboration among universities, industrial labs, and national research centers in six countries: the U.S., Mexico, Spain, Argentina, China, and India. The program

accepted its first group of 18 students from FIU and FAU in 2008, 12 of whom were graduate students. All participants took part in a pre-travel semester-long cultural and language training program for their country of interest, and then completed collaborative research/internships at an overseas institute or company over the summer.

The semester-long cultural and language training course was the primary cultural education component unique to this program. Although the course was not conducted overseas, it was an experiential learning component that utilized interpersonal and online teaching of "survival vocabulary" including technical terms and useful phrases, and "proper cultural behavior, business etiquette, and differences in manners" (Sadjadi et al., 2009, pg. 68). Each student also worked intensively with at least one faculty member at their home university and a faculty or executive mentor at their international host institute. Faculty developers of this program strongly emphasized a student centric, integrative approach to education, featuring local-international, basic-applied, and academic-industrial linkages. Additionally, students and faculty were chosen for diversity in disciplines, gender, race, and ethnicity, with the aim of building an international network of researchers. While this program is so young that no formal educational evaluation has been published, participant comments indicate that both the mentorship and international research experience affected their cultural growth in that they "develop[ed] quite a bit as a person, researcher, and professional" and that, "being able to interact with new cultures, new people, new places… has enriched my mind (Deng et al., 2009).

Despite fundamental differences in program models, disciplines, components, and objectives, it appears that the three programs presented as case studies managed to achieve definite, although informal, indicators of intellectual and cultural growth from their participants, indicating that no

one model or instructional format could be considered ideal. Even within this limited sample, it is possible to identify some shared elements or considerations important to facilitating personal and professional cultural growth in these science-centric, short-term graduate training opportunities. These include: 1) a recognition, whether explicit or implicit, of the importance of cultural knowledge for graduate students; 2) the development of strong inter-cultural, personal relationships between advisors and students, or within a diverse group of participants; 3)the building of a constructive environment in which these relationships can develop; and 4) a well-designed program framework with built-in support systems for institutional, student, and faculty development and evaluation of cultural knowledge.

Implications for Program Development and Research

Based on the critical factors identified through our review of the literature on graduate training programs in the health and STEM disciplines, we suggest a simple set of general guidelines for integrating and enhancing cultural knowledge through international programs for graduate students in the sciences and health sciences.

- Recognize and include cultural knowledge as a specific program goal.
- Build the program around staff, faculty, host and partner institutions, and student participants who represent and support cultural diversity and intercultural relations.
- Construct educational, research, living, and recreational environments in which
- Participants have the time and space to develop meaningful cultural relationships.
- Require staff, faculty, and student participants to do formal or informal evaluations prior to and during the program, and formal short and long-term surveys after the program.

- Use these reflections and evaluations to determine whether graduate participants have developed their cultural, and make improvements in the program accordingly.
- Publish or otherwise communicate widely the structure and successes of your program to colleagues, educators, and other institutions, so that others in the larger scientific community might also benefit from your experience.

In order to build these international networks and collaborations for graduate students who will become the next generation of global scientists, current researchers must take a step back and work on cataloguing, documenting, and evaluating their own international programs to determine the available resources and how they may be improved. To do so, they may have to reach out beyond their specific disciplines, perhaps to the more experienced program planners in the social sciences or the health and engineering disciplines that already have a history of incorporating cultural knowledge in their graduate training programs. Consultation with specialists in adult education or community development with practical experience in cultural knowledge and experiential international education may also be helpful. Realistically, funding agencies may have to either require or jointly manage and facilitate the development of a network or database to organize websites, publications, and other information from international training and education programs across scientific disciplines and educational levels, which would itself be a valuable resource for researchers, teachers, and students alike.

Further studies must also be conducted to determine how different component of program planning models, such as duration, participant diversity, location, and activities, can affect the development of cultural knowledge. (Caffarella, 2003, 2009; Green & Kreiter, 2004) As the review we present here was a singular cross-sectional assay of the available programs, it would

be interesting to also conduct a series of longitudinal studies following various programs to determine if the factors identified as important in promoting cultural knowledge are similar in both studies. Finally, this study focused mainly on the effect of short-term international graduate programs in sparking greater cultural knowledge and sensitivity among graduate students. A complimentary comparative study evaluating the longer-term international sandwich degree programs and their effect on professional development of cultural mobility and adaptivity would be the next logical investigation in this area of adult experiential education.

REFERENCES

Caffarella, R.S. (2002). *Planning programs for adult learners.* San Francisco: Jossey Bass.

Caffarella, R.S. (2003). Crossing borders: A conceptual framework. Proceedings of the Western Region Adult Education Research Conference, Bellingham,WA, Western Washington University.

Fenwick, T. (2008). Workplace learning: Emerging trends and new perspectives. In S.B. Merriam (Ed.). *Third update on adult learning theory* (pp. 17-26). New Directions for Adult and Continuing Education, 119.

Green, LW., & Kreiter, M.W. (2004*). Health program planning: An ecological approach (4th Ed.).*Boston: McGraw Hill.

McElmurry, B.J., Misner S. & Buseh A. (2003). Minority international research training program: Global collaboration in nursing research. *Journal of Professional Nursing 19* (1), 22–31.

Merriam, S.B., & Associates (2007). *Non-western perspectives on learning and knowing.* San Francisco: Jossey-Bass.

Merriam S.B. Caffarella, R.S., & Baumgartner, L. (2007). *Learning in adulthood. A comprehensive guide*. San Francisco: Jossey-Bass.

Parkinson, A. (2007). Engineering study abroad programs: Formats, challenges, best practices. *Online Journal for Global Engineering Education, 2*(2), 1-14.

Phillips, R., Magor, N., Shires, D., Leung, H., McCouch, S., & Macintosh, D. (2008). Student opportunity: Short-term exposure to international agriculture. *Rice, 1*(1), 11-15.

Regan, T. (2005). Non-western and indigenous educational traditions. Indigenous approaches to educational thought and practice (3rd Ed.). Mahwah, NJ: Laurence Erlbaum.

Sadjadi, S.M., Chen, S., Graham, S., Luis, S., Deng, Y., Furht, B., Martinez, P., Bowen, N., Caraballo, J. (2009). PIRE: A Global Living Laboratory for Cyberinfrastructure Application Enablement. *Proceedings of the 2009 Richard Tapia Celebration of Diversity in Computing Conference,* Portland, Oregon. 64-69.

Deng, Y., Furht, B., Martinez, P., Chen, S., Sadjadi, S.M. (2009). PIRE: A Global Living Laboratory for Cyberinfrastructure Application Enablement. Poster presented at the *2009 Richard Tapia Celebration of Diversity in Computing Conference,* Portland, Oregon.

Spencer, S.E. and Tuma, K. (Eds.). (2002). *The guide to successful short-term programs abroad*. Washington D.C.: NAFSA.

APPENDIX A. POPULATION DYNAMICS AMONG SIX MAJOR GROUPS OF THE *ORYZA RUFIPOGON* SPECIES COMPLEX, WILD RELATIVE OF CULTIVATED ASIAN RICE

HyunJung Kim[1*], Janelle Jung[1*], Namrata Singh[1], Anthony Greenberg[1], Jeff J. Doyle[1], Wricha Tyagi[1,2], Jong-Wook Chung[1,3], Jennifer Kimball[1,4‡], Ruaraidh Sackville Hamilton[5], and Susan R. McCouch[1]

1 Section of Plant Breeding and Genetics, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA
2 *Current address:* School of Crop Improvement, College of PG Studies, Central Agricultural University, Umroi Road, Umiam, Meghalaya, India
3 *Current address:* Department of Industrial Plant Science and Technology, Chungbuk National University, Cheongju, Chungbuk 28644, Republic of Korea
4 *Current address:* Department of Crop Science, North Carolina State University, Raleigh, NC 27695–762, USA
5 TT Chang Genetics Resources Center and International Rice Genebank, International Rice Research Institute, Los Baños, Laguna, Philippines
* These authors contributed equally to this work.

Corresponding Author:
Susan R. McCouch
Section of Plant Breeding & Genetics
School of Integrative Plant Science
Cornell University
162 Emerson Hall
Ithaca, NY 14853
Phone: +1 (607) 255-0420
Email: srm4@cornell.edu

Abstract
Background: The *Oryza rufipogon* species complex (*ORSC*) is the wild progenitor of Asian cultivated rice, *O. sativa* L. Understanding population structure in the *ORSC* is of interest to plant breeders and contributes to our understanding of domestication in rice. A collection of 286 diverse *ORSC* accessions was evaluated for nuclear variation using genotyping-by-sequencing (113,739 SNPs) and indel analysis, and for chloroplast variation using Sanger sequencing (25 polymorphic sites).

Results: Six wild subpopulations were identified by both model-based and distance-based clustering methods, and 25% of accessions were classified as admixed. Three wild groups were genetically and geographically closely related to the *O. sativa* subpopulations, *indica*, *aus* and *japonica* and carried *O. sativa* introgressions; the other three wild groups were genetically divergent, had unique chloroplast haplotypes, and were located at the geographical extremes of the species range. The genetic subpopulations were significantly correlated (r^2=0.562) with traditional species designations, *O. rufipogon* (perennial) and *O. nivara* (annual), historically differentiated based on morphology and life history. A wild diversity panel consisting of 95 purified (inbred) accessions was developed as the basis for future genetic studies.

Conclusions: Genetic relationships between domesticated and wild subpopulations suggest that the cultivated *aus* subpopulation is most closely related to an annual wild relative, *japonica* to a perennial wild relative, and modern *indica* to an admixed population of annual and perennial ancestors. Gene flow between *ORSC* and *O. sativa* contributes to admixture, confounds the interpretation of domestication, and threatens the identity and diversity of wild *ORSC* populations.

Keywords: Population structure, domestication, annual-perennial, chloroplast diversity, phylogeography

Background

The *Oryza rufipogon* species complex (*ORSC*) is the wild progenitor of Asian cultivated rice, *O. sativa* (Oka 1988b; Barbier *et al.* 1991b; Khush 1997b), a crop that provides staple food for three billion people (Elert 2014). Both the *ORSC* and *O. sativa* are widely distributed across South, Southeast and Eastern Asia, but the wild stands exist mostly as small, isolated populations, adjoining or intermingling with cultivated fields (Vaughan *et al.* 2003b). As such, wild stands are threatened by habitat destruction, admixture with *O. sativa*, and genetic erosion (Song *et al.* 2005). Seeds from thousands of crop wild relatives have been collected and preserved in gene banks around the world (Plucknett *et al.* 1983; Tanksley and McCouch 1997; Meilleur and Hodgkin 2004). These collections contribute to the conservation of natural variation, provide the foundation for biological research and insights into the domestication process, and they offer a genetically tractable source of novel variation for breeding (Brar and Singh, 2011; McCouch et al. 2013). Yet little has been done to characterize them genetically or phenotypically (McCouch

et al. 2012b). The lack of information makes it difficult to focus conservation and research efforts, or to utilize these crop wild relatives for variety improvement (Gepts 2006).

Historically, the species found within the *ORSC* are classified as either perennial (*O. rufipogon*) or annual (*O. nivara*), based on morphology, life/mating habit, and the ecological habitat in which they are found. The perennial form, *O. rufipogon*, is reportedly photoperiod sensitive and cross-pollinated; it is aquatic and found in areas with year-round standing water, such as swamps, river beds, and marshes. In contrast, *O. nivara* is considered to be annual, upright, photoperiod insensitive, and predominantly self-fertilized; it is found in seasonally wet habitats such as lake shores and river banks, which undergo periodic flooding with the monsoon rains (BARBIER 1989b; Li *et al.* 2006a; Vaughan *et al.* 2008b). A third designation, *Oryza spontanea,* is a mistaken contraction of *O. sativa* L. f. *spontanea* Roschev which refers to accessions derived from hybrids between *O. sativa* X *O. nivara* or *O. rufipogon* (Morishima *et al.* 1961; Chang 1976; Vaughan *et al.* 2001).

Previous studies have sought to interpret the genetic and geographical relationships among accessions in the *ORSC*, but differences in size of collections, geographical representation of germplasm, and/or marker coverage have led to different conclusions (Wang *et al.* 1992; Cheng *et al.* 2003; Londo *et al.* 2006a; Molina *et al.* 2011; Huang *et al.* 2012b; e; Banaticla-Hilario *et al.* 2013a; Gross and Zhao 2014). In this study, we evaluate a panel of diverse *ORSC* accessions collected from 15 countries, including 56 accessions that overlap with previous reports, using genotyping-by-sequencing (GBS) and indel analysis for nuclear DNA and Sanger sequencing for chloroplast DNA to: 1) characterize the population structure of the *ORSC*, 2) determine the relationship between the subpopulations of the *ORSC* and *O. sativa,* 3) elucidate the relationship between *ORSC* population structure, geographical distribution, annual-perennial life habit (based on traditional species designations), and archaeo-botanical history, and 4) select a subset of diverse accessions as the basis for developing an immortal wild diversity panel for future genetic studies.

Results and Discussion

Population structure and geographical distribution of the *ORSC*

A collection of 286 geographically and genetically diverse accessions from the *ORSC* (Additional File 1: Table S1) was genotyped using GBS to generate a dataset consisting of 113,739 SNPs. Model-based analysis using marginal likelihoods predicted the optimal number of subpopulations to be K=6 (Fig. 1A), but there was little difference between K-values of 5-9 (Additional File 2: Fig. S1). Based on *fastStructure* results at K=6, 25% of the *ORSC* accessions were classified as admixed because they had less than 75% shared ancestry with one of the major subpopulation groups. The subpopulations were identified based on the order in which they diverged from the original population group (W1) with increasing values of K, such that Wild Group 2 (W2) diverged at K=2, W3 diverged at K=3, etc. (Additional File 3: Fig. S2A). When the Neighbor Joining (NJ) method was used to analyze the same data, results were largely consistent with the model-based analysis at K=6 (Additional File 4: Fig. S3).

To determine whether the subpopulation groups identified by *fastStructure* were associated with a nonrandom geographical distribution, we mapped them onto a geographical map of Asia (Fig. 1B-C) and used the Mantel test to evaluate isolation-by-distance. An east-west axis separated the two most geographically isolated populations, W2 (Nepal) and W3 (Papua New Guinea), while a north-south axis (straddling the Himalayan Mountains) separated W6 (China and Taiwan) from a majority of the W1, W4 and W5 subpopulations (SE Asia) (Fig. 1C). W1 was the most widely distributed subpopulation, with accessions geographically co-mingled with other groups across both continental and archipelagic SE Asia. Consistent with its broad geographical distribution, W1 was also the most admixed subpopulation; it shared ancestry with a majority (93%) of individuals classified as admixed in this study (n=71). W2 accessions were also widely distributed across South and SE Asia, but were the predominant group in southern India and Sri Lanka. W3 accessions were found only in the geographically isolated Papua New Guinea region and were not found on the mainland. W4 accessions were widely distributed across SE Asia, extending west into northern India and east into southern China and Taiwan. W5 accessions were mainly from Nepal and western India, and were closely related to W2. W6 accessions were the predominant group in eastern Asia, found mostly in China and Taiwan. Interestingly, of the 16 W1/W6 admixed accessions in our collection, seven were from China or northern Vietnam, and nine were collected in Myanmar, NE India or Bangladesh (Additional File 1: Table 1).

At higher K-values, the emergence of W7 and W8 brought greater geographical definition to the subpopulations identified in SE Asia (Fig. 1A & 1D). At K=7, a cluster of four accessions, previously classified as W1/W5 admixtures, was identified as a subpopulation from Myanmar. At K=8, approximately half of the previously identified W4 accessions along with some admixed W1/W4 accessions, clustered as a separate subpopulation in SE Asia, geographically well differentiated from the remaining W4 samples found in E. India and Bangladesh (Fig. 1D).

Using the Mantel test to determine whether genetic distance was significantly associated with geographical distance, we found a small but significant correlation for the *ORSC* as a whole (not including admixed samples) (r^2=0.10, $p<0.001$) (Additional File 4: Fig. S3). When the Mantel test was run separately on W2, W3 and W5 accessions, the most geographically isolated and least admixed among the *ORSC*, the association between genetic and geographical distance was significantly greater ($r^2 = 0.439$, $p<0.0003$), and contrasted sharply with test results in W1, W4 and W6 accessions, the most widely distributed and most highly admixed subpopulations of the *ORSC* ($r^2 = 0.0531$, $p<0.001$).

Genetic relationship between *O. rufipogon* and *O. sativa*

We next re-analyzed the *ORSC* samples along with 45 *O. sativa* control varieties using Bayesian clustering based on the 113,739-SNP dataset. At K=6, the same *ORSC* subpopulation groups were observed as when the data were analyzed without the *O. sativa* samples, but the cultivated samples allowed us to identify wild populations that clustered with specific *O. sativa* subpopulations (Additional File 3: Fig. S2A). At K=5 or K=6, the W1 population shared >75% ancestry with *indica* (black) accessions, the W4 population with *aus* (orange) accessions, and the

W6 population with *japonica* (blue) accessions (*temperate japonica*, *tropical japonica* and *aromatic*). In contrast, W2, W3 and W5 did not cluster with any of the cultivated groups. These data support the hypothesis that the *aus*, the *indica* and the *japonica* subpopulations of *O. sativa* evolved from genetically distinct *ORSC* lineages. Further, they underscore the finding that the *aus* subpopulation is distinct from both *indica* and *japonica* and represents one of three domestication foci for rice in Asia (Garris et al., 2005; Londo et al., 2006; Schatz et al., 2014; Civáň et al. 2015).

To further examine the relationships between the *ORSC* and *O. sativa*, we compared pairwise genetic distance (GD) and *Fst* values to determine the degree of genome-wide divergence between wild and cultivated groups. These comparisons supported the close relationship between W1 and the *indica* subpopulation, W4 and *aus*, and W6 and *japonica*, while W2, W3 and W5 were maximally differentiated from the *O. sativa* subpopulations (Additional File 5: Table S2).

When the NeighborNet method was used to analyze both wild and cultivated accessions, results were largely consistent with the model-based analysis (Fig. 2). At K=6, *O. sativa, indica* (red) accessions were nested within one of the W1 clusters, *aus* accessions (yellow) emerged from one branch of the W4 cluster corresponding to samples from Bangladesh and India, the *temperate japonica*, *tropical japonica* and *aromatic* subpopulations (shades of blue and pink) emerged from the W6 group with long branch-lengths, and the three independent groups, W2, W3 and W5, were highly divergent based on long branch lengths with strong bootstrap support in the rooted NJ tree. W1 was found at the root position, and clustered with the *O. officinalis* (CC) outgroup, suggesting that the root position is among the W1 lineages. This interpretation was supported by the NJ dendrogram (Additional File 6: Fig. S4) where nearly all groups in the *ORSC* had one or more W1 accessions as their sister group. Further, when the genetic divergence of *ORSC* subpopulations was compared, W1 had the lowest mean pairwise *Fst* and genetic distance (GD) (Additional File 5: Table S2B).

The presence of the *O. sativa* samples in the analysis also revealed increased levels of admixture within the *OSRC*, particularly in the W1 (*indica*-like) and W6 (*japonica*-like) groups (Additional File 2: Fig. S1D). While the cultivated *indica* and *japonica* subpopulations were clearly differentiated from each other, they each shared significant levels of ancestry with both W1 and W6 *ORSC* accessions. This suggested that complex patterns of migration had impacted the geographical distribution of both wild and cultivated groups, offering repeated opportunities for gene flow among and between them over the course of their history. If this were the case, we should be able to document regions of introgression from *O. sativa* in the *ORSC* genome, and vice versa.

To address this possibility, we surveyed the *ORSC* accessions for domestication-related seed and grain phenotypes where the genes underlying those phenotypes had been cloned and characterized, and then analyzed the genomic regions within and around the target genes in *ORSC* and *O. sativa* accessions to determine the origin of the DNA in accessions with wild-type or domestication-related phenotypes. We focused our analysis on two domestication-related

phenotypes that could be measured in seeds, hull color and pericarp color, to determine whether any of the *ORSC* accessions carried white hull and/or white pericarp, traits that were likely to have been inherited from *O. sativa*.

Of the 157 accessions analyzed for these phenotypes, 22 (13%) were found to carry one or both domestication traits (Additional File 7: Table S3). To determine whether the phenotypes were the result of domestication-related mutations, we analyzed DNA samples from a subset of the 22 *ORSC* accessions with white hull or white pericarp and and a control set of 19 black hull, red pericarp accessions representing all wild subpopulation groups to determine whether they carried the wild type allele (conferring color) or the non-functional allele (associated with domestication) at the *BH4* gene (for hull color) and the *RC* gene (for pericarp color). Both genes had been previously cloned and the functional polymorphisms associated with the loss of color in *O. sativa* were determined to be a 22bp deletion in *BH4* (Zhu *et al.* 2011) and a 14bp deletion in *RC* (Sweeney et al., 2007). PCR-based analysis of the 22 white hull and/or pericarp accessions and the set of 19 control demonstrated that all but one of the *ORSC* accessions with white hull and/or white pericarp carried the knock-out mutation associated with domestication – accession NSF_ID 474 had white hull color, but a wild-type non-deletion *Bh4* allele. All but two of the *ORSC* accessions with black hull and red pericarp carried the wild type alleles; the exception being NEF-ID 540 and 460, both of which had black hulls but carried the 22bp deletion Bh4 allele (Additional File 1: Table S1). The discrepancies are likely due to the use of plant tissue from the Biobank in the McCouch lab and the heterogeneity and heterozygosity of seed stocks, which is a common occurrence in *ORSC* accessions. It is noteworthy that all *ORSC* accessions carrying the domestication-related alleles corresponded to W1, W6 or admixtures containing one or both of these subpopulations. This is consistent with the high levels of admixture observed in these two *ORSC* groups, and the low *Fst* and GD statistics summarized in Additional File 5: Table S2A.

To further confirm the origin of the domestication-related traits in *ORSC* accessions, we analyzed the SNP haplotypes surrounding the *RC* gene using ancestrally informative polymorphisms (Sweeney et al., 2007; Kovach et al., 2009; Lam et al., 2010; Takano-Kai et al., 2011). For this analysis, we included the same set of *ORSC* accessions that had been phenotyped and genotyped for the functional indel polymorphisms described above as part of a larger subset of 81 *ORSC* accessions (Fig 3). We observed that the *ORSC* accessions carrying the knock-out (14-bp deletion) allele at *RC* carried an *O. sativa* extended haplotype around the *RC* locus while accessions carrying the wild type allele carried an *ORSC*-specific haplotype around *RC.* (Fig. 3, Additional File 8: Table S4). We also find that pericarp color scores for haplotype groups 2, 8, 10, 11 and 12 with the W7 allele in the two SNPs within RC (Fig 3) are higher (more red) than those in haplotypes 1, 3, 4, 6 and 7, which all carry cultivated allele at both SNPs (Additional File 9: Figure S5). This analysis supports the conclusion that the presence of domestication – related phenotypes in *ORSC* accessions are the result of gene flow and introgression from *O. sativa,* rather than standing variation in the wild.

Comparison of subpopulation and species classification

Several different species names are used by gene banks to refer to accessions within the *ORSC*. When the six wild subpopulations identified in this study were analyzed in relation to the two primary species designations, *O. rufipogon* (perennial) and *O. nivara* (annual), we observed a significant correlation (r^2=0.562; Chi-square p<0.0001) (Additional File 10: Table S3). Ninety one percent of W1, 100% of W3 accessions, and 50% of W6 accessions were classified as *O. rufipogon*, while a majority of W2 (56%), W4 (64%), and W5 (83%) accessions were classified as *O. nivara* (Fig. 1A). Both species were found throughout mainland SE Asia, but *O. rufipogon* was predominant in the Indonesian archipelago and in China, while *O. nivara* was concentrated on the Indian subcontinent and across SE Asia (Additional File 11: Fig. S6). Cultivated *aus* shares most recent ancestry with annual forms of W4, *japonica* with perennial forms of W6, and the *indica* subpopulation, which is geographically the most widely distributed, is closely related to forms of W1 that show admixture with W4 on the one hand, and W6 on the other (Fig. S2B). This ancestral dichotomy, where both annual and perennial ancestors are recombined in W1 accessions, undoubtedly contributes to the high levels of recombinational diversity and broad adaptation observed within the *indica* subpopulation (Garris *et al.* 2005b; Huang *et al.* 2012a).

This is the first report documenting the highly admixed nature of the most recent wild ancestor of *indica* (W1) where significant admixture is observed between W1 and the annual *aus*-like ancestor, W4 in India, Bangladesh and SE Asia, as well as between W1 and the perennial, *japonica*-like ancestor, W6 across SE Asia and into southern China. In this study, *ORSC* samples collected from Guangdong and Guangxi in southern China were related to both *indica* and *japonica*, while samples collected north of the Nanling mountains, in the central sub-tropical zone, were most closely related to *japonica*, consistent with previous reports (Wang *et al.* 2008c). The idea that *indica* evolved as a complex derivative from divergent ancestral groups is parallel to the scenario recently reported for barley (Pourkheirandish *et al.* 2015) but with the added dimension of coalescing annual and perennial life habits.

The 18 *O. spontanea* accessions shared >75% ancestry with individuals in diverse subpopulations; half the samples were classified as W6, 22% as W1, 17% as W4, and 11% as W5, and one as an admixture (W1/W4) (Additional File 10: Table 5 and Fig. S7). Because they did not cluster into a single genetic group, nor were they generally diagnosed as admixtures, we conclude that the species classification for these samples should be reconsidered, given that it would be more informative to identify each sample in association with its most closely related wild subpopulation.

Chloroplast haplotype network

We assayed chloroplast sequence from five different regions of the rice chloroplast genome and identified 59 haplotypes among 268 *ORSC* accessions, 44 *O. sativa* accessions, five AA genome wild accessions and three non-AA genome outgroups. We generated a statistical parsimony haplotype network from these haplotypes, which clustered them into eight chloroplast groups (*cpGroup I – VIII*) (Fig. 4; Fig. 5; Additional File 1: Table S1). Not surprisingly, haplotypes from many of these groups were found in W1 individuals, consistent with nuclear data in suggesting that W1 comprises an ancestral, genetically diverse subpopulation; admixed

individuals were similarly variable, sharing haplotypes from different subpopulations presumably due to gene flow. Excluding W1 and admixed individuals, there was good correspondence between chloroplast haplotype groups and subpopulations, particularly wild subpopulations: *cpGroup IV* was unique to W3, and *cpGroup VI* was unique to W5 accessions. These chloroplast haplotypes provide evidence of distinct maternal lineages in wild subpopulation groups and lend support to the results of the *fastStructure* analyses. At the same time, several haplotype groups were shared by several different wild and cultivated subpopulations, suggesting gene both ancient and (in the case of cultivated accessions) more recent gene flow (Fig. 5: note *cpGroups I, III,* and *VIII*).

Haplotypes of outgroups (*O. officinalis* (CC) and *O. australiensis* (EE)) were very distinct from those of *ORSC.* The outgroup haplotypes joined the network at *cpGroup V*, a haplotype found almost exclusively in W1 and admixed individuals, further supporting the ancestral nature of the W1 group. The network had several loops; given the historically non-recombining nature of the chloroplast genome, loops are interpreted as being due to substitutional parallelisms and reversals rather than to recombination. This reticulate structure complicates interpretation of the network; however, outgroup rooting clearly split the network into two large groups strongly associated with the two major *O. sativa* varietal groups, *JAPONICA* (*tropical japonica, temperate japonica, aromatic*) and *INDICA* (*indica, aus*), referred to as *cpGroup I* (or the *JAPONICA-cpGroup*) and *cpGroup VIII* (or the *INDICA-cpGroup*), respectively. *cpGroup I* haplotypes were found in 87.5% of cultivated *japonica* cultivars and 58.8% of W6 accessions, the most closely related ancestral group, while *cpGroup VIII* haplotypes were found in 77.8% of cultivated *indica,* 80% of cultivated *aus* cultivars, and only 47.6% and 48.4% of the related W1 and W4 accessions, respectively. The divergence in frequency between these two chloroplast groups is not as obvious in the *ORSC* accessions as it is in the *O. sativa* groups. This is consistent with the results of the Mantel test suggesting that geographical dispersion of *ORSC* populations and admixture with *O. sativa* (particularly for S1, W4 and W6) has eroded the genetic composition of the ancestral populations from which *O. sativa* was originally domesticated.

Along one path from the outgroup to the *JAPONICA-cpGroup I,* the first group of accessions to diverge was *cpGroup IV,* found primarily in the geographically isolated W3 accessions from Papua New Guinea and Australia and the closely related AA genome species, *O. meridionalis*. Along the alternative path toward *JAPONICA- cpGroup I,* the *cpGroup III* diverged; this group was most common in admixed and W1 individuals. In the other half of the network, along the path leading to the *INDICA-cpGroup VIII* were *cpGroups VI* and *VII*; haplotypes of the former group were found exclusively in individuals of subpopulation W5, from Nepal (colored light green), whereas haplotypes of the latter group were found only in W1 accessions (Fig. 4; Additional File 12: Figure S8).

The *aus* and *indica* subpopulations of *O. sativa* both have primarily haplotypes from *cpGroup VIII* suggesting that they share a more recent maternal ancestor than either does with *japonica*, although both include individuals with *cpGroup I* haplotypes, shared with *japonica* individuals. This is consistent with previous findings (Garris et al. 2005; Londo et al., 2006). Interestingly,

the analysis also supports the conclusion that when hybridization occurred between early *indica, aus* and *japonica* domesticates, individuals from the *indica* and *aus* subpopulations were more likely to have served as the maternal parents. This is based on the observation that *indica* and *aus* varieties carry *cpGroup I* (*JAPONICA)* haplotypes at frequencies of 10% (aus) or 22% (*indica*), while none of the *japonica* cultivars carried *cpGroup VIII* (*INDICA)* haplotypes.

We next examined specific chloroplast sequence polymorphisms that were shared between *ORSC* and *O. sativa* (Fig. 4; Additional File 13: Table 6B). One of the *indica/aus*-specific derived variants corresponds to a 69bp deletion (#6) which is widely used to differentiate *japonica* (ancestral, non-deletion type) from *indica/aus* in phylogenetic studies (Kanno *et al.* 1993; Garris *et al.* 2005b). In addition to the 69bp deletion, we discovered a single derived SNP located inside the indel (at 8,599bp) that was found in non-deletion types, predominantly in *japonica* ("G"), while the ancestral SNP ("A") was exclusively found in all out-groups and other AA genome species (Additional File 13: Table S6B). Within the *ORSC*, two geographically divergent subpopulations, W3 (from Papua New Guinea) and W5 (from Nepal) both harbored the "G" SNP within the non-deletion allele (at frequencies of 100% and 90.0%, respectively), while the rest of the wild subpopulations collected across South and SE Asia and southern China, contained a mixture of all three chloroplast genotypes: 69bp non-deletion type with SNP-A, 69bp non-deletion type with SNP-G, and the 69bp deletion type.

The fact that chloroplast haplotype patterns are not identical to the nuclear genome groups in either wild or cultivated rice is not unexpected; rather it underscores the complex population dynamics in both the *ORSC* and *O. sativa*, where deep coalescence (incomplete lineage sorting) and recent hybridization (admixture) both play a role. Because these two processes produce the same signature of incongruence, it is difficult to disentangle them or to accurately interpret the timing of events that contribute to the patterns of diversity among and between populations.

Development of Wild Rice Diversity Panel (W-RDP)

Based on these studies of nuclear and chloroplast variation, 95 *ORSC* accessions were selected to represent the major subpopulation groups as part of the Wild Rice Diversity Panel 1 (W-RDP1) (Fig. 1A; Additional File 1: Table S1). As the basis for replicated phenotypic evaluation and genome wide association mapping, a single individual from each accession was selfed for three generations to genetically purify the lines. Seed production in the greenhouse on these wild, shattering plants was very limited in the Ithaca environment, and with successive generations of inbreeding, there was a noticeable reduction in the quantity and quality of seed set on many of the plants, most notably those in the W3 subpopulation. The result was that none of the W3 individuals generated viable S_3 seed. Nonetheless, we were able to generate S_3 seed on a diverse collection of 95 *ORSC* accessions representing the W1, W2, W4, W5 and W6 subpopulations. These purified (self-pollinated) seed stocks represent a valuable genetic resource as the basis for future genetic studies in this crop wild ancestor.

Evolutionary history and population dynamics

To gain further insight into the evolutionary history and population dynamics of the wild subpopulations, we compared levels of nucleotide diversity (π) and linkage disequilibrium (LD) decay among groups. Of the wild accessions not closely related to any cultivars, W3 and W5 are found at the two opposite extremes of the geographical range of the *ORSC* and behave as expected for small isolated populations: their within-population diversity is low, and divergence from all other groups is high, likely due to a combination of genetic drift and local adaptation (Additional File 14: Fig. S9). However, these two populations are distinguished by their levels of LD (Fig. 6; Additional File 15: Table S7); the population from Papua New Guinea, W3, contains individuals that are exclusively classified as *O. rufipogon* using the traditional annual-perennial nomenclature system, and has relatively rapid LD decay, consistent with the out-crossing nature that is characteristic of most perennials, while W5 (mainly from Nepal) has >80% of individuals classified as *O. nivara* and maintains LD over larger distances than any other subpopulation, in keeping with its predicted inbreeding habit.

Population W2 is unusual. It is the first group to be differentiated from W1 in *fastStructure* analysis, its level of nucleotide diversity (π) is the highest of all populations, yet it has extensive LD (Additional File 14: Fig. S7; Fig. 6). This suggests that while the effective population appears to be large, there is not much recombination among individuals. Similar to W5, W2 accessions are predominantly identified as *O. nivara*, which suggests a high level of self-pollination, but W2 is more widely distributed geographically, being abundant in eastern India and isolated parts of southern India and Sri Lanka. This raises interesting questions about the potential for the annual habit to have arisen multiple times in response to diverse climatic factors across a broad geographical range. We hypothesize that the high level of π, combined with the extensive LD observed in the W2 population may be the result of a rapid evolutionary process that favored survival of numerous geographically dispersed and genetically isolated populations that were independently able to transition to an annual, inbreeding habit in response to a dramatic change in climate, such as that which has been described as global warming at the end of the Pleistocene era (Fuller et al., 2010).

The W4 subgroup, with its high estimates of π, rapid LD decay, and its distinctive relationship with the *aus* subpopulation, is also predominantly comprised of *O. nivara* accessions, again suggesting a strong annual growth habit. W4 is distributed throughout Bangladesh, northern Myanmar and Eastern India (Khush 1997b; Garris *et al.* 2005b; Londo *et al.* 2006a). It's distinctive subpopulation structure offers further evidence that the annual growth habit may have evolved multiple times from different ancestral populations. The W4 subgroup and its *aus* relatives are increasingly recognized as a source of unique, stress-tolerance traits of interest to plant breeders for developing new, climate-resilient rice varieties. With its unique geographic, genetic and ecological history, the cultivated *aus* subpopulation and its wild ancestors (W4) represent an underappreciated genetic resource.

W6 represents a group of *O. rufipogon* accessions collected in China and Taiwan, the presumed center of domestication for the *japonica* subspecies of *O. sativa* (Londo *et al.* 2006a; Kovach *et al.* 2009; Huang *et al.* 2012e). This group has low to intermediate levels of π and LD decay,

consistent with its recent expansion into the temperate region in eastern Asia, the northern-most tip of the zone inhabited by the *ORSC*. Low diversity would be expected at the forefront of a range expansion or in isolated colonizing groups, as is the case for *temperate japonica*. Some wild diversity, particularly the ancestral populations from which the earliest *japonica* cultivars were domesticated, has surely also been lost as human civilization encroaches on its habitat (Song *et al.* 2005). W6 samples from southern China were more likely to share ancestry with W1 wild accessions than were samples from farther north, contributing to the loss of identity of the ancestral *japonica* gene pool (Wang et al., 2008).

Within the *ORSC*, W1 is a heterogeneous group that is at the center of the network of relationships (Fig. 2). It has the most diverse representation of chloroplast haplotypes, the most rapid LD decay, and is geographically the most widely distributed wild subpopulation. It has hybridized extensively with several other groups to produce admixed individuals. The geographic distribution and genetic closeness of W1 to other wild and domestic populations suggest the possibility that it may be ancestral to the entire *ORSC*. Under this scenario, it is interesting to speculate how ecological, genetic, and climatic changes may have contributed to the differentiation of the other groups.

The surprising observation that W1 has only intermediate π (Additional File 14: Fig. S9) suggests that, rather than being ancestral to the entire *ORSC*, it may actually be a product of secondary hybridization between an assortment of wild and cultivated populations. A high level of admixture is characteristic of a majority of *ORSC* gene bank accessions. While exhibiting numerous "wild" phenotypic characteristics, these accessions also carry numerous "cultivated" alleles inherited from *O. sativa* , as demonstrated for hull and pericarp color in this study. The value of the W1 population for plant breeding is that it provides a wealth of recombinant options whereby natural variation, including valuable forms of disease and insect resistance, abiotic stress tolerance, and grain quality traits, has been massively shuffled and exposed to both natural and artificial selection over many thousands of years. Thus, valuable allele combinations conferring adapted networks of quantitatively inherited traits of interest to breeders are likely to be found uniquely in W1 accessions.

Climate and species range

The current range of the *ORSC* extends across a northwest (W2 and W5) to southeast (W3) axis, with the subpopulations most closely affiliated with *O. sativa* (W1, W4, W6), bracketed by those extremes and oriented along a north - south axis. (Fig. 1C). This observation is consistent with Fuller *et al.*'s (2010) hypothesized climate-based shifts in the ranges of ancestral wild rice habitat since the Pleistocene. This hypothesis asserts that 20,000 years ago, during the Last Glacial Maximum, wild rice populations were limited to wet tropical refugia such as Eastern India, Southern China, and continental Southeast Asia, which extended down into the then-interconnected northern Indonesian peninsula. Subsequent changes in climate, characterized by increased temperatures, a rise in atmospheric CO2, and periodic dry seasons followed by monsoon rainfalls helped to expand the range of the *ORSC* and alter the population dynamics. Increasing temperatures in the northern hemisphere would be predicted to support the expansion of wild rice populations northwards, consistent with the identification of the W6 subpopulation located as far north as the Yangtze River basin in China and the W5 subpopulation in the

highlands of Nepal. The emerging monsoon climate with its long, hot, dry summers, particularly pronounced on the Indian subcontinent and across into SE Asia, would have selected for new, wild, annual forms of *O. nivara*, such as those observed in the dispersed W2 subpopulation in this study. In the southernmost ranges, rising sea levels would have inundated low-lying land bridges and created islands of reproductively isolated *ORSC* populations, consistent with the W3 subpopulation documented from Papua, New Guinea. Into this scenario of wild rice population dynamics, humans began to experiment with early domestication efforts, introducing an additional agent of change that contributed to population movement and helped to obfuscate the wild subpopulation structure that once existed across South and SE Asia. While our study detects the impact of these events, documented in the observed patterns of admixture, we make no claims as to the timing of population expansion because it is unclear how biases in calling SNPs from GBS data would affect the site frequency spectrum and thus obscure any demographic signal.

Geographically isolated *ORSC* populations provide a unique opportunity to document the genetic composition of ancient subpopulations of wild rice. In this study, we document an unusual case of a chloroplast haplotype shared between accessions of W3 (Papua, New Guinea), W5 (Nepal) and two outgroups, *O. officinalis* (CC-genome) and *O. australiensis* (EE-genome), suggesting the possibility that the geographically isolated W5 and W3 subpopulations may have radiated from a common ancestor at about the same time. Isolated populations such as these that survive in natural refugia are of great interest for genetic studies and pre-breeding applications in rice improvement because they are likely to harbor variation rarely seen in cultivated rice. They also warrant special conservation efforts because they are increasingly threatened by habitat destruction.

Research aimed at exploring the diversity and population structure of other *Oryza* species, particularly those native to Australia and New Guinea, is of interest to expand our understanding of both the AA genome and more distantly related *Oryza* relatives that exist in isolated populations in that part of the world (Waters *et al.* 2012; Sotowa *et al.* 2013). In this study we found an Australian accession of *O. rufipogon* corresponding to subpopulation W3 that shared a chloroplast haplotype with three *O. meridionalis* accessions, suggesting either shared ancestry or gene flow between the two species (Cai *et al.* 2008). Such findings can help clarify the evolutionary history of the *Oryza* genus.

Reports of admixed accessions being found far from the geographical regions occupied by their immediate ancestors support the idea that small subsets of the *ORSC* likely traveled (and continue to be moved) along with cultivated *O. sativa* in the form of mixed/contaminated seed lots through commercial trade and human migration. This, along with back-introgression from *O. sativa* to *ORSC* in the field, could explain the presence of such geographically unexpected admixed subpopulations. The fact that W1/W6 admixed accessions are found in eastern China and as far west as NE India is consistent with dissemination by humans and with genetic and archeological evidence documenting hybridization between *japonica* rice from Southern China and proto-*indica* rice in North India (Fuller 2011). In addition, there are several reports of key

domestication traits being introgressed from domesticated *japonica* varieties into *indica* (Sweeney *et al.* 2007; Takano-Kai *et al.* 2009; Kovach *et al.* 2009; Yang *et al.* 2011). These observations suggest that humans have contributed to the complex hybridization and introgression patterns observed in the *ORSC* over thousands of years and across a wide geographical range. Further, in this study of the *ORSC,* we see that humans have left their mark not only on the populations they domesticated, but also on the wild relatives they left behind.

Conclusions

Six wild subpopulations were identified in a collection of 286 diverse *ORSC* accessions originating from 15 countries. Three of the wild groups were genetically and geographically closely related to the three major *O. sativa* subpopulations, *indica*, *aus* and *japonica,* while three other wild groups were genetically divergent, each with unique chloroplast haplotypes. The three divergent wild subpopulations were located at the geographical extremes of the species range, while the wild relatives most closely related to *O. sativa* were located across S. Asia, continental SE Asia, and southern China and shared significant levels of admixture with each other and with *O. sativa*. A significant correlation was observed between the *ORSC* subpopulations defined based on molecular variation in this study and the two traditionally recognized species groups, *O. rufipogon* (perennial) and *O. nivara* (annual), classified based on morphology, mating habit, and ecological habitat. Our results suggest that the cultivated *japonica* subpopulation derives from a perennial ancestor, the *aus* subpopulation from an annual wild relative, and that *indica* is the result of admixture between divergent annual and perennial wild ancestors. Our findings are consistent with the hypothesis that the annual habit likely arose multiple times in response to diverse climatic factors across a broad geographical range. Understanding the relationship between subpopulation structure, ecology and geography is crucial for breeding programs seeking to harness the wealth of natural variation that resides in crop wild relatives. As part of this study, we also developed a wild diversity panel consisting of 95 purified (inbred) accessions representing the range of variation in the *ORSC* as the basis for future genetic studies.

Materials and Methods

Germplasm

Seeds from 286 *ORSC* accessions were imported from the International Rice Germplasm Collection (IRGC; n=283) at the International Rice Research Institute in the Philippines and from the National Institute of Genetics (n=3) in Japan (Additional File 1: Table S1; Additional File 6: Fig. S4). Fifty accessions of *O. sativa* from the Rice Diversity Panel 1 (RDP1) (Eizenga *et al.* 2014) were used to evaluate the relationship between wild and cultivated rice (Additional File 1: Table S1).

Phenotyping

Hull and pericarp color phenotyping

Hull and pericarp color were phenotyped on all 286 *ORSC* accessions grown out at the Guterman Bioclimatic Laboratory from 2006-2007. Hull color and pericarp color were scored on three seeds each, produced by three individuals from each accession. Hull color was scored as follow: black hull-8.0; white hull-1.0. Pericarp color was scored as follow: red pericarp = 1.0, white

pericarp = 0. Hull and pericarp color scores were then averaged across individuals to determine the accession mean for each trait. Only 157 of these 286 accessions with complete data for both hull and pericarp color were included in the final analysis.

Genotyping

DNA extraction:

Young leaf tissue was collected from single plants for DNA extraction using a modified potassium acetate-SDS protocol (Dellaporta et al. 1983) and DNeasy Plant Mini Kit (Qiagen).

Genotyping-By-Sequencing:

96-plex GBS libraries were prepared using the *Ape*KI restriction enzyme; libraries were sequenced using an Illumina HiSeq 2500 (Elshire *et al.* 2011). SNP calling and filtering was done using the Tassel 3 GBS Plugin (Glaubitz *et al.* 2014). The sequence tags were aligned to the Nipponbare reference genome (MSU v6) using Bowtie2 (Langmead and Salzberg 2012). A set of 113,739 SNPs with call rates greater than 50% per SNP locus (average 72%) and with Minor Allele Count (MAC) >3 well distributed across the *ORSC* and *O. sativa* genomes were used for analyses of wild materials. More detailed information about Materials and Methods is provided as Additional File 16.

Chloroplast markers:

Sequence information for two EE genome and four AA genome wild control accessions were selected from Genbank; two *O. australiensis* (GU592209 and KJ830774, EE genome), three *O. meridionalis* (NC_016927, JN005831, and GU592208, AA genome) and one Australian *O. rufipogon* (JN005833, AA genome). Sequence data were aligned to the reference genome, NC_001320, implemented by Geneious v7.1.7.

A total of 36 sequence variants were selected from 4,127bp of concatenated chloroplast sequence representing 5 different regions in the *O. rufipogon, O. sativa, O. meridionalis, O. officinalis* and *O. australiensis.* Of these, 25 variants were polymorphic within *ORSC* (Additional File 10: Table S5) and were selected for diversity analysis (Additional File 6: Table S4A). Chloroplast sequence data were generated as described in Kim *et al.* (2014) (Kim *et al.* 2014).

Data Analysis

Nuclear data

***Population structure and genetic relationships*:** Population structure was investigated using *fastStructure* with a simple prior (Raj *et al.* 2014) and visualized in *distruct* (Rosenberg, 2004). The range of optimal K (number of populations) values to be tested was determined based on model complexity using marginal likelihood and model components to explain the structure of the data. Genetic relationships were also investigated as a network using an unrooted Neighbor Joining (NJ) algorithm implemented in SplitsTree v4 (Huson and Bryant 2006) and a rooted NJ dendrogram with 100 bootstrap replicates in Geneious v7.1.7. Genomic diversity between individuals and subpopulations was visualized based on NJ genetic distance as a heatmap using the devtools package in R 3.0.1. The chi-square statistic, implemented by JMP Pro V10 (SAS

Institute Inc.), was used to determine whether the subpopulation designations for the *ORSC* accessions based on GBS data corresponded to taxon names used by the IRGC.

Isolation by distance:
The relationship between geographical and genetic distance was analyzed based on Mantel's test using Isolation By Distance v3.23 (Jensen *et al.* 2005) with 1,000 randomization cycles. Nineteen accessions from China with unknown geographical location within the country were excluded from these analyses.

Calculation of Fst, π, and d:
Pairwise *Fst* statistics among subpopulations were calculated based on the average value over non-overlapping sliding windows of 100 SNPs across the genome with 95% empirical Confidence Interval (CI) (Weir and Cockerham 1984). Using the same 100 SNP windows, we calculated π and *d*. We enumerated the sequence differences between a given pair of DNA segments and calculated sequence differentiation using the Jukes-Cantor model (Li 1997). Genetic distances between population pairs and nucleotide diversity within populations were estimated based on NEI (1973). For estimates of within-population π for *ORSC* populations, we used the full set of 113,739 SNPs; for calculating each pairwise genetic distance, only polymorphic SNPs were used. To enable comparisons between different analyses, we estimated per-kb values of π and *d* by dividing the total value for a window by the reference map distance (in kb) between the first and last SNP.

***Haplotype analysis*:**
Extended haplotypes spanning a 580 kb region flanking the *Rc* locus on chromosome 7 were constructed on 81 *ORSC* accessions from the Wild Rice Diversity Panel (WRD-P) (Table S1) and on 406 *O. sativa* accessions from RDP1 genotyped with the HDRA (McCouch *et al*, 2016). The HDRA carries a total of 1021 SNPs in the 580 kb region. SNPs with a MAF > 0.05 and < 3% missing data were initially selected. SNPs were then filtered based on a frequency test; only SNPs with a significant frequency difference between *O. sativa* accessions with white pericarp and *ORSC* accessions with red pericarp (P value cutoff: 1.0e -05) were used. The final set of SNPs used to construct the haplotype map in Fig. 3 consisted of 40 SNPS (Additional File 17: Table S8).

***Linkage disequilibrium*:**
Linkage disequilibrium (LD; estimated as r^2 between SNPs) within populations was calculated in 10 Megabase windows using Plink v1.9 (Purcell et al., 2007). We retained SNPs with no more that 30% missing data and at least two individuals carrying the minor allele. Raw pairwise estimated were binned by distance range. We present the LD estimates as means within a bin. Because W5 was the smallest population (with 12 samples), we sub-sampled 12 accessions from the other groups 100 times each and re-ran the LD analyses to account for any effect of sample size on the r^2 statistic. Figure 5 thus shows a mean and 95% confidence interval of LD decay rates for each population, with the exception of W5 which has not been sub-sampled and thus has only one value per distance bin.

PCR analysis of Rc and Bh4 indel polymorphisms
PCR primer pairs were designed to amplify a 236bp region spanning the functional 14bp indel of *Rc* (Sweeney et al., 2006) and a 227 bp region spanning the functional 22bp indel of *Bh4* (Zhu et al, 2011), with product sizes optimized for indel resolution via agarose gel electrophoresis. DNA was extracted from tissue of 41 *ORSC* accessions from McCouch lab W-RDP biobank samples. PCR was done with a Tm of 56°C for BH4-M22 primer set and 57°C for Rc-1 primer set. Reactions were run out on a 5% agarose gel for 3 hours and scored. Primer sequences are as follows: BH4-M22F 3'-TCTGGTGCATAATCAGAATGG-5'; BH4-M22R 3'-TCGTGTATATGGCGACCTTG-5'; Rc-1 F 3'-CTTGCCAGTTTCAGAGAAATCA-3'; Rc-1 R 3'-CTCTTTCAGCACATGGTTGG-5'

Haplotype analysis across Rc region
EHs spanning a 580 kb region flanking the *Rc* locus on chromosome 7 were constructed on 81 *ORSC* accessions from the Wild Rice Diversity Panel (WRD-P; Table S1) and on 405 *O. sativa* accessions from RDP1 genotyped with the HDRA (McCouch et al, 2016). The HDRA carries a total of 1021 SNPs in the 580 kb region. SNPs with a MAF > 0.05 and <3% missing data were initially selected. SNPs were then filtered based on a frequency test to include only the SNPs with a significant frequency difference between *O. sativa* accessions with white pericarp and *ORSC* accessions with red pericarp (P value cutoff: 1.0e -05). The final set of SNPs used in constructing the haplotype map in Fig. 3 consisted of 40 SNPs.

Chloroplast data
A statistical parsimony haplotype network was generated for 268 *ORSC* accessions, 44 *O. sativa* accessions, five AA genome wild accessions and three non-AA genome outgroups, one *O. officinalis* (CC) and two *O. australiensis* (EE), based on chloroplast sequence information using TCS v1.21 (Clement *et al.* 2000), implemented by POPART (Leigh and Bryant 2015). Sequence data were aligned to the reference genome, NC_001320. Every polymorphism was given the same weight, under the assumption that each represented a single evolutionary event. Chloroplast groups were defined as a continuum of haplotypes at 97% parsimony connection (Ray *et al.* 2013) and haplotypes not belonging to any cpGroup were considered independent haplotypes (*ln*).

List of Abbreviations
ORSC: *Oryza rufipogon* species complex; GBS: Genotyping-by-sequencing; *cpGroup*; Chloroplast group; W-RDP: Wild Rice Diversity Panel; π: Nucleotide diversity; LD: Linkage disequilibrium; IRGC: International Rice Germplasm Collection; MAC: Minor Allele Count; NJ: Neighbor Joining; CI: Confidence Interval

Competing Interests
The author(s) declare that they have no competing interests.

Availability of Data and Materials
Germplasm:

All rice accessions have an International Rice Genebank Collection (IRGC) number or Wild Identification Number as described in Additional File 1: Table S1. The wild rice accessions from diverse species used for generating the chloroplast sequence information are in Table S1C.

Genotype:
1) Chloroplast sequence data is in process of submission to Genbank. Additional File 18: Table S9 contains chloroplast SNP locations.
2) Nuclear SNP data was submitted to NCBI in April 2016 and has been assigned NCBI Batch ID 10632455. Accession ID's will be provided as soon as available.

Author's Contributions
HJK, JJ, AG and SMc wrote the manuscript. RSH advised on selection of wild accessions from the IRGC. NS generated GBS data. WT, JWC, and JK generated nuclear SSR data. HJK generated chloroplast sequence data, conducted genetic analyses. AG generated π, d, and Fst statistics. JJ purified the wild diversity panel and conducted the *BH4* and *Rc* phenotype, genotype and haplotype analysis. WT, JK and NS managed *ORSC* plants and DNA samples. JJD and AG provided critical interpretive insights re evolutionary analysis. SMc conceived of and provided overall intellectual guidance for the project.

Acknowledgements
We gratefully acknowledge Sandra Harrington for managing tissue and seed stocks of the wild *Oryza* germplasm, Yuxin Shi for submission of data to NCBI, Genevieve DeClerck and Sandra Harrington for developing the Laboratory Information Management System (LIMS) for the McCouch lab, Fumio Onishi for greenhouse support, Namrata Singh for designing the primers used in the Bh4 and Rc gene indel analysis, Kyeong Oh Kim for Python and R consulting, Diane Wang, and Margaret Smith for constructive comments and help editing the manuscript, Sang-Nag Ahn for historical interpretation of population structure in the *ORSC*, and Jeanne Kisacky for help with formatting. The wild rice accessions used in this study were distributed from the International Rice Genebank at the International Rice Research Institute in the Philippines and from the National Institute of Genetics supported by the National Bioresource Project, MEXT, Japan. This project was supported by the National Science Foundation (NSF) with a grant from the Plant Genome Research Program, Award #0606461 and #1026555 to SMc, and a grant from the Government of Norway entitled 'Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives' implemented in partnership with the Global Crop Diversity Trust, the Millennium Seed Bank of the Royal Botanic Gardens, Kew Gardens, national and international gene banks (SMc & RSH).

References

Banaticla-Hilario M. C. N., McNally K. L., Berg R. G. van den, Sackville Hamilton N. R., 2013 Crossability patterns within and among *Oryza* series Sativae species from Asia and Australia. Genet. Resour. Crop Evol. **60**: 1899–1914.

Barbier P., 1989 Genetic variation and ecotypic diferentiation in the wild rice species *Oryza* rufipogon. II. Influence of the mating system and life-history traits on the genetic structure of populations. Japanese J. Genet. **64**: 273–285.

Barbier P., Morishima H., Ishihama A., 1991 Phylogenetic relationships of annual and perennial wild rice: probing by direct DNA sequencing. Theor. Appl. Genet. **81**: 693–702.

Cai H.-W., Akimoto M., Morishima H., 2008 Genetic diversity in wild relatives of rice and domestication events. In: *Rice Biology in the Genomics Era*,, pp. 261–275.

Chang T., 1976 The origin, evolution, cultivation, dissemination, and diversification of Asian and African rices. Euphytica **25**: 425–441.

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. Mol. Biol. Evol. **20**: 67–75.

Civáň P., Craig H., Cox C. J., Brown T. A., 2015 Three geographically separate domestications of Asian rice. Nat. Plants **1**: 15164.

Clement M., Posada D., Crandall K. A., 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**: 1657–1659.

Eizenga G. C., Ali M. L., Bryant R. J., Yeater K. M., McClung A. M., McCouch S. R., 2014 Registration of the rice diversity panel 1 for genomewide association studies. J. Plant … **8**: 109.

Elert E., 2014 Rice by the numbers: A good grain. **514**: S50–S51.

Elshire R. J., Glaubitz J. C., Sun Q., Poland J. a, Kawamoto K., Buckler E. S., Mitchell S. E., 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One **6**: e19379.

Fuller D. Q., Sato Y.-I., Castillo C., Qin L., Weisskopf A. R., Kingwell-Banham E. J., Song J., Ahn S.-M., Etten J., 2010 Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol. Anthropol. Sci. **2**: 115–131.

Fuller D. Q., 2011 Finding Plant Domestication in the Indian Subcontinent. Curr. Anthropol. **52**: S347–S362.

Garris A. J., Tai T. H., Coburn J., Kresovich S., McCouch S., 2005 Genetic structure and diversity in *Oryza sativa* L. Genetics **169**: 1631–8.

Gepts P., 2006 Plant genetic resources conservation and utilization. Crop Sci. **46**: 2278.

Glaubitz J. C., Casstevens T. M., Lu F., Harriman J., Elshire R. J., Sun Q., Buckler E. S., 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. (NA Tinker, Ed.). PLoS One **9**: e90346.

Gross B. L., Zhao Z., 2014 Archaeological and genetic insights into the origins of domesticated rice. Proc. Natl. Acad. Sci. U. S. A. **111**: 6190–7.

Huang X., Zhao Y., Wei X., Li C., Wang A., Zhao Q., Li W., Guo Y., Deng L., Zhu C., Fan D., Lu Y., Weng Q., Liu K., Zhou T., Jing Y., Si L., Dong G., Huang T., Lu T., Feng Q., Qian Q., Li J., Han B., 2012a Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. **44**: 32–9.

Huang P., Molina J., Flowers J. M., Rubinstein S., Jackson S. A., Purugganan M. D., Schaal B. A., 2012b Phylogeography of Asian wild rice, *Oryza rufipogon*: a genome-wide view. Mol. Ecol. **21**: 4593–604.

Huang X., Kurata N., Wei X., Wang Z.-X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L., Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A., Toyoda A., Lu T., Feng Q., Qian Q., Li J., Han B., 2012c A map of rice genome variation reveals the origin of cultivated rice. Nature.

Huson D. H., Bryant D., 2006 Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23**: 254–67.

Jensen J. L., Bohonak A. J., Kelley S. T., 2005 Isolation by distance, web service. BMC Genet. **6**: 13.

Kanno A., Watanabe N., Nakamura I., Hirai A., 1993 Variations in chloroplast DNA from rice (*Oryza* sativa): differences between deletions mediated by short direct-repeat sequences within a single species. Theor. Appl. Genet. **86**: 579–584.

Khush G. S., 1997 Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. **35**: 25–34.

Kim H., Jeong E. G., Ahn S.-N., Doyle J., Singh N., Greenberg A. J., Won Y. J., McCouch S. R., 2014 Nuclear and chloroplast diversity and phenotypic distribution of rice (*Oryza* sativa L.) germplasm from the democratic people's republic of Korea (DPRK; North Korea). Rice (N.

Y). **7**: 7.

Kovach M. J., Calingacion M. N., Fitzgerald M. a, McCouch S. R., 2009 The origin and evolution of fragrance in rice (*Oryza* sativa L.). Proc. … **106**: 14444–9.

Langmead B., Salzberg S. L., 2012 Fast gapped-read alignment with Bowtie 2. Nat. Methods **9**: 357–9.

Leigh J. W., Bryant D., 2015 popart : full-feature software for haplotype network construction (S Nakagawa, Ed.). Methods Ecol. Evol. **6**: 1110–1116.

Li W.-H., 1997 *Molecular Evolution*. Sinauer Associates.

Li C., Zhou A., Sang T., 2006 Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza* nivara. New Phytol. **170**: 185–93.

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*. Proc. Natl. Acad. Sci. U. S. A. **103**: 9578–9583.

McCouch S. S. R., McNally K. K. L., Wang W., Sackville Hamilton R., 2012 Genomics of gene banks: A case study in rice. Am. J. Bot. **99**: 407–23.

McCouch S., Baute G. J., Bradeen J., Bramel P., Bretting P. K., Buckler E., Burke J. M., Charest D., Cloutier S., Cole G., Dempewolf H., Dingkuhn M., Feuillet C., Gepts P., Grattapaglia D., Guarino L., Jackson S., Knapp S., Langridge P., Lawton-Rauh A., Lijua Q., Lusty C., Michael T., Myles S., Naito K., Nelson R. L., Pontarollo R., Richards C. M., Rieseberg L., Ross-Ibarra J., Rounsley S., Hamilton R. S., Schurr U., Stein N., Tomooka N., Knaap E. van der, Tassel D. van, Toll J., Valls J., Varshney R. K., Ward J., Waugh R., Wenzl P., Zamir D., 2013 Agriculture: Feeding the future. Nature **499**: 23–24.

Meilleur B. A. B., Hodgkin T., 2004 In situ conservation of crop wild relatives: status and trends. Biodivers. Conserv. **13**: 663–668.

Molina J., Sikora M., Garud N., Flowers J. M., Rubinstein S., Reynolds A., Huang P., Jackson S., Schaal B. a, Bustamante C. D., Boyko A. R., Purugganan M. D., Molinaa J., Sikorab M., Garudb N., 2011 Molecular evidence for a single evolutionary origin of domesticated rice. Proc. Natl. Acad. Sci. U. S. A. **108**: 8351–8356.

Morishima H., OKA H.-I., Chang W.-T., 1961 Directions of Differentiation in Populations of Wild Rice , *Oryza* perennis and O . sativa f . spontanea. Evolution (N. Y). **15**: 326–339.

Nei M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U. S. A. **70**: 3321–3323.

Oka H.-I., 1988 *Origin of cultivated rice.* Elsevier.

Plucknett D., Smith N., Williams J., Anishetty N., 1983 Crop germplasm conservation and developing countries. Science (80-.).

Pourkheirandish M., Hensel G., Kilian B., Senthil N., Chen G., Sameri M., Azhaguvel P., Sakuma S., Dhanagond S., Sharma R., Mascher M., Himmelbach A., Gottwald S., Nair S. K., Tagiri A., Yukuhiro F., Nagamura Y., Kanamori H., Matsumoto T., Willcox G., Middleton C. P., Wicker T., Walther A., Waugh R., Fincher G. B., Stein N., Kumlehn J., Sato K., Komatsuda T., 2015 Evolution of the Grain Dispersal System in Barley. Cell **162**: 527–539.

Raj A., Stephens M., Pritchard J. K., 2014 fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics **197**: 573–89.

Ray D. K., Mueller N. D., West P. C., Foley J. A., 2013 Yield Trends Are Insufficient to Double Global Crop Production by 2050. (JP Hart, Ed.). PLoS One **8**: e66428.

Schatz M. C., Maron L. G., Stein J. C., Hernandez Wences A., Gurtowski J., Biggers E., Lee H., Kramer M., Antoniou E., Ghiban E., Wright M. H., Chia J., Ware D., McCouch S. R., McCombie W. R., 2014 Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica*. Genome Biol. **15**: 506.

Song Z., Li B., Chen J., Lu B.-R., 2005 Genetic diversity and conservation of common wild rice (*Oryza rufipogon*) in China. Plant Species Biol. **20**: 83–92.

Sotowa M., Ootsuka K., Kobayashi Y., Hao Y., Tanaka K., Ichitani K., Flowers J. M., Purugganan M. D., Nakamura I., Sato Y.-I., Sato T., Crayn D., Simon B., Waters D. LE, Henry R. J., Ishikawa R., 2013 Molecular relationships between Australian annual wild rice, *Oryza* meridionalis, and two related perennial forms. Rice (N. Y). **6**: 26.

Sweeney M. T., Thomson M. J., Cho Y. G., Park Y. J., Williamson S. H., Bustamante C. D., McCouch S. R., 2007 Global dissemination of a single mutation conferring white pericarp in rice. (J Pritchard, Ed.). PLoS Genet. **3**: 1418–1423.

Takano-Kai N., Jiang H., Kubo T., Sweeney M., Matsumoto T., Kanamori H., Padhukasahasram B., Bustamante C. D., Yoshimura A., Doi K., McCouch S., 2009 Evolutionary History of *GS3*, a Gene Conferring Grain Length in Rice. Genetics **182**: 1323–1334.

Tanksley S. D., McCouch S. S. R., 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. Science (80-.). **277**: 1063–1066.

Vaughan L. K., Ottis B. B., 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 Sci. **49**: 468–476.

Vaughan D. A., Morishima H., Kadowaki K., 2003 Diversity in the *Oryza* genus. Curr. Opin. Plant Biol. **6**: 139–146.

Vaughan D. a., Lu B.-R., Tomooka N., 2008 The evolving story of rice evolution. Plant Sci. **174**: 394–408.

Wang Z. Y., Second G., Tanksley S. D., 1992 Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. Theor. Appl. Genet. **83**: 565–581.

Wang M. X., Zhang H. L., Zhang D. L., Qi Y. W., Fan Z. L., Li D. Y., Pan D. J., Cao Y. S., Qiu Z. E., Yu P., Yang Q. W., Wang X. K., Li Z. C., 2008 Genetic structure of *Oryza rufipogon* Griff. in China. Heredity (Edinb). **101**: 527–35.

Waters D. D. L. E., Nock C. J. C., Ishikawa R., Rice N., Henry R. J., 2012 Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. Ecol. Evol. **2**: 211–7.

Weir B., Cockerham C. C., 1984 Estimating F-Statistics for the Analysis of Population Structure on JSTOR. Evolution (N. Y). **38**: 1358–1370.

Yang C. -c., Kawahara Y., Mizuno H., Wu J., Matsumoto T., Itoh T., 2011 Independent Domestication of Asian Rice Followed by Gene Flow from *japonica* to *indica*. Mol. Biol. Evol. **29**: 1471–1479.

Zhu B.-F., Si L., Wang Z., Zhou Y., Zhu J., Shangguan Y., Lu D., Fan D., Li C., Lin H., Qian Q., Sang T., Zhou B., Minobe Y., Han B., 2011 Genetic control of a transition from black to straw-white seed hull in rice domestication. Plant Physiol. **155**: 1301–11.

Figures

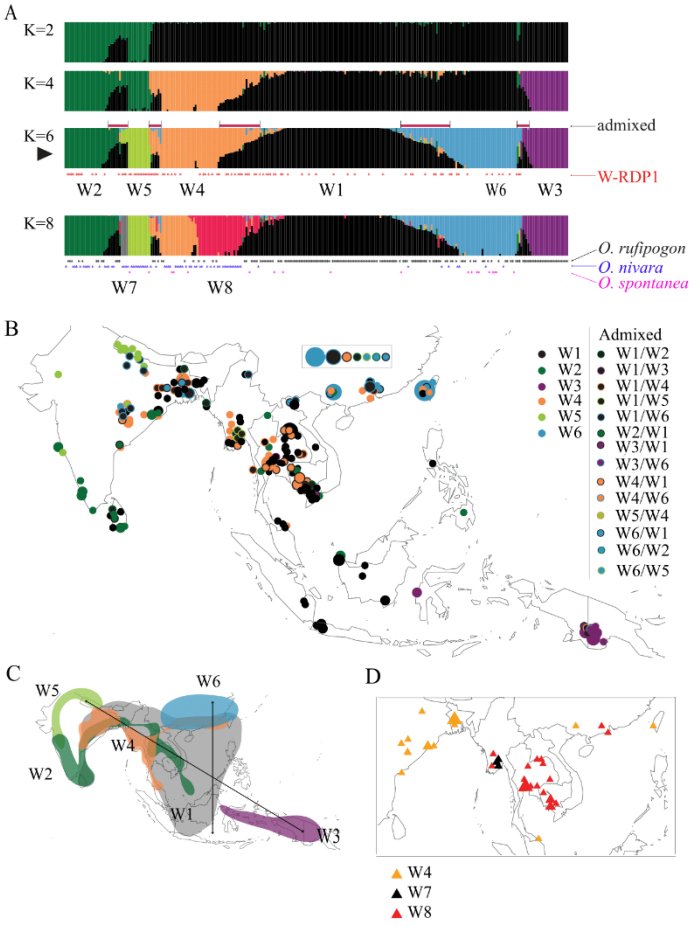


Figure 1. Population structure in the *ORSC*. (A) *fastStructure* analysis for 286 *ORSC* samples based on 113,739 SNPs where black arrow indicates optimal number of populations at K=6 (see Fig. S1A); admixed accessions sharing <75% ancestry with any one subpopulation are highlighted by red rectangles above K=6 panel; wild group numbers, W1-W6, correspond to order of divergence (as shown in Fig. S1C); accessions included in the Wild Rice Diversity Panel (Wild RDP, n=95) indicated as red stars under K=6 panel; traditional species designations, *O. rufipogon* (perennial), *O. nivara* (annual), and *O. spontanea* indicated by black, blue and pink stars, respectively, under K=8 panel. (B) Geographical map showing distribution of samples from each subpopulation group where circle size corresponds to number of samples; fill color indicates subpopulation designation (K=6); For admixed accessions, the first mentioned subpopulation represents the major proportion of ancestry; Chinese accessions lacking location detail indicated in closed rectangle; further detail provided in Table S1A. (C) Simplified geographical map showing regional distribution of six subpopulation groups (K=6); (D) Detail view of geographical distribution of subpopulation groups (K=8) highlighting relationship between W4 and W8.

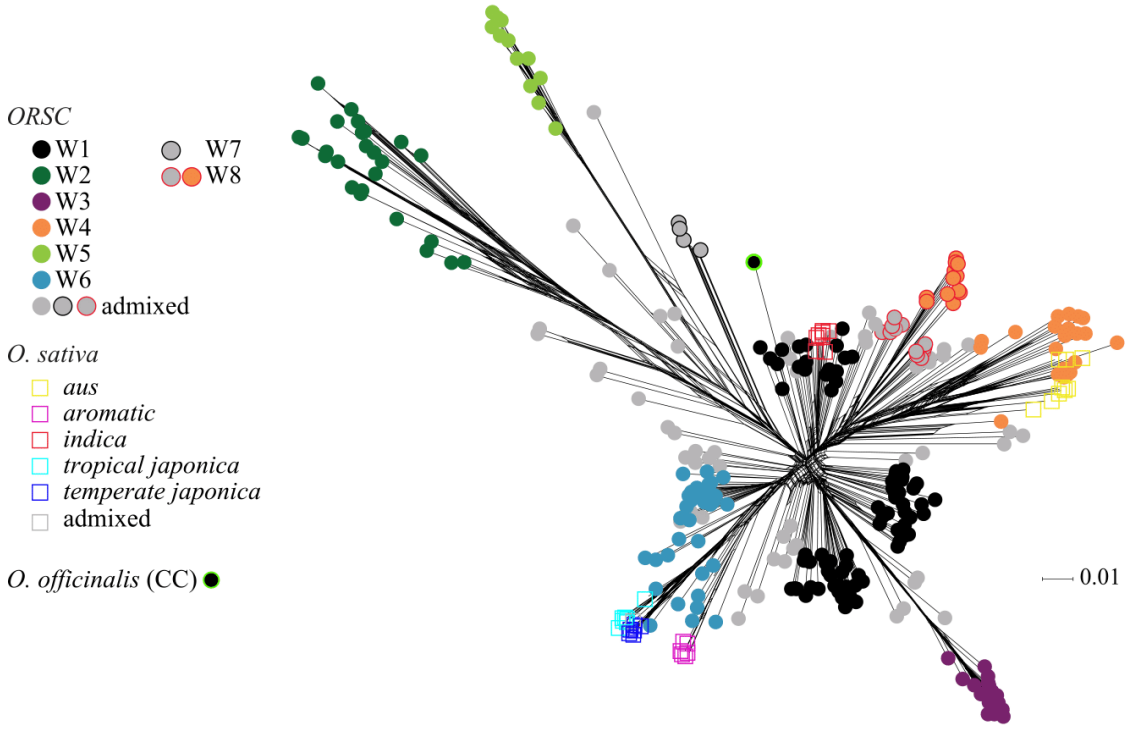


Figure 2.
Phylogenetic network based on SNP data from the *ORSC* and *O. sativa* samples. Circle color corresponds to subpopulation identity as in Fig. 1A.

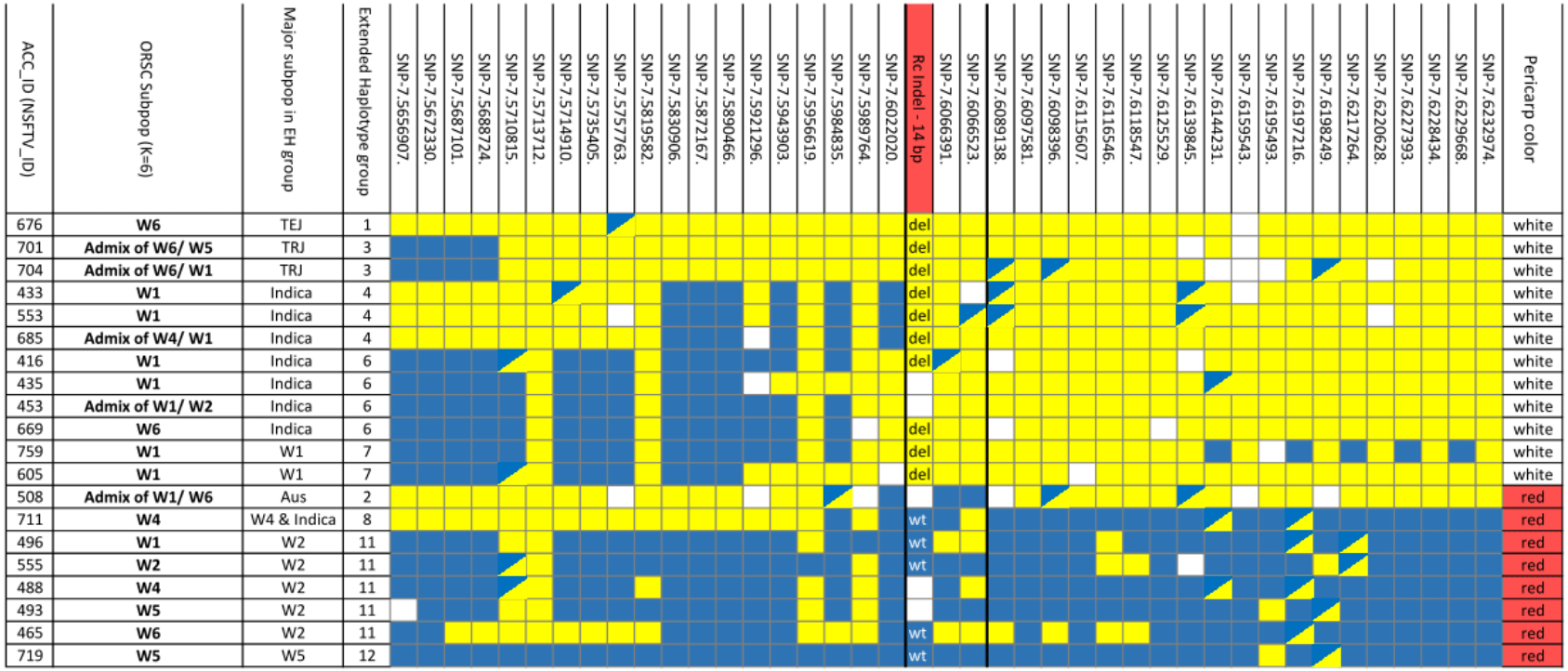


Figure 3. Rc extended haplotypes for representative *ORSC* accessions. Extended haplotypes across a 576-KB window around the Rc gene for 12 white pericarp and 8 red pericarp *ORSC* accessions. The two SNPs and 14-bp indel within the Rc gene are outlined in black. Yellow = cultivated allele; blue = wild type allele; blue/yellow =heterozygous; white = missing data. Note that all but one (NSFTV_ID 508) of the white pericarp accessions carries the cultivated allele at all three markers within the Rc gene. NSF ID corresponds to accession number in Table 1.

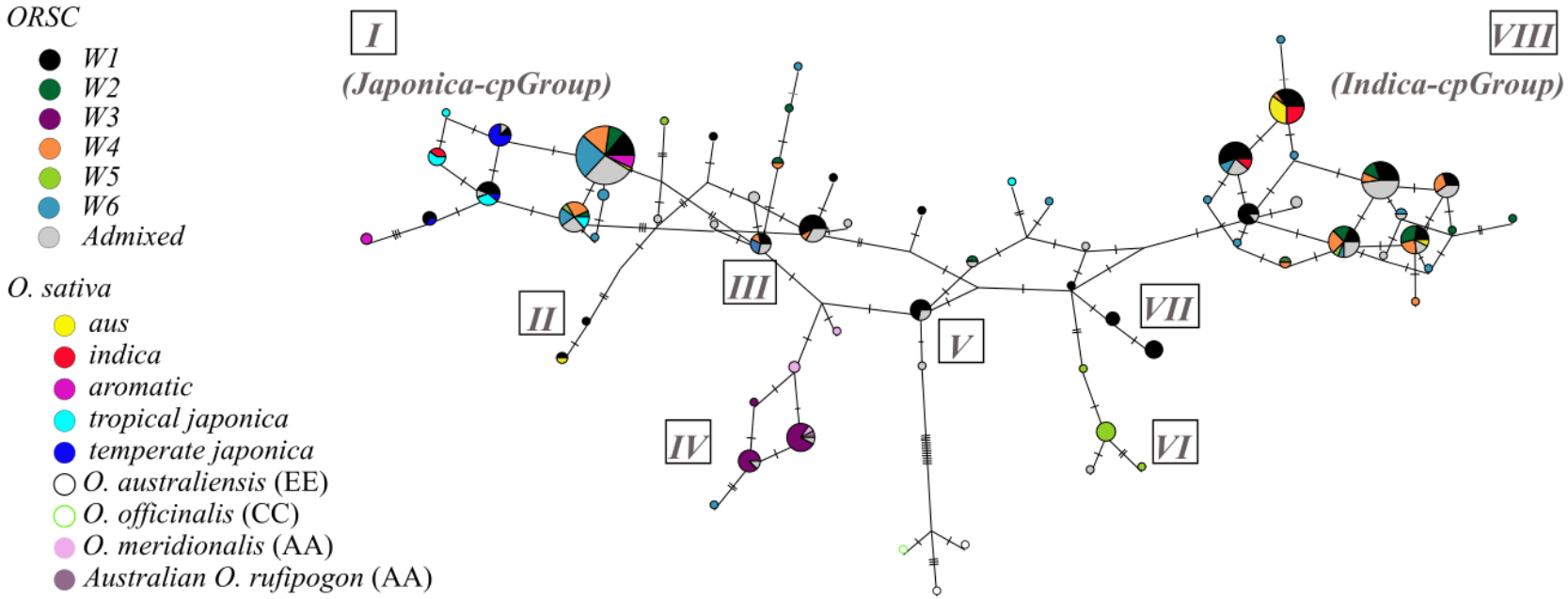


Figure 4. Chloroplast haplotype network. (A) Haplotype network for the *ORSC* and *O. sativa* samples based on 25 chloroplast variants**;** single mutations indicated as hatches between haplotypes; chloroplast groups (*cpGroup*) *I* to *VIII* indicated in rectangles*;* size of nodes (circles) is proportional to haplotype frequency; colors indicate proportion of individuals from each subpopulation (based on GBS data at K=6 in Fig. 1A) that carry the haplotype; gray indicates admixed accessions; for more detail, see Fig. S3 and Table S1. (B) Pie chart showing the frequency of nuclear subpopulations represented by individuals in each chloroplast haplotype group.

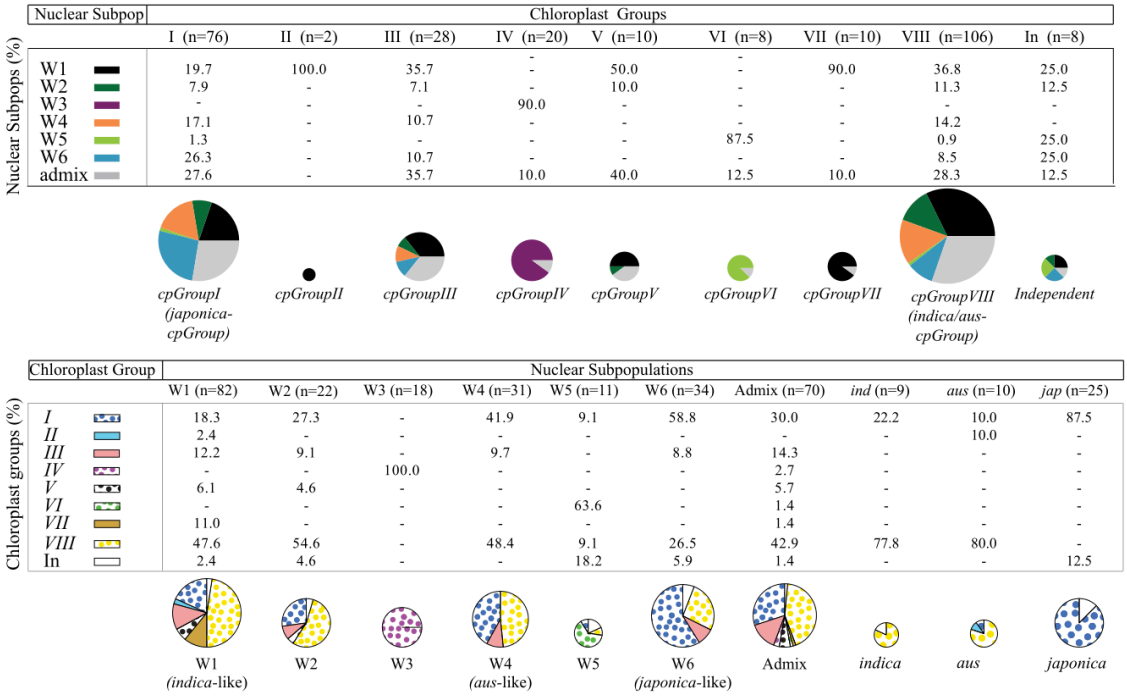


Figure 5. Chloroplast haplotype network. Incongruence between chloroplast group and nuclear subpopulation; top pie chart shows the frequency of nuclear subpopulations represented by individuals in each chloroplast haplotype group; bottom pie chart shows the frequency of chloroplast groups represented by individuals in each nuclear subgroup.

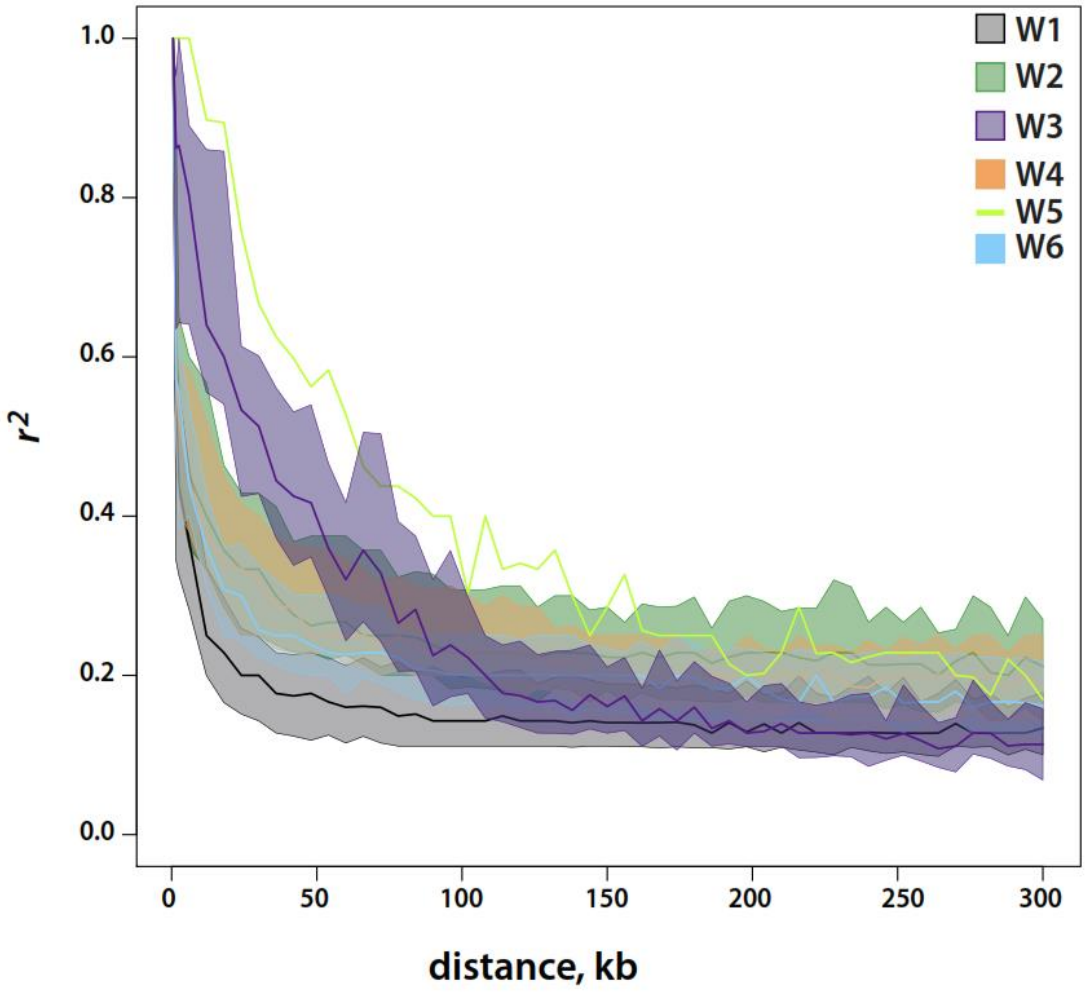


Figure 6. LD decay for each subpopulation.

Supplementary Figures

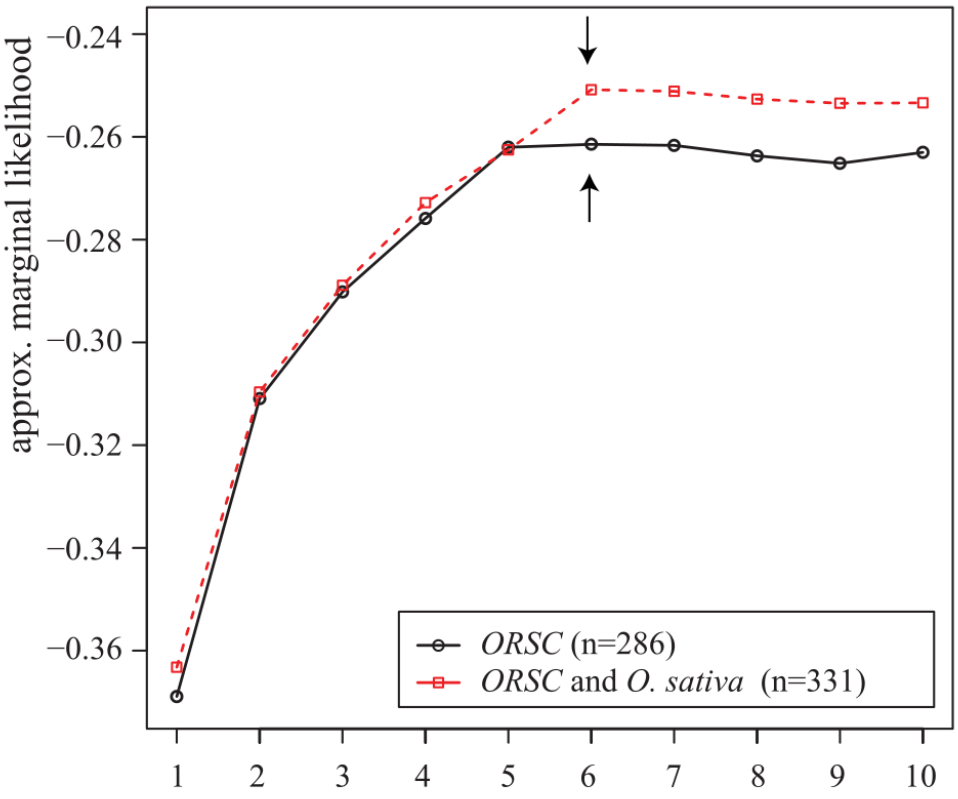


Figure S1. Population structure in the *ORSC* and with *O. sativa*. Analysis of model complexity (K); K value evaluated based on approximations of the marginal likelihood of the data computed by *fastStructure* among the *ORSC* and the *ORSC* with *O. sativa*; black arrows indicate the highest value of marginal likelihood.

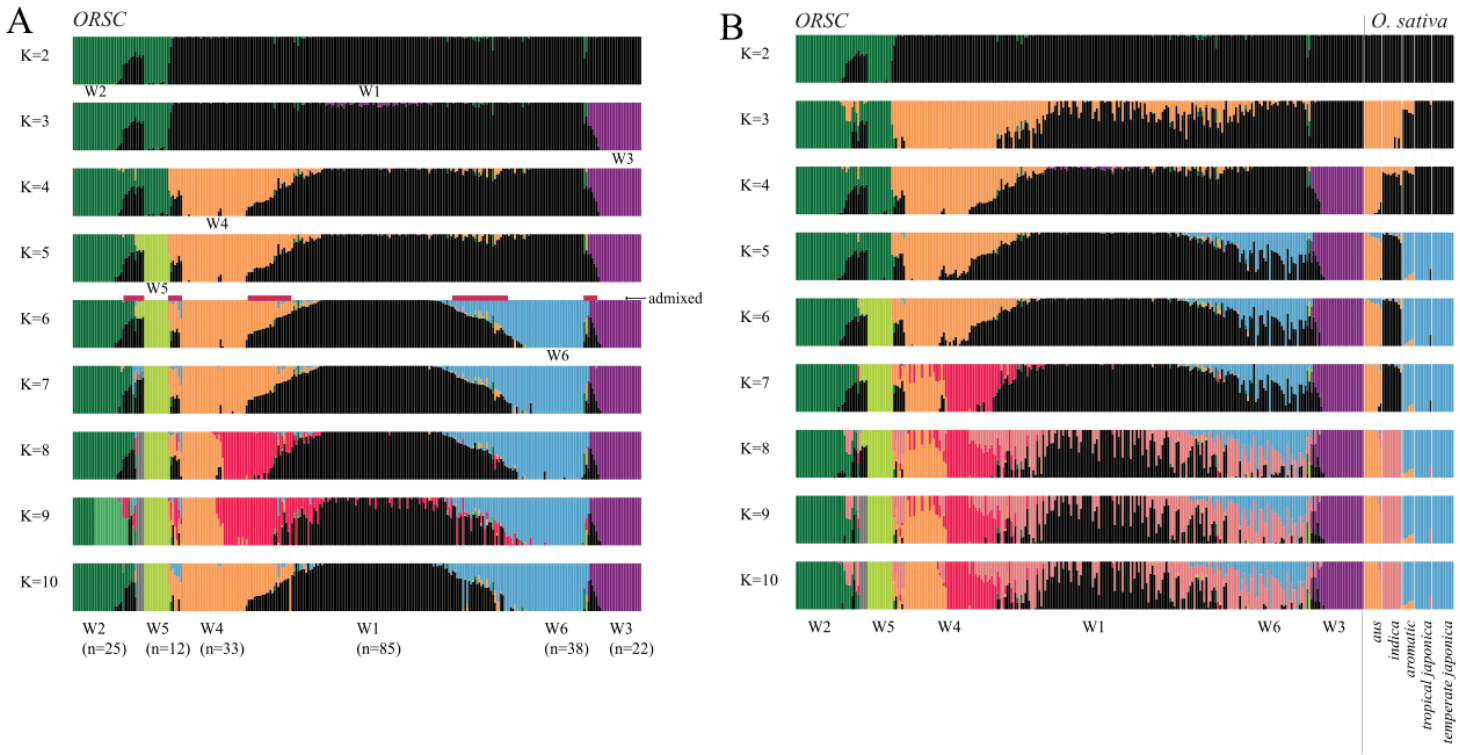


Figure S2. Population structure in the *ORSC* and with *O. sativa*. (A) *fastStructure* analysis for 286 *ORSC* samples from K=2 to K=10 based on 113,996 SNPs; subpopulation identity defined at K=6 (see Fig. 1A); wild group numbers, W1-W6, correspond to order of divergence; admixed accessions indicated by red rectangles above K=6 panel. (B) *fastStructure* analysis for 286 *ORSC* samples and 45 *O. sativa* accessions from K=2 to K=10 based on same 113,996 SNPs; the three *O. sativa* subpopulations correspond to *aus* (orange), *indica* (pink) and *japonica* (blue).

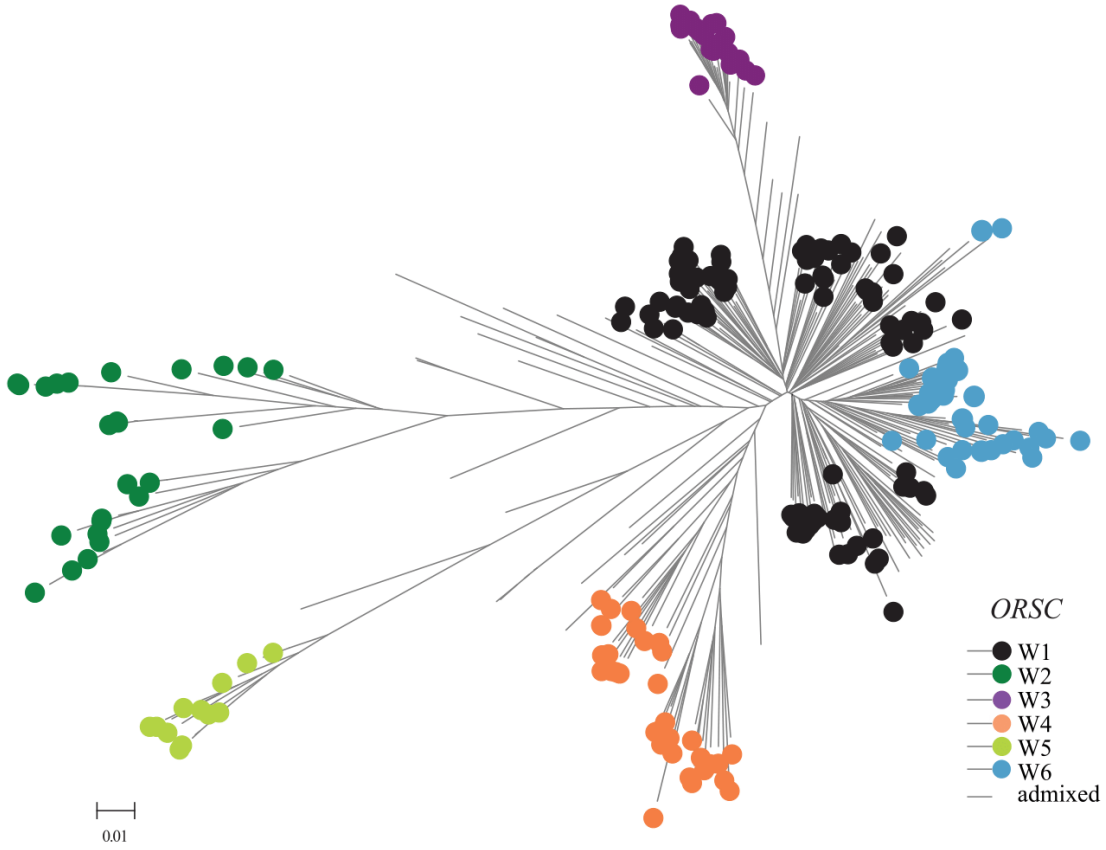


Figure S3. Neighbor Joining (NJ) tree from the *ORSC* based on SNP data. Circle color corresponds to subpopulation identity as in Fig. 1A.

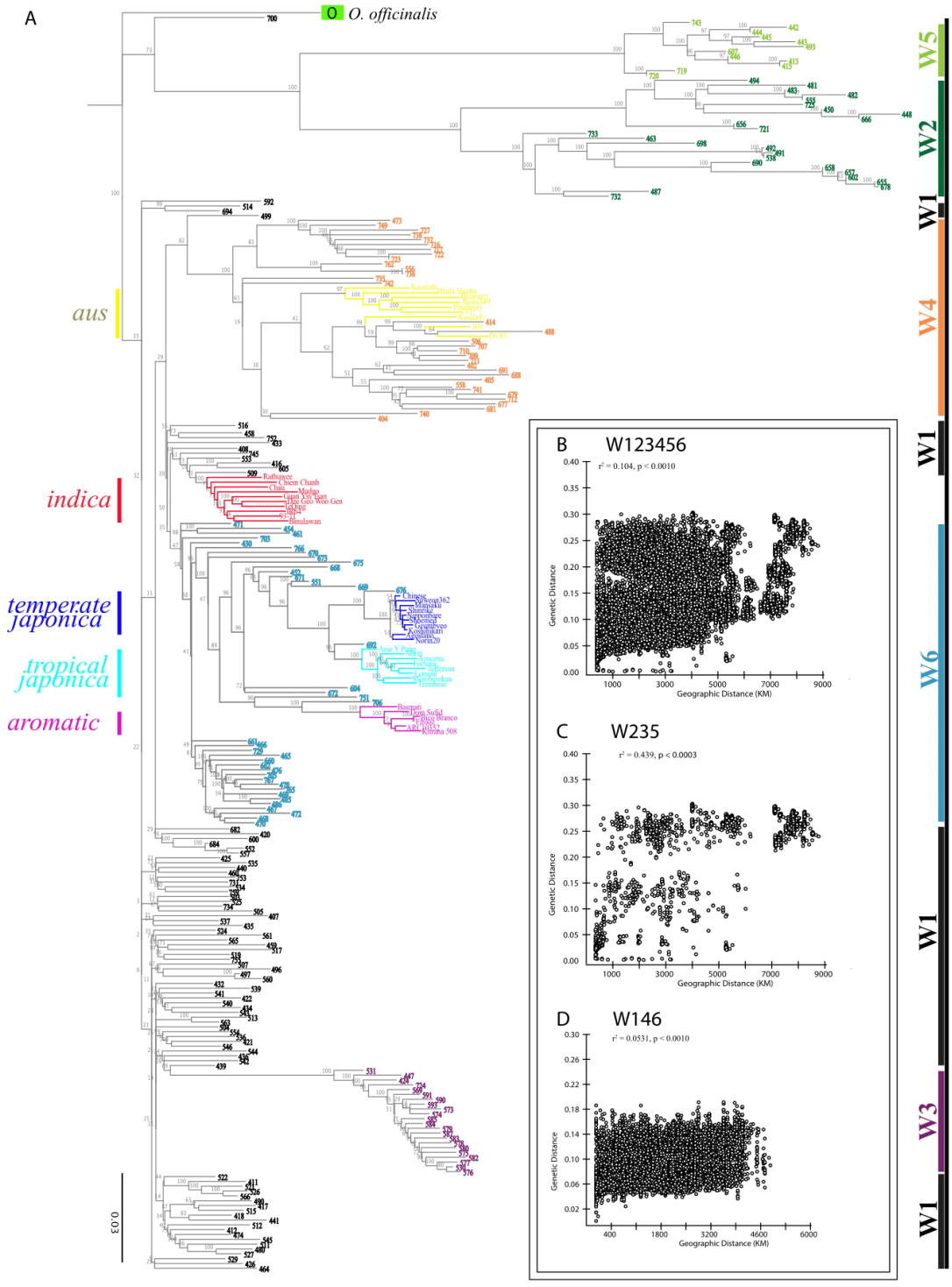


Figure S4. The relationship between geographical and genetic distance of the *ORSC*. Consensus Neighbor Joining (NJ) dendrogram of 262 individuals (215 wild accessions of groups W1-W6, 45 *O. sativa* accessions, and one *O. officinalis* outgroup) based on 100 bootstrap replications. Colors indicate subpopulation identity; colored bars on left correspond to *O. sativa* subpopulation groups; colored bars on right correspond to *ORSC* subpopulations. Right panel graphs show results of the Mantel test between genetic distance and geographical distance for individuals in: all six *ORSC* subpopulations (S4B); the three independent *ORSC* subpopulations (S4C); and the three *O. sativa*-like *ORSC* subpopulations (S4D).

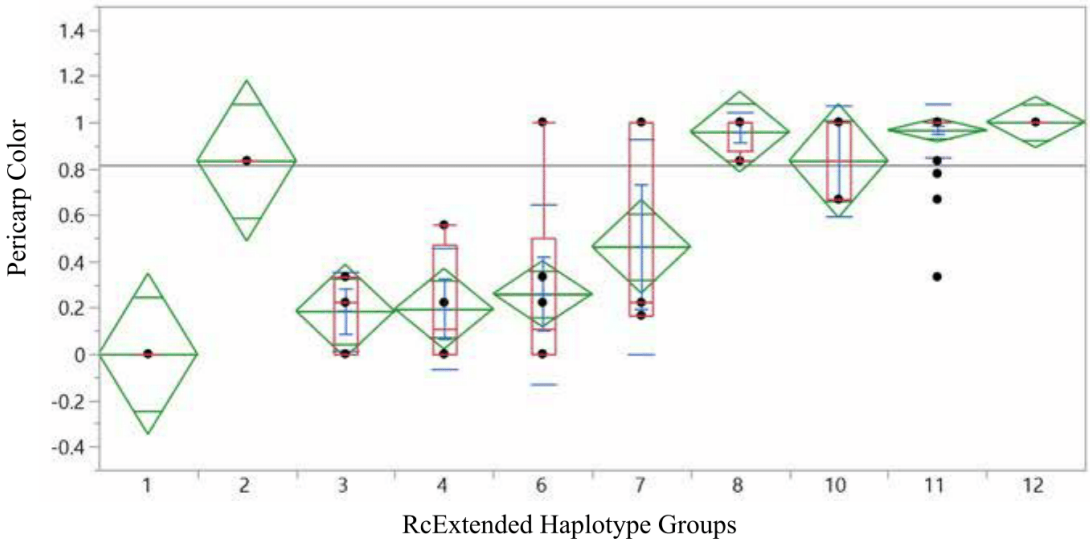


Figure S5. Pericarp Color Associated with *Rc* Extended Haplotype Groups. Quantile boxplots showing the pericarp color of 81 accessions across 12 Rc extended haplotype groups. Points of green diamonds indicate upper and lower 95% confidence intervals. Mean error and standard deviations are indicated by the inner and outer pairs of blue horizontal lines, respectively. Pericarp color was scored according to the following scale: red pericarp=1.0; white pericarp=0.

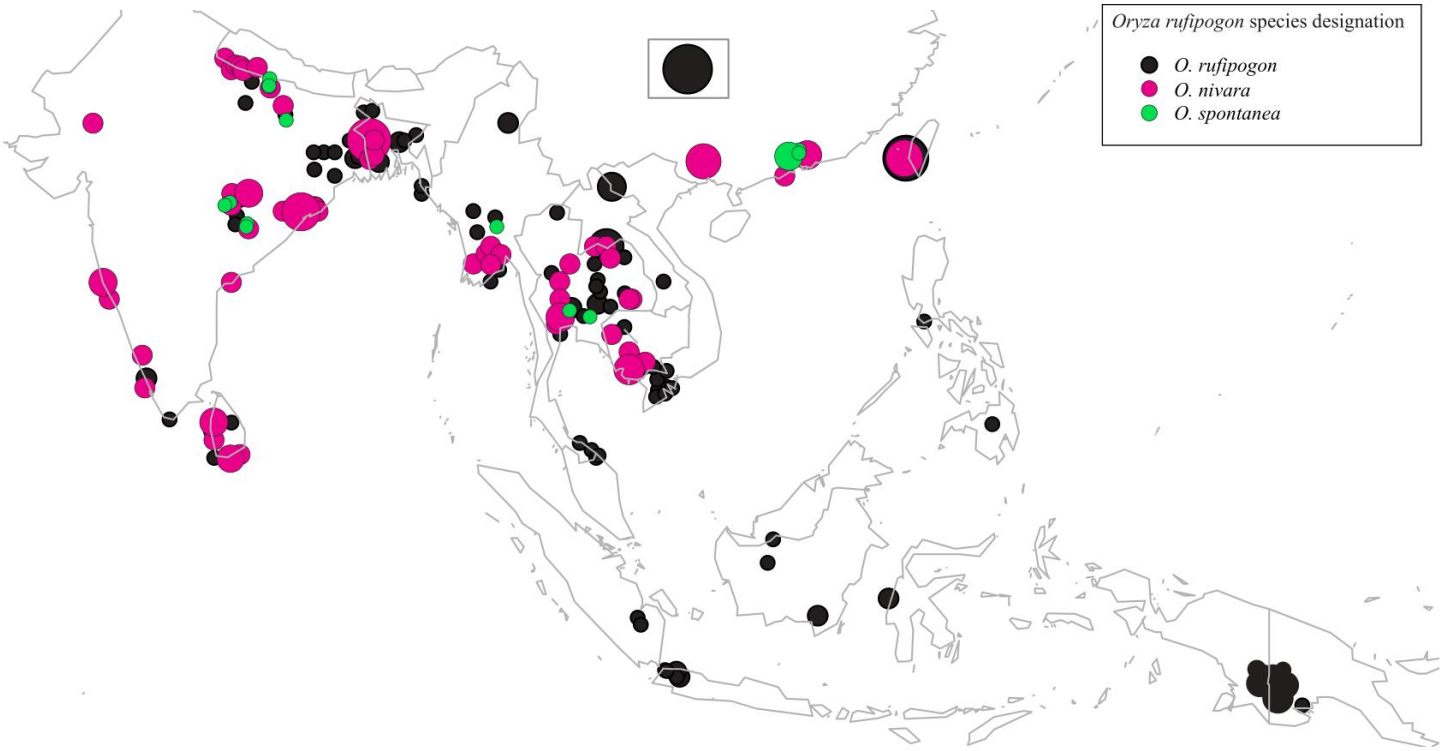


Figure S6: Traditional species nomenclature in the *ORSC*. Geographical map showing distribution of samples from each species designation where size of circle corresponds to relative number of samples; Chinese accessions lacking location detail indicated in closed rectangle.

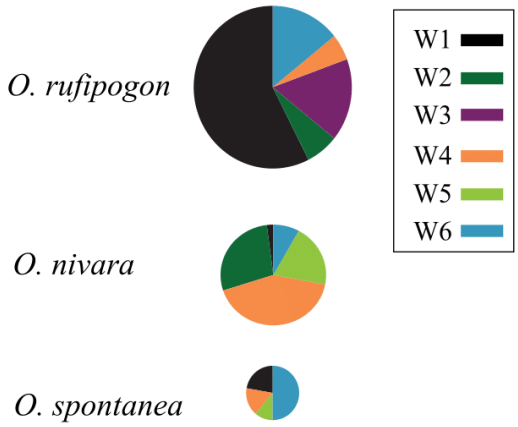


Figure S7: Distribution of nuclear subpopulations within traditional species groups in the *ORSC*; *rufipogon* (perennial), *O. nivara* (annual), and *O. spontanea*.

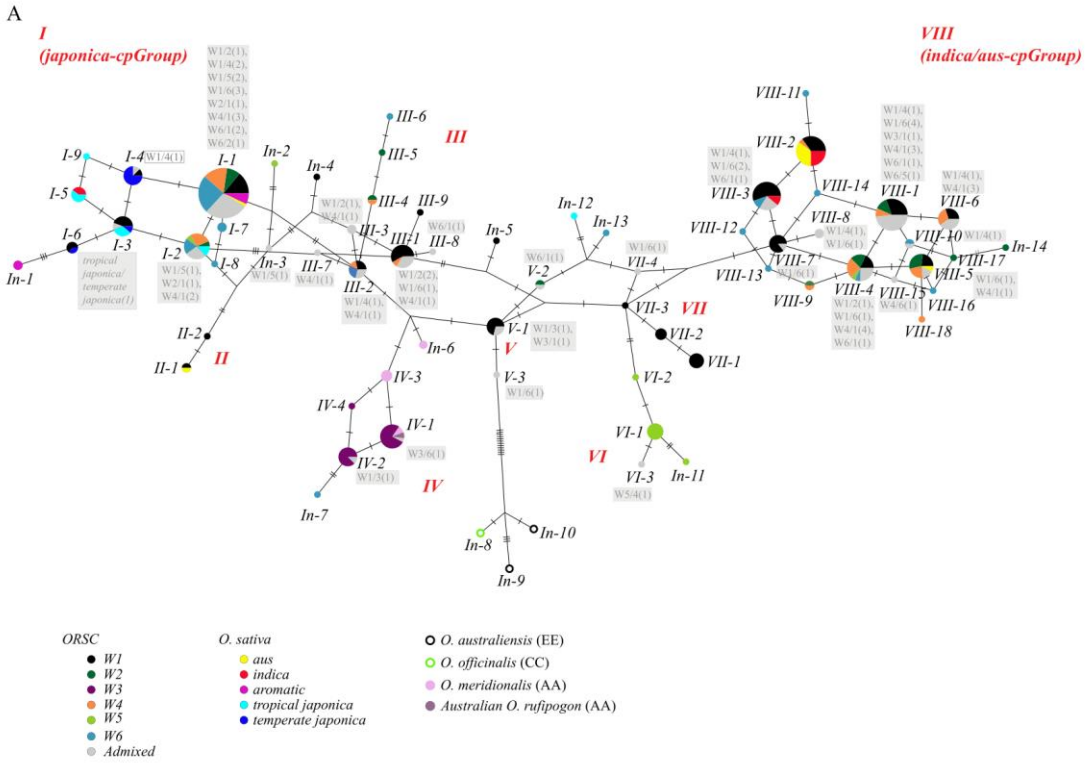


Figure S8. Chloroplast haplotype network. (A) Haplotype network for the *ORSC* and *O. sativa* samples based on 25 chloroplast variants; single mutations indicated as hatches between haplotypes; chloroplast groups (*cpGroup*) *I* to *VIII* indicated in rectangles; size of nodes (circles) is proportional to haplotype frequency; colors indicate proportion of individuals from each subpopulation (based on GBS data at K=6 in Fig. 1a) that carry the haplotype; text highlighted in gray indicates admixed accessions with genetic information.

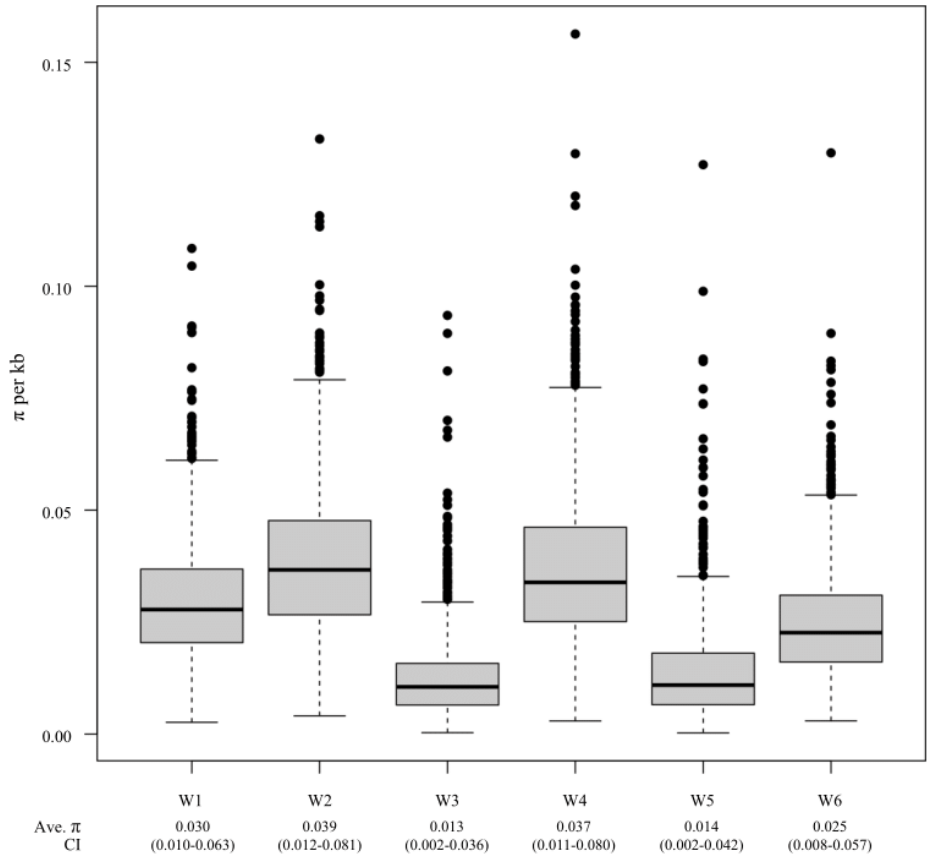


Figure S9. Average DNA sequence diversity (π) within each *ORSC* subpopulation. Average π estimated based on pairwise sequence diversity; CI = 95% confidence interval. Note: These values represent underestimates by roughly one to two orders of magnitude compared to full sequencing data because only a subset of the DNA between any pair of SNPs was actually sequenced in our GBS data set. While they are not expected to be consistent with data sets that are based on complete sequences, our estimates are useful for among-poulation comparisons within our data set.

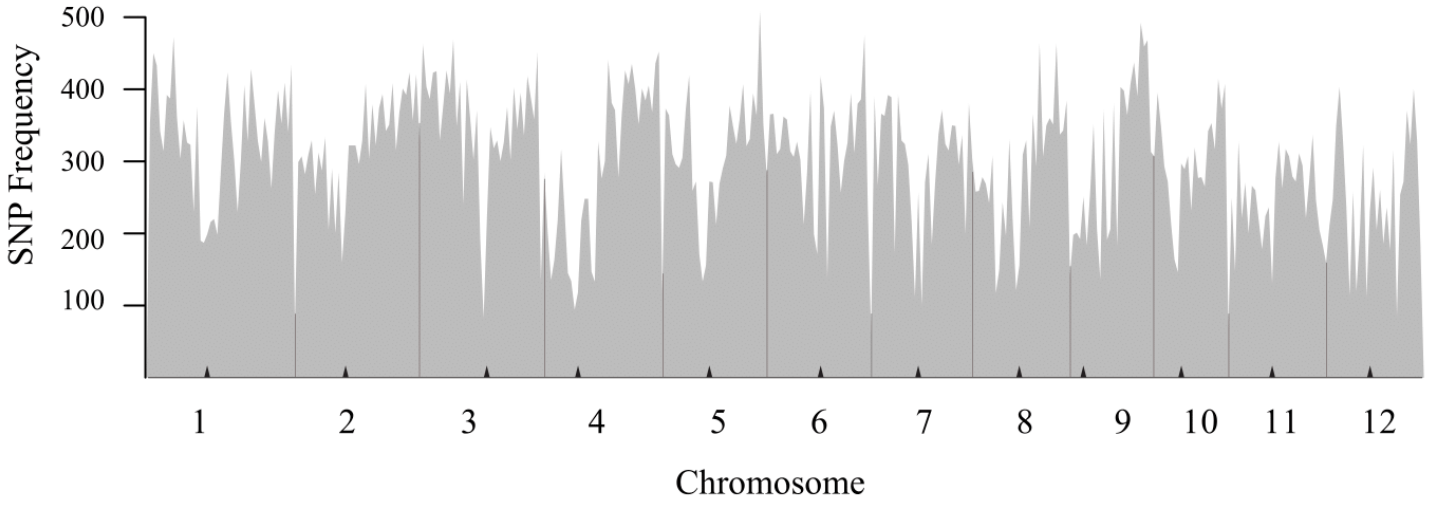


Figure S10. Distribution of GBS SNPs along the twelve chromosomes of rice. SNP frequency calculated in 1Mb window; black triangles indicate centromere positions.

Supplementary Tables

Table S1. Germplasm information.

A. *ORSC* accessions

NSF ID[1]	IRGC ID[2] /Wild ID[3]	Species	Country	Wild-RDP[4]	Subpopulation K=6	Subpopulation K=8	Ancestry at K=6 W1	W2	W3	W4	W5	W6	Chloroplast Haplotype	Previous genetic information Cheng's et al. (2003)[5]	Previous genetic information Huang's et al. (2012)[6]	Pericarp color scores	*Rc* 14bp deletion
499	105767	*O. rufipogon*	Thailand	+	*W1*		0.77	0.00	0.00	0.23	0.00	0.00	*VIII-3*	-	-	1.000	-
700	103423	*O. rufipogon*	Sri Lanka	-	*W1*		0.79	0.10	0.00	0.06	0.00	0.05	*VIII-6*	-	-	-	-
553	100926	*O. rufipogon*	Myanmar	+	*W1*		0.79	0.00	0.00	0.12	0.00	0.08	*II-2*	-	Or-II	0.000	14bp del
557	105567	*O. rufipogon*	Indonesia	-	*W1*		0.84	0.16	0.00	0.00	0.00	0.00	*II-1*	-	Or-II	1.000	-
433	83795	*O. rufipogon*	India	+	*W1*		0.84	0.00	0.00	0.10	0.00	0.06	*VII-1*	-	-	0.000	14bp del
592	80671	*O. rufipogon*	India	+	*W1*		0.86	0.02	0.00	0.04	0.00	0.08	*VIII-5*	-	-	1.000	-
496	105720	*O. rufipogon*	Cambodia	+	*W1*		0.86	0.02	0.00	0.12	0.00	0.00	*VIII-7*	-	-	1.000	WT
752	105901	*O. rufipogon*	Bangladesh	-	*W1*		0.86	0.00	0.00	0.12	0.00	0.02	*I-1*	-	-	-	-
605	100911	*O. spontanea*	Thailand	+	*W1*		0.86	0.00	0.00	0.11	0.00	0.03	*I-3*	-	Or-I	0.167	14bp del
600	100187	*O. sat. x O. ruf.*	Malaysia	+	*W1*		0.90	0.00	0.00	0.01	0.00	0.09	*I-4*	-	Or-II	1.000	-
745	105738	*O. rufipogon*	Cambodia	-	*W1*		0.92	0.00	0.00	0.05	0.00	0.03	*VIII-1*	-	-	-	-
507	105881	*O. rufipogon*	Bangladesh	-	*W1*		0.92	0.00	0.00	0.08	0.00	0.00	*III-1*	-	-	-	-
420	81984	*O. rufipogon*	Laos	+	*W1*		0.94	0.00	0.00	0.06	0.00	0.00	*VIII-6*	-	-	1.000	-
416	81970	*O. spontanea*	Thailand	+	*W1*		0.97	0.00	0.00	0.02	0.00	0.00	*I-1*	-	-	0.000	14bp del
497	105726	*O. rufipogon*	Cambodia	-	*W1*		0.99	0.00	0.00	0.01	0.00	0.00	*VIII-7*	-	-	1.000	-
753	105960	*O. rufipogon*	Bangladesh	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	-	-
407	80742	*O. rufipogon*	Myanmar	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-6*	-	-	1.000	-
417	81976	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VII-2*	-	-	1.000	WT
539	106412	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*V-1*	-	-	-	-
563	105951	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	1.000	-
512	105942	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-7*	-	-	-	-
411	81801	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-3*	-	-	-	-
425	82040	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VII-2*	-	-	-	-
441	92605	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VII-3*	-	-	1.000	-
490	105567	*O. rufipogon*	Indonesia	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-2*	-	Or-II	1.000	-
511	105909	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	-	-
543	106420	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-1*	-	-	-	-
422	81993	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*V-1*	-	-	-	-
505	105855	*O. rufipogon*	Thailand	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-9*	-	-	1.000	-
545	106453	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
421	81990	*O. rufipogon*	Myanmar	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
426	82979	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-7*	-	-	-	-
435	86448	*O. rufipogon*	Thailand	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	0.000	-
464	104602	*O. rufipogon*	Sri Lanka	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
480	105250	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
513	105951	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	1.000	-
521	106145	*O. rufipogon*	Laos	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VII-1*	-	-	-	-
522	106150	*O. rufipogon*	Laos	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-1*	-	-	-	-
527	106166	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	-	-	-	1.000	-
535	106342	*O. rufipogon*	Myanmar	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	0.000	WT
537	106384	*O. spontanea*	Myanmar	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	-	-
541	106414	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*V-1*	-	-	-	-
759	106336	*O. rufipogon*	Cambodia	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-4*	-	-	0.222	14bp del
434	83823	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*V-1*	-	-	-	-
436	86454	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*In-4*	-	-	-	-
544	106452	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-
526	106163	*O. rufipogon*	Laos	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-2*	-	-	-	-
536	106357	*O. rufipogon*	Myanmar	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-2*	-	-	-	-
542	106415	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	-	-	-	1.000	-
560	105726	*O. rufipogon*	Cambodia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-7*	-	-	-	-
503	105843	*O. rufipogon*	Thailand	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VII-2*	-	-	1.000	-
515	105958	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
540	106413	*O. rufipogon*	Vietnam	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*V-1*	-	-	1.000	WT
566	106163	*O. rufipogon*	Laos	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-2*	-	-	-	-
682	100904	*O. rufipogon*	Thailand	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-1*	-	Or-II	1.000	-
546	106509	*O. rufipogon*	Myanmar	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*In-5*	-	-	-	-
554	103305	*O. rufipogon*	Philippines	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-1*	-	-	-	-
418	81977	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
534	106332	*O. rufipogon*	Cambodia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-7*	-	-	1.000	-
474	104714	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-3*	-	-	1.000	WT
525	106161	*O. rufipogon*	Laos	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-4*	-	-	-	-
412	81802	*O. rufipogon*	Indonesia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-3*	-	-	-	-
432	83794	*O. rufipogon*	Thailand	+	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-1*	-	-	-	-
439	86486	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
460	103848	*O. rufipogon*	India	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-5*	-	Or-III	1.000	WT
504	105847	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*III-2*	-	-	-	-
734	105487	*O. rufipogon*	Thailand	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-2*	-	Or-II	-	-
684	100920	*O. ruf. x O. niv.*	Malaysia	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-6*	-	Or-I	-	-
561	105868	*O. rufipogon*	Bangladesh	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-3*	-	-	-	-
459	103847	*O. rufipogon*	India	-	*W1*		1.00	0.00	0.00	0.00	0.00	0.00	*VIII-4*	-	Or-II	-	-

A. *ORSC* accessions cont'd

NSF ID[1]	IRGC ID[2] /Wild ID[3]	Species	Country	Wild-RDP[4]	Subpopulation K=6	Subpopulation K=8	Ancestry at K=6 W1	W2	W3	W4	W5	W6	Chloroplast Haplotype	Previous genetic information Cheng's et al. (2003)[5]	Previous genetic information Huang's et al. (2012)[6]	Pericarp color scores	*Rc* 14bp deletion
514	105956	*O. rufipogon*	Indonesia	+	*W1*		0.99	0.00	0.01	0.00	0.00	0.00	*VIII-1*	-	-	1.000	-
516	106036	*O. rufipogon*	Malaysia	-	*W1*		0.99	0.00	0.00	0.00	0.00	0.01	*III-2*	-	-	1.000	WT
731	105424	*O. rufipogon*	Sri Lanka	-	*W1*		0.99	0.00	0.00	0.00	0.00	0.01	*I-1*	-	-	-	-
440	88787	*O. rufipogon*	Bangladesh	-	*W1*		0.99	0.00	0.00	0.00	0.00	0.01	*I-1*	-	-	1.000	WT
524	106156	*O. rufipogon*	Laos	-	*W1*		0.98	0.00	0.00	0.00	0.00	0.02	*VIII-1*	-	-	1.000	-
565	106144	*O. rufipogon*	India	-	*W1*		0.97	0.00	0.00	0.00	0.00	0.03	*VIII-3*	-	-	1.000	-
755	106103	*O. rufipogon*	India	-	*W1*		0.93	0.00	0.00	0.00	0.00	0.07	*I-6*	-	-	-	-
519	106115	*O. rufipogon*	India	-	*W1*		0.91	0.00	0.00	0.00	0.00	0.09	*VII-1*	-	-	-	-
552	100920	*O. ruf. x O. niv.*	Malaysia	-	*W1*		0.91	0.09	0.00	0.00	0.00	0.00	*VIII-2*	-	Or-I	1.000	-
458	103844	*O. rufipogon*	Bangladesh	-	*W1*		0.89	0.00	0.00	0.00	0.00	0.11	*VII-1*	-	Or-I	0.500	-
408	80745	*O. spontanea*	Myanmar	-	*W1*		0.87	0.00	0.00	0.03	0.00	0.10	*VII-1*	-	-	1.000	-
529	106169	*O. rufipogon*	Vietnam	-	*W1*		0.87	0.00	0.00	0.00	0.00	0.13	*I-1*	-	-	1.000	-
517	106057	*O. rufipogon*	India	-	*W1*		0.82	0.00	0.00	0.05	0.01	0.12	*VIII-5*	-	-	1.000	-
509	105897	*O. rufipogon*	Bangladesh	+	*W1*		0.77	0.00	0.00	0.04	0.00	0.19	*VIII-2*	-	-	0.333	-
694	103306	*O. nivara*	Taiwan	-	*W1*		0.76	0.00	0.00	0.00	0.00	0.24	*VIII-1*	-	-	0.000	-
453	103404	*O. rufipogon*	Bangladesh	+	*Admix W1/ W2*		0.63	0.37	0.00	0.00	0.00	0.00	*III-1*	-	-	0.222	-
510	105898	*O. rufipogon*	Bangladesh	-	*Admix W1/ W2*		0.63	0.37	0.00	0.00	0.00	0.00	*III-1*	-	-	0.000	WT
533	106327	*O. rufipogon*	Cambodia	-	*Admix W1/ W2*		0.56	0.44	0.00	0.00	0.00	0.00	*III-3*	-	-	1.000	-
696	103415	*O. nivara*	Sri Lanka	-	*Admix W1/ W2*		0.50	0.50	0.00	0.00	0.00	0.00	*VIII-4*	-	Or-III	1.000	-
714	104058	*O. ruf. x O. niv.*	China	-	*Admix W1/ W2*		0.36	0.30	0.00	0.22	0.00	0.12	*I-1*	-	-	-	-
547	105908	*O. rufipogon*	Thailand	-	*Admix W1/ W3*		0.65	0.00	0.35	0.00	0.00	0.00	*V-1*	-	-	-	-
570	106266	*O. rufipogon*	Papua New Guine	-	*Admix W1/ W3*		0.54	0.00	0.46	0.00	0.00	0.00	*IV-2*	-	-	1.000	-
500	105785	*O. nivara*	Thailand	-	*Admix W1/ W4*		0.72	0.00	0.00	0.28	0.00	0.00	*VIII-1*	-	-	0.667	-
498	105735	*O. rufipogon*	Cambodia	+	*Admix W1/ W4*		0.71	0.00	0.00	0.29	0.00	0.00	*I-1*	-	-	1.000	-
484	105388	*O. rufipogon*	Thailand	+	*Admix W1/ W4*		0.68	0.00	0.00	0.32	0.00	0.00	*VIII-3*	-	-	1.000	-
520	106144	*O. rufipogon*	India	-	*Admix W1/ W4*		0.68	0.00	0.00	0.27	0.00	0.06	*III-2*	-	-	1.000	-
462	104501	*O. rufipogon*	India	-	*Admix W1/ W4*		0.65	0.00	0.00	0.28	0.00	0.07	*VIII-8*	-	-	1.000	-
477	104967	*O. spontanea*	China	+	*Admix W1/ W4*		0.65	0.00	0.00	0.35	0.00	0.00	*I-4*	-	-	1.000	WT
665	100203	*O. ruf. x O. sat.*	Myanmar	+	*Admix W1/ W4*		0.64	0.00	0.00	0.18	0.00	0.18	*VIII-10*	-	Or-I	1.000	WT
568	106263	*O. rufipogon*	Papua New Guine	+	*Admix W1/ W4*		0.63	0.00	0.00	0.37	0.00	0.00	*VIII-6*	-	-	1.000	WT
518	106078	*O. rufipogon*	India	-	*Admix W1/ W4*		0.47	0.12	0.00	0.41	0.00	0.00	*I-1*	-	-	1.000	-
664	100196	*O. nivara*	Myanmar	-	*Admix W1/ W5*	W7	0.65	0.00	0.00	0.00	0.33	0.02	*I-1*	-	Or-III	-	-
449	100195	*O. nivara*	Myanmar	+	*Admix W1/ W5*	W7	0.62	0.00	0.00	0.00	0.33	0.05	*In-3*	-	-	1.000	-
410	80759	*O. nivara*	Myanmar	+	*Admix W1/ W5*	W7	0.62	0.00	0.00	0.00	0.35	0.03	*I-2*	-	-	1.000	-
760	106345	*O. nivara*	Myanmar	+	*Admix W1/ W5*	W7	0.61	0.00	0.00	0.00	0.37	0.02	*I-1*	-	-	1.000	-
702	103814	*O. niv. x O. ruf.*	China	-	*Admix W1/ W6*		0.75	0.00	0.00	0.00	0.00	0.25	*VIII-4*	-	-	1.000	-
686	100926	*O. rufipogon*	Myanmar	+	*Admix W1/ W6*		0.74	0.00	0.00	0.11	0.01	0.14	*VIII-1*	-	Or-II	1.000	-
438	86476	*O. rufipogon*	India	+	*Admix W1/ W6*		0.71	0.00	0.00	0.00	0.03	0.26	*VII-4*	-	-	0.778	-
562	105890	*O. rufipogon*	Bangladesh	-	*Admix W1/ W6*		0.68	0.00	0.00	0.04	0.00	0.28	*III-8*	-	-	-	-
429	82990	*O. rufipogon*	China	-	*Admix W1/ W6*		0.67	0.00	0.00	0.00	0.00	0.33	*VIII-7*	-	-	1.000	-
663	99556	*O. rufipogon*	China	-	*Admix W1/ W6*		0.66	0.00	0.00	0.00	0.00	0.34	*I-1*	-	-	-	-
437	86475	*O. rufipogon*	India	-	*Admix W1/ W6*		0.65	0.00	0.00	0.00	0.04	0.31	*VIII-1*	-	-	1.000	-
406	80592	*O. rufipogon*	India	-	*Admix W1/ W6*		0.64	0.06	0.00	0.06	0.01	0.23	*VIII-3*	-	-	-	-
455	103823	*O. rufipogon*	China	-	*Admix W1/ W6*		0.63	0.00	0.00	0.00	0.00	0.37	*VIII-1*	-	-	-	-
549	81881	*O. rufipogon*	India	+	*Admix W1/ W6*		0.62	0.00	0.00	0.06	0.09	0.24	*VIII-8*	-	-	1.000	-
599	100183	*O. ruf. x O. sat.*	India	-	*Admix W1/ W6*		0.61	0.01	0.00	0.12	0.00	0.26	*V-3*	-	-	1.000	-
550	81887	*O. rufipogon*	India	-	*Admix W1/ W6*		0.61	0.00	0.00	0.00	0.04	0.35	*VIII-3*	-	-	-	-
713	104056	*O. ruf. x O. niv.*	China	-	*Admix W1/ W6*		0.61	0.00	0.00	0.00	0.00	0.39	*VIII-1*	-	-	-	-
403	80562	*O. rufipogon*	India	-	*Admix W1/ W6*		0.57	0.02	0.00	0.17	0.00	0.24	*VIII-5*	-	-	-	-
427	82988	*O. rufipogon*	China	+	*Admix W1/ W6*		0.56	0.00	0.00	0.00	0.00	0.44	*I-1*	-	-	1.000	-
508	105890	*O. rufipogon*	Bangladesh	+	*Admix W1/ W6*		0.51	0.00	0.00	0.00	0.00	0.49	*III-1*	-	-	0.833	-
528	106168	*O. rufipogon*	Vietnam	-	*Admix W1/ W6*		0.43	0.25	0.00	0.00	0.00	0.32	*I-1*	-	-	-	-
448	100189	*O. nivara*	Malaysia	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	-	-	Or-III/Or-I	-	-
482	105349	*O. rufipogon*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
483	105375	*O. rufipogon*	Thailand	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
666	100211	*O. rufipogon*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-1*	-	-	1.000	-
481	105343	*O. nivara*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
725	105319	*O. nivara*	India	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*I-2*	-	-	1.000	-
721	104705	*O. nivara*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*III-4*	-	-	1.000	-
555	105349	*O. rufipogon*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*V-2*	-	-	1.000	WT
450	100916	*O. nivara*	China	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-4*	-	-	1.000	-
494	105711	*O. rufipogon*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-
656	104443	*O. nivara*	Thailand	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-9*	-	-	-	-
655	100898	*O. nivara*	India	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-17*	II	Or-I	1.000	-
678	100898	*O. nivara*	India	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	-	II	Or-I	1.000	-
657	104705	*O. nivara*	India	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*In-14*	-	-	1.000	-
690	101942	*O. ruf. x O. niv.*	Malaysia	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-4*	-	Or-III	-	-
602	100900	*O. nivara*	India	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-5*	-	Or-III	0.833	-
658	100912	*O. niv. x O. ruf.*	Thailand	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*III-5*	-	Or-I	1.000	-
492	105569	*O. rufipogon*	Cambodia	+	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-5*	-	Or-I	1.000	WT
538	106410	*O. rufipogon*	Vietnam	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	-	-	-	1.000	WT
698	103418	*O. nivara*	Sri Lanka	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-5*	-	Or-II	-	-
491	105568	*O. rufipogon*	Philippines	-	*W2*		0.00	1.00	0.00	0.00	0.00	0.00	*VIII-5*	-	Or-III	1.000	-
463	104599	*O. rufipogon*	Sri Lanka	-	*W2*		0.12	0.88	0.00	0.00	0.00	0.00	*VIII-1*	-	-	1.000	-
487	105428	*O. nivara*	Sri Lanka	+	*W2*		0.12	0.88	0.00	0.00	0.00	0.00	*I-1*	-	-	1.000	-

A. *ORSC* accessions cont'd

NSF ID[1]	IRGC ID[2] /Wild ID[3]	Species	Country	Wild-RDP[4]	Subpopulation K=6	Subpopulation K=8	Ancestry at K=6 W1	W2	W3	W4	W5	W6	Chloroplast Haplotype	Previous genetic information Cheng's et al. (2003)[5]	Previous genetic information Huang's et al. (2012)[6]	Pericarp color scores	*Rc* 14bp deletion
732	105431	*O. nivara*	Sri Lanka	-	*W2*		0.17	0.83	0.00	0.00	0.00	0.00	*VIII-4*	-	-	-	-
733	105444	*O. nivara*	Sri Lanka	-	*W2*		0.21	0.79	0.00	0.00	0.00	0.00	*VIII-1*	-	-	1.000	-
695	103407	*O. nivara*	Sri Lanka	-	*Admix W2/ W1*		0.50	0.50	0.00	0.00	0.00	0.00	*I-2*	-	Or-III	-	-
680	100902	*O. niv. x O. ruf.*	India	-	*Admix W2/ W1*		0.41	0.55	0.00	0.04	0.00	0.00	*I-1*	-	Or-III	-	-
573	106269	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	-	-	-	-	-
579	106276	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
582	106279	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
575	106272	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
576	106273	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
530	106273	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
577	106274	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
578	106275	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
580	106277	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	-	-	-	-	-
583	106280	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
587	106285	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
590	106289	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	-	-	-	-	-
574	106270	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
724	104999	*O. rufipogon*	Indonesia	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	IV	Or-III	-	-
584	106282	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
424	81996	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-1*	-	-	-	-
591	106290	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	-	-	-	1.000	-
593	105757	*O. rufipogon*	Thailand	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-4*	-	-	-	-
585	106283	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
569	106264	*O. rufipogon*	Papua New Guine	-	*W3*		0.00	0.00	1.00	0.00	0.00	0.00	*IV-2*	-	-	-	-
447	93274	*O. rufipogon*	Indonesia	-	*W3*		0.08	0.00	0.92	0.00	0.00	0.00	*IV-1*	-	-	-	-
531	106283	*O. rufipogon*	Papua New Guine	-	*W3*		0.12	0.00	0.88	0.00	0.00	0.00	*IV-2*	-	-	-	-
581	106278	*O. rufipogon*	Papua New Guine	-	*Admix W3/ W1*		0.28	0.00	0.69	0.00	0.00	0.04	*VIII-1*	-	-	-	-
594	106412	*O. rufipogon*	Vietnam	-	*Admix W3/ W1*		0.42	0.00	0.58	0.00	0.00	0.00	*V-1*	-	-	-	-
588	106286	*O. rufipogon*	Papua New Guine	-	*Admix W3/ W6*		0.23	0.00	0.45	0.00	0.00	0.32	*IV-1*	-	-	-	-
488	105491	*O. rufipogon*	Malaysia	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	Or-I	1.000	-
691	101967	*O. nivara*	India	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*III-2*	-	Or-I	0.833	-
677	100897	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-9*	-	Or-I	-	-
711	103841	*O. nivara*	Bangladesh	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	-	1.000	WT
712	103845	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-2*	-	-	1.000	-
402	80539	*O. spontanea*	India	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	-	1.000	-
414	81903	*O. spontanea*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	-	1.000	-
506	105879	*O. nivara*	Bangladesh	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	-	0.778	-
679	100899	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-5*	-	Or-II	1.000	-
688	101450	*O. nivara*	Taiwan	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-18*	-	-	1.000	-
707	103835	*O. nivara*	Bangladesh	+	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	Or-I	1.000	-
709	103837	*O. nivara*	Bangladesh	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*III-1*	II	Or-I	-	-
558	105616	*O. rufipogon*	China	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-4*	III	-	1.000	-
405	80586	*O. spontanea*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*III-4*	-	-	1.000	-
710	103840	*O. nivara*	Bangladesh	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-2*	-	-	-	-
681	100903	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-4*	-	Or-III	1.000	-
740	105622	*O. rufipogon*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	-	-	-	-	-
404	80582	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*VIII-4*	-	-	1.000	WT
741	105624	*O. nivara*	India	-	*W4*		0.00	0.00	0.00	1.00	0.00	0.00	*I-2*	-	-	0.667	-
735	105493	*O. rufipogon*	Myanmar	-	*W4*		0.03	0.00	0.00	0.97	0.00	0.00	*I-1*	II	Or-I	1.000	-
742	105625	*O. nivara*	India	-	*W4*		0.18	0.00	0.00	0.82	0.00	0.00	*VIII-1*	II	Or-I	-	-
717	104650	*O. nivara*	Thailand	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-6*	-	-	1.000	-
737	105599	*O. nivara*	Thailand	-	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-5*	-	Or-I	1.000	-
716	104647	*O. rufipogon*	Thailand	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-4*	-	-	1.000	-
738	105601	*O. ruf. x O. niv.*	Thailand	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-6*	II	Or-I	1.000	-
473	104644	*O. nivara*	Thailand	-	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-2*	-	-	1.000	-
722	104962	*O. ruf. x O. niv.*	China	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*I-2*	-	-	0.889	-
727	105391	*O. nivara*	Thailand	-	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-5*	-	-	1.000	-
736	105494	*O. rufipogon*	Myanmar	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	-	-	Or-I	0.667	-
749	105867	*O. nivara*	Thailand	-	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-1*	-	-	1.000	-
556	105494	*O. rufipogon*	Myanmar	-	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*I-1*	-	Or-I	1.000	-
723	104969	*O. nivara*	China	+	*W4*	W8	0.00	0.00	0.00	1.00	0.00	0.00	*VIII-6*	-	-	1.000	WT
762	106396	*O. nivara*	Myanmar	+	*W4*	W8	0.17	0.00	0.00	0.83	0.00	0.00	*I-1*	-	-	1.000	-
693	102116	*O. ruf. x O. niv.*	Cambodia	-	*Admix W4/ W1*	W8	0.29	0.00	0.00	0.71	0.00	0.00	*I-1*	-	-	1.000	-
747	105742	*O. nivara*	Cambodia	-	*Admix W4/ W1*	W8	0.29	0.00	0.00	0.71	0.00	0.00	*VIII-1*	-	-	-	-
748	105763	*O. nivara*	Thailand	-	*Admix W4/ W1*	W8	0.29	0.00	0.00	0.71	0.00	0.00	*VIII-6*	-	-	0.000	-
683	100918	*O. nivara*	Cambodia	+	*Admix W4/ W1*	W8	0.35	0.00	0.00	0.65	0.00	0.00	*III-7*	-	-	0.778	-
746	105740	*O. nivara*	Cambodia	+	*Admix W4/ W1*	W8	0.36	0.00	0.00	0.64	0.00	0.00	*VIII-1*	-	-	0.333	-
718	104670	*O. nivara*	Thailand	-	*Admix W4/ W1*	W8	0.37	0.00	0.00	0.63	0.00	0.00	*VIII-4*	-	-	-	-
757	106148	*O. nivara*	Laos	+	*Admix W4/ W1*	W8	0.38	0.00	0.00	0.62	0.00	0.00	*I-1*	-	-	1.000	-
744	105716	*O. nivara*	Cambodia	-	*Admix W4/ W1*	W8	0.39	0.00	0.00	0.61	0.00	0.00	*I-2*	-	-	-	-
501	105821	*O. nivara*	Thailand	+	*Admix W4/ W1*	W8	0.41	0.00	0.00	0.59	0.00	0.00	*VIII-6*	-	-	1.000	-
475	104823	*O. nivara*	Thailand	-	*Admix W4/ W1*	W8	0.43	0.00	0.00	0.57	0.00	0.00	*III-3*	-	-	1.000	-

A. *ORSC* accessions cont'd

NSF ID[1]	IRGC ID[2] /Wild ID[3]	Species	Country	Wild-RDP[4]	Subpopulation K=6	Subpopulation K=8	Ancestry at K=6 W1	W2	W3	W4	W5	W6	Chloroplast Haplotype	Previous genetic information Cheng's et al. (2003)[5]	Previous genetic information Huang's et al. (2012)[6]	Pericarp color scores	*Rc* 14bp deletion
495	105717	*O. nivara*	Cambodia	+	*Admix W4/ W1*	W8	0.43	0.00	0.00	0.57	0.00	0.00	*III-1*	-	-	1.000	-
523	106155	*O. nivara*	Laos	+	*Admix W4/ W1*	W8	0.49	0.00	0.00	0.51	0.00	0.00	*VIII-1*	-	-	1.000	-
708	103836	*O. nivara*	Bangladesh	+	*Admix W4/ W1*		0.33	0.00	0.00	0.63	0.00	0.04	*I-1*	-	Or-I	1.000	-
428	82989	*O. rufipogon*	China	+	*Admix W4/ W1*		0.16	0.19	0.00	0.60	0.00	0.06	*III-2*	-	-	1.000	-
685	100923	*O. rufipogon*	Myanmar	+	*Admix W4/ W1*		0.41	0.00	0.00	0.57	0.00	0.02	*VIII-6*	II	Or-I	0.222	14bp del
715	104497	*O. ruf. x O. niv.*	Thailand	+	*Admix W4/ W1*		0.44	0.00	0.00	0.56	0.00	0.00	*VIII-5*	-	-	1.000	-
401	80433	*O. rufipogon*	India	+	*Admix W4/ W1*		0.37	0.00	0.00	0.54	0.00	0.09	*I-1*	-	-	1.000	-
728	105397	*O. ruf. x O. niv.*	China	-	*Admix W4/ W1*		0.32	0.00	0.00	0.46	0.00	0.22	*I-2*	-	-	-	-
674	100657	*O. rufipogon*	Taiwan	-	*Admix W4/ W6*		0.00	0.00	0.00	0.60	0.00	0.40	*VIII-15*	-	-	-	-
415	81909	*O. spontanea*	India	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	WT
493	105706	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	-
607	102178	*O. nivara*	India	-	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-2*	-	-	1.000	-
443	93183	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*In-2*	-	-	1.000	-
444	93188	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	WT
442	93181	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	-
445	93189	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	-	-	-	1.000	-
719	104687	*O. nivara*	India	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VIII-4*	-	-	1.000	WT
720	104703	*O. nivara*	India	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*I-2*	-	-	1.000	-
743	105705	*O. nivara*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	-
413	81850	*O. nivara*	India	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*VI-1*	-	-	1.000	-
446	93224	*O. spontanea*	Nepal	+	*W5*		0.00	0.00	0.00	0.00	1.00	0.00	*In-11*	-	-	1.000	-
451	101508	*O. nivara*	India	+	*Admix W5/ W4*		0.00	0.00	0.00	0.42	0.58	0.00	*VI-3*	-	-	1.000	WT
669	100593	*O. nivara*	Taiwan	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-2*	-	-	0.000	14bp del
485	105400	*O. rufipogon*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*III-2*	-	-	-	-
668	100588	*O. rufipogon*	Taiwan	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*III-6*	-	-	-	-
604	100907	*O. ruf. x O. niv.*	Taiwan	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-2*	-	-	0.000	14bp del
676	100692	*O. rufipogon*	Taiwan	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-10*	-	-	0.000	14bp del
486	105402	*O. rufipogon*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	1.000	WT
476	104959	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	-	-
672	100639	*O. rufipogon*	Taiwan	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-16*	-	-	-	-
673	100647	*O. rufipogon*	Taiwan	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*In-7*	-	-	1.000	-
692	101979	*O. nivara*	India	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	-
467	104624	*O. rufipogon*	China	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	1.000	-
470	104632	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-3*	-	-	1.000	-
662	99555	*O. rufipogon*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	1.000	-
465	104620	*O. spontanea*	China	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	1.000	WT
469	104628	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	1.000	WT
478	104971	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-12*	-	-	1.000	-
765	W1943[c]	*O. rufipogon*	-	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-14*	I	Or-III	-	-
551	100596	*O. ruf. x O. niv.*	Taiwan	+	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*In-13*	-	-	1.000	-
729	105403	*O. ruf. x O. niv.*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-13*	-	-	-	-
468	104626	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-3*	-	-	1.000	-
466	104621	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	-
705	103822	*O. ruf. x O. niv.*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*VIII-4*	-	-	-	-
767	W1945[c]	*O. rufipogon*	-	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	I	Or-III	-	-
661	99554	*O. rufipogon*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-1*	-	-	-	-
452	103308	*O. rufipogon*	Taiwan	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-2*	-	-	1.000	WT
671	100599	*O. rufipogon*	Taiwan	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-8*	-	-	-	-
471	104634	*O. spontanea*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*III-2*	-	-	1.000	-
660	82993	*O. rufipogon*	China	-	*W6*		0.00	0.00	0.00	0.00	0.00	1.00	*I-7*	-	-	-	-
472	104636	*O. spontanea*	China	+	*W6*		0.01	0.00	0.00	0.00	0.00	0.99	*I-1*	-	-	0.667	-
706	103825	*O. niv. x O. ruf.*	China	-	*W6*		0.00	0.00	0.00	0.07	0.00	0.93	-	-	Or-III	0.000	14bp del
675	100678	*O. rufipogon*	Taiwan	-	*W6*		0.00	0.07	0.00	0.00	0.00	0.93	*I-7*	-	-	-	-
766	W1944[c]	*O. rufipogon*	-	-	*W6*		0.09	0.00	0.00	0.00	0.00	0.91	*I-1*	-	-	-	-
670	100597	*O. rufipogon*	Taiwan	-	*W6*		0.12	0.00	0.00	0.00	0.00	0.88	*VIII-11*	-	-	0.000	14bp del
751	105895	*O. nivara*	Bangladesh	+	*W6*		0.16	0.00	0.00	0.00	0.00	0.84	*I-1*	-	-	1.000	WT
454	103821	*O. nivara*	China	+	*W6*		0.00	0.00	0.00	0.16	0.00	0.84	*I-1*	-	Or-I	0.333	-
430	82991	*O. rufipogon*	China	-	*W6*		0.19	0.00	0.00	0.00	0.00	0.81	*I-1*	-	-	1.000	WT
703	103817	*O. ruf. x O. niv.*	China	-	*W6*		0.25	0.00	0.00	0.00	0.00	0.75	-	-	Or-II	1.000	-
461	104057	*O. rufipogon*	China	+	*W6*		0.25	0.00	0.00	0.00	0.00	0.75	*I-1*	-	-	0.556	-
687	101193	*O. rufipogon*	Taiwan	+	*Admix W6/ W1*		0.27	0.00	0.00	0.00	0.00	0.73	*VIII-4*	-	-	1.000	-
456	103824	*O. nivara*	China	-	*Admix W6/ W1*		0.31	0.00	0.00	0.01	0.00	0.68	*VIII-1*	-	Or-I	1.000	-
567	106167	*O. rufipogon*	Vietnam	-	*Admix W6/ W1*		0.37	0.00	0.00	0.00	0.00	0.63	*I-1*	-	-	-	-
564	106138	*O. rufipogon*	India	-	*Admix W6/ W1*		0.42	0.00	0.00	0.04	0.00	0.54	-	-	-	-	-
704	103818	*O. ruf. x O. niv.*	China	+	*Admix W6/ W1*		0.37	0.00	0.00	0.13	0.00	0.51	*I-1*	-	Or-I	0.222	14bp del
457	103838	*O. nivara*	Bangladesh	+	*Admix W6/ W1*		0.38	0.00	0.00	0.13	0.00	0.49	*VIII-3*	-	Or-I	1.000	-
559	105618	*O. rufipogon*	China	-	*Admix W6/ W1*		0.30	0.00	0.00	0.24	0.00	0.46	*V-2*	-	-	1.000	-
431	82992	*O. rufipogon*	China	+	*Admix W6/ W2*		0.13	0.42	0.00	0.00	0.00	0.45	*I-1*	-	-	1.000	-
701	103813	*O. niv. x O. ruf.*	China	+	*Admix W6/ W5*		0.29	0.00	0.00	0.04	0.31	0.36	*VIII-1*	-	Or-II	0.000	14bp del

1 National Science Foundation -"Exploring the Genetic Basis of Transgressive Variation in Rice" project accession Identification Number

2 International Rice Germplasm Collection Identification Number

3 Wild Identification Number. Seed source from National Institute of Genetics, Japan

4 Wild Rice Diversity Panel

5 Cheng C, Motohashi R, Tsuchimoto S, et al (2003) Polyphyletic Origin of Cultivated Rice: Based on the Interspersion Pattern of SINEs. Mol Biol Evol 20:67–75.

6 Huang X, Kurata N, Wei X, et al (2012) A map of rice genome variation reveals the origin of cultivated rice. Nature 490:497–501.

Table S1. Germplasm information, cont'd

B. *O. sativa* accessions						
NSF ID	IRGC ID	Subpopulation	Country	Chloroplast Haplotype	Sweeney et al, (2007)[2] Rc phenotype	Rc haplotype
13	117605	*aus*	Pakistan	*VIII-2*	-	-
18	117661	*aus*	India	*I-1*	red	7
28 [1]	-	*aus*	Thailand	*VIII-2*	white	8
44	117710	*aus*	Bangladesh	*VIII-2*	white	2
49	117725	*aus*	Bangladesh	*VIII-5*	red	10
78	117769	*aus*	India	*VIII-2*	-	-
85	117617	*aus*	India	*VIII-2*	red	10
88	117781	*aus*	Thailand	*VIII-2*	white	8
131	117850	*aus*	Bhutan	*VIII-2*	white	8
152	117902	*aus*	India	*II-1*	-	-
17	117659	*indica*	Philippines	-	white	5
29	117682	*indica*	Vietnam	*VIII-3*	white	5
30	117684	*indica*	Vietnam	-	white	2
43	117705	*indica*	Taiwan	*VIII-2*	white	2
61	117745	*indica*	China	*VIII-2*	-	-
76 [1]	117766	*indica*	India	*VIII-2*	white	5
110	117818	*indica*	India	*VIII-2*	red	7
132	117859	*indica*	Sri Lanka	*I-5*	white	2
161	117912	*indica*	China	*VIII-3*	white	5
398	-	*indica*	China	*VIII-2*	white	5
612	-	*indica*	Philippines	*I-5*	white	5
7 [1]	126381	*tropical japonica*	Indonesia	*I-2*	white	2
8	117643	*tropical japonica*	Philippines	*I-3*	white	2
54	117736	*tropical japonica*	US	-	white	2
95 [1]	117790	*tropical japonica*	-	*I-5*	-	-
101	117802	*tropical japonica*	US	*I-5*	white	2
107	117815	*tropical japonica*	Bangladesh	*I-2*	white	2
108	117621	*tropical japonica*	Guinea	*I-3*	white	2
165	117921	*tropical japonica*	Indonesia	*I-3*	white	2
174	-	*tropical japonica*	Philippines	*I-9*	white	2
397	117699	*tropical japonica*	US	*In-12*	-	-
628	126385	*tropical japonica*	US	*I-5*	white	2
60	117744	*tropical japonica* x *temperate japonica*	Indonesia	*I-3*	white	2
5	117641	*aromatic*	India	*I-1*	-	-
12	117652	*aromatic*	Pakistan	*I-1*	-	-
14 [1]	117653	*aromatic*	India	*I-1*	white	8
16	117658	*aromatic*	Brazil	*In-1*	-	-
45	117721	*aromatic*	Iran	*In-1*	-	-
53	117735	*aromatic*	Iran	-	white	2
93	117787	*aromatic*	Madagascar	*I-1*	white	2
1	126380	*temperate japonica*	Italy	-	white	2
31	117686	*temperate japonica*	China	*I-4*	white	2
56	117612	*temperate japonica*	Korea	*I-4*	-	-
94	117789	*temperate japonica*	Japan	*I-4*	white	2
104	117811	*temperate japonica*	Japan	-	-	-
113	117822	*temperate japonica*	Japan	*I-4*	-	-
143	117884	*temperate japonica*	Japan	*I-4*	-	-
144	117887	*temperate japonica*	US	*I-3*	-	-
151	117900	*temperate japonica*	Korea	*I-6*	-	-
173	121592	*temperate japonica*	Japan	*I-4*	-	-

1 accession only used for chloroplast analysis

2 Sweeney MT, Thomson MJ, Cho YG, et al (2007) Global dissemination of a single mutation conferring white pericarp in rice. PLoS Genet 3:1418–1423.

Table S1. Germplasm information, cont'd

C. Wild rice accessions from different species

NSF ID	IRGC ID or Genbank ID	Species	Chloroplast Haplotype
-	GU592208	*O. meridionalis*	*In-6*
-	JN005831	*O. meridionalis*	*IV-3*
-	NC_016927	*O. meridionalis*	*IV-3*
-	JN005833	*O. rufipogon*	*IV-1*
479	105220	*O. officinalis*	*In-8*
-	GU592209	*O. australiensis*	*In-9*
-	KJ830774	*O. australiensis*	*In-10*

Table S2. Pairwise Fst and genetic distance among six *ORSC* and five *O. sativa* subpopulations based on GBS-SNP data.

A. Pairwise mean Fst and Genetic distance between subgroups of the *ORSC* and *O. sativa* [1].

	W1	W2	W3	W4	W5	W6	*aus*	*indica*	*japonica*
W1		0.074 (0.029-0.138)	0.046 (0.017-0.098)	0.046 (0.019-0.086)	0.070 (0.026-0.130)	0.033 (0.013-0.065)	0.042 (0.015-0.086)	0.028 (0.011-0.056)	0.044 (0.018-0.085)
W2	0.332 (0.182-0.530)		0.088 (0.037-0.160)	0.086 (0.031-0.157)	0.079 (0.032-0.140)	0.075 (0.028-0.137)	0.078 (0.028-0.140)	0.071 (0.028-0.127)	0.081 (0.033-0.146)
W3	0.220 (0.100-0.368)	0.550 (0.350-0.762)		0.065 (0.027-0.127)	0.083 (0.036-0.144)	0.052 (0.021-0.093)	0.061 (0.025-0.115)	0.049 (0.023-0.083)	0.062 (0.028-0.108)
W4	0.181 (0.084-0.328)	0.395 (0.211-0.614)	0.408 (0.239-0.623)		0.080 (0.033-0.136)	0.049 (0.021-0.088)	0.036 (0.013-0.073)	0.037 (0.014-0.072)	0.057 (0.024-0.100)
W5	0.338 (0.158-0.532)	0.492 (0.260-0.737)	0.769 (0.484-0.926)	0.467 (0.259-0.695)		0.069 (0.028-0.121)	0.071 (0.029-0.124)	0.066 (0.029-0.110)	0.075 (0.033-0.125)
W6	0.116 (0.045-0.208)	0.411 (0.223-0.589)	0.427 (0.215-0.638)	0.274 (0.137-0.432)	0.507 (0.271-0.697)		0.042 (0.016-0.079)	0.028 (0.012-0.050)	0.028 (0.012-0.054)
aus	0.177 (0.015-0.410)	0.468 (0.233-0.693)	0.674 (0.436-0.865)	0.133 (-0.00-0.359)	0.691 (0.359-0.914)	0.330 (0.108-0.568)		0.033 (0.013-0.065)	0.046 (0.018-0.083)
indica	0.065 (-0.03-0.189)	0.450 (0.241-0.682)	0.643 (0.425-0.828)	0.187 (0.041-0.422)	0.706 (0.377-0.906)	0.202 (0.043-0.368)	0.418 (0.109-0.742)		0.037 (0.015-0.067)
japonica	0.214 (0.079-0.409)	0.501 (0.290-0.729)	0.635 (0.353-0.890)	0.368 (0.177-0.603)	0.679 (0.414-0.939)	0.192 (0.077-0.389)	0.524 (0.211-0.844)	0.467 (0.155-0.825)	

1 Genetic distance above diagonal and Fst below diagnal with 95% Confidence Interval in parenthesis.

B. Mean *Fst* and Genetic distance of one subgroup with the rest of subgroups in the *ORSC*.

	W1	W2	W3	W4	W5	W6
Fst	0.237	0.436	0.475	0.345	0.515	0.347
Genetic Distance	0.054	0.080	0.067	0.065	0.076	0.056

Table S3. Pericarp and hull color of 157 ORSC accessions grouped by subpopulation (at K=6). Number of *ORSC* accessions with hull color and pericarp color phenotypes, grouped by subpopulation.

	ORSC subpopulation identity at K=6													
	W1	W2	W4	W5	W6	Admix W1/ W2	Admix W1/ W4	Admix W1/ W5	Admix W1/ W6	Admix W4/ W1	Admix W5/ W4	Admix W6/ W1	Admix W6/ W2	Admix W6/ W5
White hull / white pericarp	2	0	0	0	3	0	0	0	0	0	0	0	0	0
White hull / red pericarp	2	0	0	0	1	0	2	0	0	0	1	0	0	0
Black hull / white pericarp	4	0	0	0	2	2	0	0	0	1	0	1	0	1
Black hull / red pericarp	28	19	27	12	14	2	5	3	8	12	0	4	1	0
Total	36	19	27	12	20	4	7	3	8	13	1	5	1	1
Grand total	157													

Table S4: Number of accessions included in Rc extended haplotype analysis (*ORSC* N=81; *O. sativa* N=405)

	***ORSC* subpopulation identity at K=6**														***O. sativa* subpopulation identity [1]**							
Rc Extended haplotype	**W1**	**W2**	**W4**	**W5**	**W6**	**Admix W1/ W2**	**Admix W1/ W4**	**Admix W1/ W5**	**Admix W1/ W6**	**Admix W4/ W1**	**Admix W5/ W4**	**Admix W6/ W1**	**Admix W6/ W2**	**Admix W6/ W5**	***indica***	***aus***	**temperate-*japonica***	**tropical-*japonica***	**aromatic**	**admixed**	**admixed-*indica***	**admixed-*japonica***
1					1												76	17	11	3		15
2									1							2						
3					1							1		1	7		13	78		4		18
4	2				1					1					54					1	2	
5																	10	8				1
6	3				2	1									14	5	2	1				
7	2				1										1	1						
8			4												4	2					3	
9															2	45			1	1	1	
10					1							1										
11	7	12	5	1	2		4	3	2	8	1	1	1		1					1		
12				10																		

Table S5. Chi-square statistic between genetic subgroups and two major traditional species groups, *O. rufipogon* and *O. nivara*.

Subpopulation	**Species**					
	O. rufipogon	*O. nivara*	*O. spontanea*	*O. rufipogon* x *O. nivara* or *O. nivara* x *O. rufipogon*	*O. rufipogon* x *O. sativa* or *O. sativa* x *O. rufipogon*	Total
W1	77	1	4	2	1	85
W2	9	14	0	2	0	25
W3	22	0	0	0	0	22
W4	7	21	3	2	0	33
W5	0	10	2	0	0	12
W6	19	4	9	6	0	38
admix	37	22	1	9	2	71
SUM	171	72	19	21	3	286

Tests

N	DF	-LogLike	Rsquare
184	5	60.519449	0.5623

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	121.039	<.0001*
Pearson	108.103	<.0001*

Table S6A. Primers Used for Chloroplast Sequencing

No.	Primer information	Position (bp)[1]	Polymorphic sites (bp)
1	1. F: gccgctttagtccactcagccatc	7,888-	8127
2	R: tcaatgccttttttcaatggtctc	8,804	8143
3			8197-8198
4			8415
5			8538
6			8548-8616
7			8599
8			8631
9	2. F: tatttgcttctcctgatggttggt	12,013-	12170
10	R: gagcggagtagagcagtttggtag	12,913	12309-12310
11			12496
12			12548
13			12672-12675
14			12799
15			12819-12820
16	3. 1) F: agaatctggacccatcgt	56,210-	57026-57041
17	R: ttactatttctatctattcgattt	57,198	57043-57044
18			57069
19			57070-57071
20			57135-57136
21			57155
-	4. F: aaaacgttgtatttttgttt	76,604-	(Variable sites were not detected.)
-	R: ttctcgaggtataatgacag	76,773	
22	5. F: atctgcagcatttaaaagggtctgaggttgaatcat	77,584-	77730-77731
23	R:aaagatctagatttcgtaaacaacatagaggaagaa	78,118	77741
24			77793-77794
25			77903

Table S7. Numeric LD Decay. LD decay by distance between pairs of SNPs. Mean values were calculated for all between-SNP distances that fell within a distance range centering on the values listed first column. All populations except W5 were subsampled 100 times to 12 individuals (the sample size of the W5 population). The values in parentheses are 95% confidence intervals calculated from these sub-sampling distributions.

Distance, kb	W1	W2	W3	W4	W5	W6
0.5	1 (1, 1)	1 (1, 1)	1 (1, 1)	1 (1, 1)	1	1 (1, 1)
1.5	0.533 (0.344, 0.615)	0.625 (0.563, 0.891)	0.862 (0.635, 0.952)	0.571 (0.503, 0.77)	1	0.571 (0.503, 0.643)
2.5	0.438 (0.326, 0.558)	0.556 (0.429, 0.651)	0.865 (0.643, 1)	0.523 (0.4, 0.596)	1	0.556 (0.385, 0.625)
6.0	0.369 (0.283, 0.463)	0.45 (0.36, 0.6)	0.803 (0.641, 0.89)	0.444 (0.382, 0.583)	1	0.438 (0.4, 0.54)
12.0	0.25 (0.2, 0.336)	0.4 (0.333, 0.567)	0.64 (0.556, 0.86)	0.376 (0.321, 0.517)	0.898	0.36 (0.311, 0.429)
18.0	0.229 (0.166, 0.297)	0.357 (0.286, 0.464)	0.6 (0.541, 0.858)	0.342 (0.28, 0.45)	0.894	0.308 (0.25, 0.36)
24.0	0.2 (0.151, 0.259)	0.333 (0.25, 0.429)	0.533 (0.425, 0.613)	0.333 (0.25, 0.415)	0.758	0.3 (0.243, 0.365)
30.0	0.2 (0.143, 0.248)	0.333 (0.25, 0.429)	0.513 (0.429, 0.601)	0.321 (0.25, 0.4)	0.667	0.256 (0.223, 0.337)
36.0	0.177 (0.128, 0.229)	0.3 (0.238, 0.412)	0.445 (0.372, 0.561)	0.286 (0.233, 0.372)	0.625	0.25 (0.213, 0.319)
42.0	0.174 (0.124, 0.226)	0.276 (0.229, 0.368)	0.425 (0.338, 0.531)	0.286 (0.229, 0.364)	0.598	0.25 (0.205, 0.301)
48.0	0.177 (0.119, 0.229)	0.262 (0.229, 0.375)	0.417 (0.349, 0.54)	0.25 (0.226, 0.359)	0.562	0.238 (0.2, 0.3)
54.0	0.167 (0.125, 0.222)	0.267 (0.22, 0.375)	0.359 (0.298, 0.466)	0.25 (0.229, 0.357)	0.583	0.229 (0.2, 0.301)
60.0	0.16 (0.115, 0.212)	0.267 (0.225, 0.375)	0.32 (0.244, 0.417)	0.25 (0.215, 0.346)	0.527	0.226 (0.177, 0.294)
66.0	0.161 (0.123, 0.222)	0.25 (0.214, 0.357)	0.357 (0.268, 0.505)	0.25 (0.209, 0.333)	0.462	0.229 (0.198, 0.284)
72.0	0.16 (0.115, 0.21)	0.25 (0.2, 0.357)	0.329 (0.243, 0.504)	0.238 (0.2, 0.316)	0.438	0.229 (0.191, 0.286)
78.0	0.149 (0.111, 0.217)	0.25 (0.2, 0.323)	0.266 (0.205, 0.393)	0.25 (0.207, 0.325)	0.438	0.222 (0.18, 0.255)
84.0	0.151 (0.111, 0.203)	0.248 (0.2, 0.33)	0.283 (0.206, 0.375)	0.238 (0.2, 0.313)	0.423	0.208 (0.17, 0.25)
90.0	0.143 (0.111, 0.211)	0.238 (0.2, 0.327)	0.225 (0.162, 0.32)	0.232 (0.2, 0.307)	0.4	0.206 (0.167, 0.25)
96.0	0.143 (0.111, 0.201)	0.229 (0.183, 0.308)	0.238 (0.173, 0.357)	0.229 (0.2, 0.312)	0.4	0.201 (0.162, 0.25)
102.0	0.143 (0.111, 0.198)	0.229 (0.185, 0.307)	0.222 (0.177, 0.299)	0.229 (0.2, 0.3)	0.302	0.2 (0.165, 0.25)
108.0	0.143 (0.111, 0.2)	0.229 (0.184, 0.308)	0.2 (0.147, 0.25)	0.229 (0.191, 0.286)	0.4	0.2 (0.162, 0.25)
114.0	0.149 (0.111, 0.206)	0.229 (0.18, 0.312)	0.178 (0.142, 0.238)	0.229 (0.184, 0.3)	0.333	0.2 (0.167, 0.25)
120.0	0.143 (0.111, 0.207)	0.229 (0.179, 0.312)	0.175 (0.135, 0.243)	0.229 (0.168, 0.286)	0.341	0.2 (0.158, 0.244)
126.0	0.143 (0.111, 0.189)	0.222 (0.167, 0.286)	0.167 (0.128, 0.226)	0.223 (0.184, 0.286)	0.333	0.201 (0.163, 0.25)
132.0	0.143 (0.111, 0.197)	0.229 (0.184, 0.3)	0.168 (0.127, 0.23)	0.216 (0.179, 0.258)	0.357	0.2 (0.16, 0.25)
138.0	0.14 (0.11, 0.2)	0.229 (0.181, 0.3)	0.156 (0.126, 0.232)	0.218 (0.17, 0.262)	0.3	0.2 (0.15, 0.25)
144.0	0.143 (0.111, 0.194)	0.228 (0.167, 0.282)	0.175 (0.133, 0.238)	0.2 (0.16, 0.25)	0.25	0.2 (0.158, 0.238)
150.0	0.141 (0.111, 0.189)	0.223 (0.181, 0.286)	0.161 (0.127, 0.21)	0.2 (0.167, 0.25)	0.286	0.2 (0.154, 0.234)
156.0	0.141 (0.11, 0.189)	0.222 (0.167, 0.267)	0.174 (0.131, 0.222)	0.206 (0.167, 0.25)	0.327	0.2 (0.155, 0.229)
162.0	0.14 (0.11, 0.188)	0.229 (0.178, 0.29)	0.143 (0.111, 0.184)	0.2 (0.167, 0.25)	0.256	0.2 (0.151, 0.243)
168.0	0.141 (0.109, 0.188)	0.222 (0.17, 0.286)	0.158 (0.124, 0.232)	0.199 (0.155, 0.244)	0.25	0.184 (0.143, 0.229)
174.0	0.141 (0.11, 0.184)	0.229 (0.17, 0.287)	0.143 (0.106, 0.193)	0.2 (0.16, 0.25)	0.25	0.2 (0.149, 0.25)
180.0	0.138 (0.109, 0.188)	0.229 (0.168, 0.299)	0.16 (0.128, 0.217)	0.2 (0.151, 0.229)	0.25	0.197 (0.143, 0.238)
186.0	0.128 (0.109, 0.184)	0.215 (0.167, 0.259)	0.133 (0.111, 0.199)	0.2 (0.16, 0.234)	0.25	0.183 (0.143, 0.229)
192.0	0.141 (0.108, 0.177)	0.223 (0.167, 0.293)	0.143 (0.114, 0.19)	0.2 (0.16, 0.229)	0.214	0.184 (0.147, 0.234)
198.0	0.129 (0.11, 0.193)	0.229 (0.174, 0.3)	0.128 (0.11, 0.167)	0.2 (0.143, 0.25)	0.2	0.2 (0.143, 0.234)
204.0	0.139 (0.104, 0.175)	0.229 (0.167, 0.293)	0.13 (0.11, 0.187)	0.2 (0.143, 0.234)	0.202	0.182 (0.143, 0.229)
210.0	0.128 (0.11, 0.177)	0.229 (0.167, 0.28)	0.139 (0.111, 0.19)	0.2 (0.148, 0.229)	0.229	0.17 (0.143, 0.224)
216.0	0.14 (0.106, 0.199)	0.222 (0.167, 0.285)	0.128 (0.0965, 0.167)	0.2 (0.163, 0.25)	0.286	0.167 (0.143, 0.234)
222.0	0.128 (0.104, 0.177)	0.218 (0.167, 0.284)	0.128 (0.0969, 0.163)	0.2 (0.155, 0.236)	0.228	0.2 (0.143, 0.229)
228.0	0.128 (0.1, 0.172)	0.229 (0.167, 0.32)	0.128 (0.099, 0.167)	0.2 (0.143, 0.238)	0.229	0.167 (0.142, 0.223)
234.0	0.128 (0.11, 0.186)	0.229 (0.167, 0.311)	0.125 (0.0977, 0.176)	0.186 (0.143, 0.229)	0.216	0.173 (0.143, 0.226)
240.0	0.128 (0.105, 0.199)	0.213 (0.166, 0.267)	0.128 (0.0857, 0.177)	0.185 (0.143, 0.244)	0.223	0.174 (0.143, 0.229)
246.0	0.128 (0.102, 0.177)	0.213 (0.158, 0.286)	0.12 (0.0933, 0.143)	0.2 (0.146, 0.23)	0.229	0.185 (0.143, 0.229)
252.0	0.128 (0.104, 0.197)	0.214 (0.163, 0.267)	0.127 (0.1, 0.188)	0.2 (0.144, 0.248)	0.229	0.165 (0.128, 0.22)
258.0	0.127 (0.101, 0.167)	0.214 (0.154, 0.286)	0.118 (0.0924, 0.162)	0.2 (0.143, 0.238)	0.229	0.167 (0.141, 0.229)
264.0	0.128 (0.0983, 0.177)	0.2 (0.154, 0.253)	0.108 (0.0845, 0.141)	0.2 (0.143, 0.25)	0.229	0.167 (0.143, 0.223)
270.0	0.139 (0.111, 0.195)	0.218 (0.167, 0.258)	0.111 (0.0787, 0.147)	0.198 (0.143, 0.233)	0.2	0.18 (0.143, 0.228)
276.0	0.128 (0.109, 0.17)	0.229 (0.163, 0.3)	0.128 (0.101, 0.194)	0.2 (0.152, 0.25)	0.198	0.16 (0.127, 0.223)
282.0	0.128 (0.111, 0.177)	0.204 (0.143, 0.286)	0.127 (0.0958, 0.167)	0.2 (0.143, 0.25)	0.175	0.167 (0.128, 0.223)
288.0	0.128 (0.1, 0.16)	0.2 (0.144, 0.25)	0.111 (0.0862, 0.145)	0.184 (0.143, 0.229)	0.22	0.167 (0.127, 0.22)
294.0	0.128 (0.107, 0.171)	0.222 (0.16, 0.299)	0.113 (0.0816, 0.166)	0.2 (0.143, 0.25)	0.2	0.166 (0.127, 0.221)
300.0	0.134 (0.1, 0.178)	0.211 (0.143, 0.271)	0.113 (0.0685, 0.158)	0.184 (0.143, 0.25)	0.168	0.162 (0.127, 0.218)

Table S8. SNP Information for Rc Extended Haplotypes Filtered SNPs from 576 kb region around Rc gene used to create extended haplotypes. SNPs extracted from McCouch et al. (2016) dataset generated using the High Density Rice Array (HDRA) and available on www.ricediversity.org/data

NSF_ID	Variety_name	Subpopulation [1, 2]	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5725405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
1	Agostano	temperate-japonica	O. sativa	1	_	_	_	_	1																																					
2	Aichi Asahi	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
3	Ai-Chiao-Hong	indica	O. sativa	_	_	_	_	_	4																																					
4	NSFTV4	aus	O. sativa	_	_	_	_	_	9																																					
5	NSFTV5	aromatic	O. sativa	5	_	_	_	_	1																																					
6	ARC 7229	aus	O. sativa	_	_	_	_	_	9																																					
7	Arias	tropical-japonica	O. sativa	7	_	_	_	_	3																																					
8	Asse Y Pung	tropical-japonica	O. sativa	8	_	_	_	_	3																																					
9	Baber	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
10	Baghlani Nangarhar	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
11	Baguamon 14	indica	O. sativa	_	_	_	_	_	6																																					
13	NSFTV13	aus	O. sativa	13	_	_	_	_	9																																					
15	Beonjo	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
16	Bico Branco	aromatic	O. sativa	16	_	_	_	_	1																																					
17	Binulawan	indica	O. sativa	17	_	_	_	_	4																																					
18	BJ 1	aus	O. sativa	18	_	_	_	_	9																																					
19	Black Gora	aus	O. sativa	_	_	_	_	_	9																																					
20	Blue Rose	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
21	Byakkoku Y 5006 Seln	indica	O. sativa	_	_	_	_	_	4																																					
22	Caawa/Fortuna 6-103-15	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
23	Canella De Ferro	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
24	Carolina Gold 12033	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
25	Carolina Gold 12034	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
26	Carolina Gold Sel	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
27	NSFTV27	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
28	Champa Tong 54	aus	O. sativa	28	_	_	_	_	9																																					
29	Chau	indica	O. sativa	29	_	_	_	_	11																																					
30	Chiem Chanh	indica	O. sativa	30	_	_	_	_	4																																					
31	Chinese	temperate-japonica	O. sativa	31	_	_	_	_	1																																					
32	Chodongji	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
33	Chuan 4	aus	O. sativa	_	_	_	_	_	6																																					
34	NSFTV34	indica	O. sativa	_	_	_	_	_	4																																					

.

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
35	CO18	indica	O. sativa	_	_	_	_	_	4																																				
36	CS-M3	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
37	Cuba 65	tropical-japonica	O. sativa	_	_	_	_	_	1																																				
39	NSFTV39	aus	O. sativa	_	_	_	_	_	7																																				
40	Dam	admixed-japonica	O. sativa	_	_	_	_	_	1																																				
41	Darmali	admixed	O. sativa	_	_	_	_	_	3																																				
43	Dee Geo Woo Gen	indica	O. sativa	43	_	_	_	_	3																																				
44	Dhala Shaitta	aus	O. sativa	44	_	_	_	_	6																																				
45	Dom Sufid	aromatic	O. sativa	45	_	_	_	_	1																																				
46	Dourado Agulha	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
48	NSFTV48	admixed-japonica	O. sativa	_	_	_	_	_	1																																				
49	DV85	aus	O. sativa	49	_	_	_	_	9																																				
51	Early Wataribune	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
52	Eh Ia Chiu	temperate-japonica	O. sativa	_	_	_	_	_	6																																				
53	Firooz	aromatic	O. sativa	53	_	_	_	_	1																																				
54	Fortuna	tropical-japonica	O. sativa	54	_	_	_	_	3																																				
55	Gerdeh	admixed-japonica	O. sativa	_	_	_	_	_	1																																				
56	Geumobyeo	temperate-japonica	O. sativa	56	_	_	_	_	1																																				
57	NSFTV57	indica	O. sativa	_	_	_	_	_	4																																				
58	Ghati Kamma Nangarhar	aus	O. sativa	_	_	_	_	_	9																																				
59	Gogo Lempuk	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
60	Gotak Gatik	admixed-japonica	O. sativa	60	_	_	_	_	3																																				
62	Gyehwa 3	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
63	Haginomae Mochi	temperate-japonica	O. sativa	_	_	_	_	_	5																																				
64	Heukgyeong	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
65	Honduras	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
66	Hsia Chioh Keh Tu	indica	O. sativa	_	_	_	_	_	6																																				
67	Hu Lo Tao	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
69	IAC 25	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
70	Iguape Cateto	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
71	IR 36	indica	O. sativa	_	_	_	_	_	4																																				
72	IR 8	indica	O. sativa	_	_	_	_	_	4																																				
73	IRAT 177	tropical-japonica	O. sativa	_	_	_	_	_	1																																				
74	IRGA 409	indica	O. sativa	_	_	_	_	_	4																																				
75	Jambu	tropical-japonica	O. sativa	_	_	_	_	_	6																																				

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
76	Jaya	indica	O. sativa	76	_	_	_	_	4																																					
77	JC149	indica	O. sativa	_	_	_	_	_	4																																					
78	Jhona 349	aus	O. sativa	78	_	_	_	_	9																																					
79	Jouiku 393G	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
80	K 65	admixed	O. sativa	_	_	_	_	_	3																																					
81	Kalamkati	aus	O. sativa	_	_	_	_	_	9																																					
83	Kamenoo	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
84	Kaniranga	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
85	Kasalath	aus	O. sativa	85	_	_	_	_	9																																					
87	Keriting Tingii	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
89	NSFTV89	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
90	Kiang-Chou-Chiu	indica	O. sativa	_	_	_	_	_	3																																					
91	Kibi	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
92	Kinastano	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
93	NSFTV93	aromatic	O. sativa	93	_	_	_	_	1																																					
94	Koshihikari	temperate-japonica	O. sativa	94	_	_	_	_	1																																					
95	NSFTV95	admixed	O. sativa	95	_	_	_	_	het																																					
96	KU115	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
98	L-202	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
99	LAC 23	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
100	Lacrosse	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
101	Lemont	tropical-japonica	O. sativa	101	_	_	_	_	3																																					
102	Leung Pratew	indica	O. sativa	_	_	_	_	_	4																																					
103	Luk Takhar	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
104	Mansaku	temperate-japonica	O. sativa	104	_	_	_	_	5																																					
105	Mehr	aus	O. sativa	_	_	_	_	_	9																																					
106	Ming Hui	indica	O. sativa	_	_	_	_	_	4																																					
107	NSFTV107	tropical-japonica	O. sativa	107	_	_	_	_	3																																					
108	Moroberekan	tropical-japonica	O. sativa	108	_	_	_	_	5																																					
110	Mudgo	indica	O. sativa	110	_	_	_	_	8																																					
112	N12	aromatic	O. sativa	_	_	_	_	_	9																																					
113	Norin 20	temperate-japonica	O. sativa	113	_	_	_	_	3																																					
114	Nova	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
115	NPE 835	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
116	NSFTV116	tropical-japonica	O. sativa	_	_	_	_	_	3																																					

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
117	O-Luen-Cheung	indica	O. sativa	_	_	_	_	_	3																																					
118	Oro	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
119	Oryzica Llanos 5	indica	O. sativa	_	_	_	_	_	4																																					
120	OS6	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
121	Ostiglia	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
122	Padi Kasalle	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
123	Pagaiyahan	indica	O. sativa	_	_	_	_	_	4																																					
124	Pankhari 203	aromatic	O. sativa	_	_	_	_	_	1																																					
125	Pao-Tou-Hung	indica	O. sativa	_	_	_	_	_	4																																					
126	Pappaku	indica	O. sativa	_	_	_	_	_	6																																					
127	NSFTV127	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
128	Pato De Gallinazo	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
129	Peh-Kuh	indica	O. sativa	_	_	_	_	_	4																																					
130	Peh-Kuh-Tsao-Tu	indica	O. sativa	_	_	_	_	_	4																																					
131	Phudugey	aus	O. sativa	131	_	_	_	_	9																																					
132	Rathuwee	indica	O. sativa	132	_	_	_	_	7																																					
133	Rikuto Kemochi	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
134	Romeo	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
136	RTS12	indica	O. sativa	_	_	_	_	_	4																																					
137	RTS14	indica	O. sativa	_	_	_	_	_	4																																					
138	RTS4	indica	O. sativa	_	_	_	_	_	4																																					
139	S4542A3-49B-2B12	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
140	Saturn	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
141	Seratoes Hari	indica	O. sativa	_	_	_	_	_	6																																					
142	Shai-Kuh	indica	O. sativa	_	_	_	_	_	6																																					
143	Shinriki	temperate-japonica	O. sativa	143	_	_	_	_	1																																					
144	Shoemed	temperate-japonica	O. sativa	144	_	_	_	_	5																																					
145	Short Grain	indica	O. sativa	_	_	_	_	_																																						
147	Sinampaga Selection	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
148	Sintane Diofor	indica	O. sativa	_	_	_	_	_	6																																					
149	Sinaguing	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
150	Sultani	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
151	Suweon	temperate-japonica	O. sativa	151	_	_	_	_	1																																					
152	T 1	aus	O. sativa	152	_	_	_	_	9																																					
153	T26	aus	O. sativa	_	_	_	_	_	9																																					

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
154	Ta Hung Ku	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
155	Ta Mao Tsao	temperate-japonica	O. sativa	_	_	_	_	_	5																																					
156	Taichung Native 1	indica	O. sativa	_	_	_	_	_	3																																				
157	Tainan Iku 487	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
158	Taipei 309	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
159	Tam Cau 9A	indica	O. sativa	_	_	_	_	_	6																																				
160	NSFTV160	aromatic	O. sativa	_	_	_	_	_	1																																				
161	TeQing	indica	O. sativa	161	_	_	_	_	4																																				
162	TKM6	indica	O. sativa	_	_	_	_	_	4																																				
163	Taducan	indica	O. sativa	_	_	_	_	_	4																																				
164	Tondok	tropical-japonica	O. sativa	_	_	_	_	_	5																																				
165	Trembese	tropical-japonica	O. sativa	165	_	_	_	_	3																																				
166	Tsipala 421	admixed-indica	O. sativa	_	_	_	_	_	4																																				
167	B6616A4-22-Bk-5-4	tropical-japonica	O. sativa	_	_	_	_	_	1																																				
169	WC 6	temperate-japonica	O. sativa	_	_	_	_	_	5																																				
170	Wells	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
171	ZHE 733	indica	O. sativa	_	_	_	_	_	4																																				
172	Zhenshan 2	indica	O. sativa	_	_	_	_	_	4																																				
173	Nipponbare	temperate-japonica	O. sativa	173	_	_	_	_	1																																				
174	Azucena	tropical-japonica	O. sativa	174	_	_	_	_	3																																				
175	NSFTV175	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
176	583	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
177	68-2	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
178	ARC 6578	aus	O. sativa	_	_	_	_	_	9																																				
179	Bellardone	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
180	Benllok	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
181	Bergreis	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
182	Blue Rose Supreme	admixed-japonica	O. sativa	_	_	_	_	_	1																																				
183	Boa Vista	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
184	Bombon	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
185	British Honduras Creole	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
186	Bul Zo	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
187	C57-5043	tropical-japonica	O. sativa	_	_	_	_	_	1																																				

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
188	Coppocina	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
189	Criollo La Fria	indica	O. sativa	_	_	_	_	_	4																																					
190	Delrex	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
191	Dom Zard	aromatic	O. sativa	_	_	_	_	_	1																																					
192	Erythroceros Hokkaido	temperate-japonica	O. sativa	_	_	_	_	_	5																																					
193	Fossa Av	tropical-japonica	O. sativa	_	_	_	_	_	5																																					
194	Hg 24	admixed	O. sativa	_	_	_	_	_	11																																					
195	IRAT 13	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
196	JM70	indica	O. sativa	_	_	_	_	_	8																																					
198	Leah	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
199	NSFTV199	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
200	P 737	aus	O. sativa	_	_	_	_	_	9																																					
201	Pate Blanc Mn 1	tropical-japonica	O. sativa	_	_	_	_	_	5																																					
202	Pratao	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
203	Radin Ebos 33	indica	O. sativa	_	_	_	_	_	4																																					
204	Razza 77	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
205	Rinaldo Bersani	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
206	Rojofotsy 738	admixed-indica	O. sativa	_	_	_	_	_	4																																					
207	Sigadis	indica	O. sativa	_	_	_	_	_	6																																					
208	SLO 17	indica	O. sativa	_	_	_	_	_	4																																					
209	Tchibanga	indica	O. sativa	_	_	_	_	_	3																																					
211	Tokyo Shino Mochi	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
212	NSFTV212	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
213	WC 3397	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
214	WC 4419	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
215	WC 4443	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
216	Yabani Montakhab 7	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
217	YRL-1	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
218	PI 298967-1	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
219	Nucleoryza	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
220	Azerbaidjanica	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
221	Sadri Belyi	aromatic	O. sativa	_	_	_	_	_	1																																					
222	Paraiba Chines Nova	indica	O. sativa	_	_	_	_	_	4																																					

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
223	Priano Guaira	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
224	Karabaschak	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
225	Biser 1	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
226	IRAT 44	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
227	Riz Local	admixed-indica	O. sativa	_	_	_	_	_	8																																					
228	CA 902/B/2/1	aus	O. sativa	_	_	_	_	_	9																																					
230	NSFTV230	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
231	Hunan Early Dwarf No. 3	indica	O. sativa	_	_	_	_	_	3																																					
232	Shangyu 394	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
233	Sung Liao 2	temperate-japonica	O. sativa	_	_	_	_	_	5																																					
234	Aijiaonante	indica	O. sativa	_	_	_	_	_	4																																				
235	Sze Guen Zim	indica	O. sativa	_	_	_	_	_	4																																				
236	WC 521	admixed-japonica	O. sativa	_	_	_	_	_	5																																				
237	Estrela	admixed-japonica	O. sativa	_	_	_	_	_	3																																				
238	WAB 56-104	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
239	WAB 502-13-4-1	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
240	WAB 501-11-5-1	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
241	ECIA76-S89-1	indica	O. sativa	_	_	_	_	_	4																																				
242	27	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
243	Tropical Rice	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
244	Arabi	admixed-japonica	O. sativa	_	_	_	_	_	3																																				
245	Sab Ini	temperate-japonica	O. sativa	_	_	_	_	_	3																																				
246	Saraya	aus	O. sativa	_	_	_	_	_	6																																				
247	Desvauxii	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
248	Caucasica	temperate-japonica	O. sativa	_	_	_	_	_	1																																				
249	Pirinae 69	temperate-japonica	O. sativa	_	_	_	_	_	3																																				
250	Bulgare	temperate-japonica	O. sativa	_	_	_	_	_	5																																				
251	H256-76-1-1-1	tropical-japonica	O. sativa	_	_	_	_	_	3																																				
252	Djimoron	indica	O. sativa	_	_	_	_	_	6																																				
253	Guineandao	admixed-japonica	O. sativa	_	_	_	_	_	3																																				
255	Pai Hok Glutinous	indica	O. sativa	_	_	_	_	_	9																																				
256	Romanica	temperate-japonica	O. sativa	_	_	_	_	_	3																																				
257	Agusita	temperate-japonica	O. sativa	_	_	_	_	_	1																																				

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
258	Tia Bura	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
259	Sadri Tor Misri	admixed-indica	O. sativa	_	_	_	_	_	8																																							
260	NSFTV260	aromatic	O. sativa	_	_	_	_	_	1																																							
261	Shim Balte	aus	O. sativa	_	_	_	_	_	8																																							
262	Halwa Gose Red	aus	O. sativa	_	_	_	_	_	9																																							
263	Maratelli	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
264	Baldo	admixed-japonica	O. sativa	_	_	_	_	_	1																																							
265	Vialone	admixed-japonica	O. sativa	_	_	_	_	_	3																																							
266	Hiderisirazu	admixed-japonica	O. sativa	_	_	_	_	_	1																																							
267	Hatsunishiki	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
268	Vavilovi	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
269	Sundensis	indica	O. sativa	_	_	_	_	_	6																																							
270	Osogovka	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
271	M. Blatec	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
272	923	admixed	O. sativa	_	_	_	_	_	3																																							
273	Varyla	tropical-japonica	O. sativa	_	_	_	_	_	1																																							
274	Padi Pagalong	tropical-japonica	O. sativa	_	_	_	_	_	3																																							
275	Sri Malaysia Dua	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
276	Kaukau	aus	O. sativa	_	_	_	_	_	6																																							
277	Gambiaka Sebela	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
278	C1-6-5-3	admixed	O. sativa	_	_	_	_	_	1																																							
279	Kon Suito	admixed	O. sativa	_	_	_	_	_	1																																							
280	Saku	tropical-japonica	O. sativa	_	_	_	_	_	5																																							
281	Patna	admixed-japonica	O. sativa	_	_	_	_	_	3																																							
282	Triomphe Du Maroc	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
283	Chibica	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
284	IR-44595	indica	O. sativa	_	_	_	_	_	4																																							
285	Tox 782-20-1	tropical-japonica	O. sativa	_	_	_	_	_	5																																							
286	IITA 135	tropical-japonica	O. sativa	_	_	_	_	_	3																																							
287	Zerawchanica Karatalski	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
288	Italica Carolina	temperate-japonica	O. sativa	_	_	_	_	_	3																																							
289	Lusitano	temperate-japonica	O. sativa	_	_	_	_	_	1																																							
290	Amposta	temperate-japonica	O. sativa	_	_	_	_	_	1																																							

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
291	Toploea 70/76	temperate-japonica	O. sativa	_	_	_	_	_	5																																					
292	Stegaru 65	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
293	TOg 7178	admixed-indica	O. sativa	_	_	_	_	_	8																																					
294	SL 22-613	indica	O. sativa	_	_	_	_	_	8																																					
295	Bombilla	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
296	Dosel	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
297	Bahia	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
299	SML 242	indica	O. sativa	_	_	_	_	_	4																																					
300	Sml Kapuri	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
301	Melanotrix	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
302	WIR 3039	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
303	Kihogo	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
304	519	indica	O. sativa	_	_	_	_	_	9																																					
305	›ble Carolina Rinaldo Barsa	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
306	WIR 3764	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
307	Uzbekskij 2	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
308	Llanero 501	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
309	Manzano	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
310	R 101	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
311	56-122-23	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
313	BR24	indica	O. sativa	_	_	_	_	_	4																																					
314	CTG 1516	aus	O. sativa	_	_	_	_	_	9																																					
315	Dawebyan	indica	O. sativa	_	_	_	_	_	8																																					
316	DD 62	aus	O. sativa	_	_	_	_	_	9																																					
317	DJ 123	aus	O. sativa	_	_	_	_	_	9																																					
318	DJ 24	aus	O. sativa	_	_	_	_	_	9																																					
319	DK 12	aus	O. sativa	_	_	_	_	_	9																																					
320	DM 43	aus	O. sativa	_	_	_	_	_	9																																					
321	DM 56	aus	O. sativa	_	_	_	_	_	9																																					
322	DM 59	aus	O. sativa	_	_	_	_	_	9																																					
323	DNJ 140	aus	O. sativa	_	_	_	_	_	9																																					
324	DV 123	aus	O. sativa	_	_	_	_	_	9																																					
325	EMATA A 16-34	indica	O. sativa	_	_	_	_	_	3																																					

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
326	Ghorbhai	aus	O. sativa	_	_	_	_	_	6																																					
327	Goria	aus	O. sativa	_	_	_	_	_	9																																						
328	Jamir	aus	O. sativa	_	_	_	_	_	9																																					
329	Kachilon	aus	O. sativa	_	_	_	_	_	9																																					
330	Khao Pahk Maw	aus	O. sativa	_	_	_	_	_	2																																					
331	Khao Tot Long 227	aus	O. sativa	_	_	_	_	_	9																																					
332	KPF-16	admixed-indica	O. sativa	_	_	_	_	_	9																																					
333	Leuang Hawn	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
334	Lomello	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
335	Okshitmayin	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
336	Paung Malaung	aus	O. sativa	_	_	_	_	_	9																																					
337	Sabharaj	indica	O. sativa	_	_	_	_	_	6																																					
338	Sitpwa	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
339	Yodanya	indica	O. sativa	_	_	_	_	_	4																																					
340	Berenj	admixed	O. sativa	_	_	_	_	_	9																																					
341	Shirkati	aus	O. sativa	_	_	_	_	_	9																																					
342	Cenit	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
343	Victoria F.A.	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
344	Habiganj Boro 6	admixed	O. sativa	_	_	_	_	_	1																																					
345	DZ 193	aus	O. sativa	_	_	_	_	_	9																																					
346	Karkati 87	aus	O. sativa	_	_	_	_	_	9																																					
347	Creole	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
348	China 1039	indica	O. sativa	_	_	_	_	_	4																																					
349	Chang Ch'Sang Hsu Tao	indica	O. sativa	_	_	_	_	_	4																																					
350	Ligerito	tropical-japonica	O. sativa	_	_	_	_	_	5																																					
351	NSFTV351	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
352	Guatemala 1021	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
353	ARC 10376	aus	O. sativa	_	_	_	_	_	9																																					
354	BALA	indica	O. sativa	_	_	_	_	_	4																																					
355	ASD 1	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
356	JC 117	indica	O. sativa	_	_	_	_	_	4																																					
357	9524	aus	O. sativa	_	_	_	_	_	9																																					
358	ARC 10086	tropical-japonica	O. sativa	_	_	_	_	_	1																																					

NSE_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
359	Surjamkuhi	aus	O. sativa	_	_	_	_	_	9																																					
360	PTB 30	aus	O. sativa	_	_	_	_	_	9																																					
362	NSFTV362	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
363	Edomen Scented	temperate-japonica	O. sativa	_	_	_	_	_	5																																					
364	Rikuto Norin 21	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
365	Shirogane	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
366	Kiuki No. 46	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
367	Sanbyang-Daeme	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
368	Deokjeokjodo	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
369	Sathi	aus	O. sativa	_	_	_	_	_	8																																					
370	Coarse	aus	O. sativa	_	_	_	_	_	9																																					
371	Santhi Sufaid	aus	O. sativa	_	_	_	_	_	9																																					
372	Sufaid	aus	O. sativa	_	_	_	_	_	9																																					
373	Lambayeque 1	aromatic	O. sativa	_	_	_	_	_	1																																					
374	NSFTV374	temperate-japonica	O. sativa	_	_	_	_	_	1																																					
375	Upland	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
376	Breviaristata	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
377	PR 304	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
378	Kalubala Vee	aus	O. sativa	_	_	_	_	_	2																																					
379	Wanica	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
380	Tainan-Iku No. 512	temperate-japonica	O. sativa	_	_	_	_	_	6																																					
381	325	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
383	NSFTV383	temperate-japonica	O. sativa	_	_	_	_	_	3																																					
384	318	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
385	Nira	indica	O. sativa	_	_	_	_	_	4																																					
386	Palmyra	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
387	M-202	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
388	Nortai	admixed-japonica	O. sativa	_	_	_	_	_	1																																					
389	CI 11011	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
390	CI 11026	admixed	O. sativa	_	_	_	_	_	4																																					
391	Della	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
392	Edith	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
394	Lady Wright Seln	tropical-japonica	O. sativa	_	_	_	_	_	1																																					

NSF_ID	Variety_name	Subpopulation 1, 2	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
395	OS 6 (WC 10296)	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
396	Cocodrie	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
397	Cybonnet	tropical-japonica	O. sativa	397	_	_	_	_	3																																					
398	93-11	indica	O. sativa	398	_	_	_	_	4																																					
399	Spring	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
400	Yang Dao 6	indica	O. sativa	_	_	_	_	_	4																																					
401	IRGC80433	*Admix W4/ W1*	*O. rufipogon*	401	1	_	8	_	11																																					
402	IRGC80539	*W4*	*O. spontanea*	402	1	_	8	_	8																																					
407	407_C2_S1	*W1*	*O. rufipogon*	407	1	_	2.67	_	11																																					
410	IRGC80759	*Admix W1/ W5*	*O. nivara*	410	1	_	8	_	11																																					
413	IRGC81850	*W5*	*O. nivara*	413	1	_	8	_	12																																					
415	IRGC81909	*W5*	*O. spontanea*	415	1	WT	8	WT	12																																					
416	IRGC81970	*W1*	*O. spontanea*	416	0	14bp del	2.33	22bp del	6																																					
431	IRGC82992	*Admix W6/ W2*	*O. rufipogon*	431	1	_	7	_	11																																					
433	IRGC83795	*W1*	*O. rufipogon*	433	0	14bp del	2.25	22bp del	4																																					
435	IRGC86448	*W1*	*O. rufipogon*	435	0	_	7	_	6																																					
438	438_B2_1_S2	*Admix W1/ W6*	*O. rufipogon*	438	0.7777778	_	6.67	_	11																																					
442	IRGC93181	*W5*	*O. nivara*	442	1	_	8	_	12																																					
443	IRGC93183	*W5*	*O. nivara*	443	1	_	8	_	12																																					
444	IRGC93188	*W5*	*O. nivara*	444	1	WT	8	WT	12																																					
445	IRGC93189	*W5*	*O. nivara*	445	1	_	8	_	12																																					
446	IRGC93224	*W5*	*O. spontanea*	446	1	_	8	_	12																																					
449	449_A1_2_S2	*Admix W1/ W5*	*O. nivara*	449	1	_	6.67	_	11																																					
450	IRGC100916	*W2*	*O. nivara*	450	1	_	8	_	11																																					
451	IRGC101508	*Admix W5/ W4*	*O. nivara*	451	1	WT	2	22bp del	11																																					
453	IRGC103404	*Admix W1/ W2*	*O. rufipogon*	453	0.2222222	_	5.5	_	6																																					
454	IRGC103821	*W6*	*O. nivara*	454	0.3333333	_	8	_	3																																					
457	457_B3_1_S2	*Admix W6/ W1*	*O. nivara*	457	1	_	4.33	_	11																																					
461	461_A1_1_S2	*W6*	*O. rufipogon*	461	0.5555556	_	6.89	_	4																																					
465	IRGC104620	*W6*	*O. spontanea*	465	1	WT	8	WT	11																																					
467	IRGC104624	*W6*	*O. rufipogon*	467	1	_	7.56	_	11																																					
472	IRGC104636	*W6*	*O. spontanea*	472	0.6666667	_	8	_	10																																					
477	IRGC104967	*Admix W1/ W4*	*O. spontanea*	477	1	WT	2	22bp del	11																																					

NSE_ID	Variety_name	Subpopulation [1, 2]	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
481	IRGC105343	*W2*	*O. nivara*	481	1	_	8	_	11																																					
482	IRGC105349	*W2*	*O. rufipogon*	482	1	_	8	_	11																																					
483	483_C2_1_S2	*W2*	*O. rufipogon*	483	1	_	7.56	_	11																																					
487	487_C2_S2	*W2*	*O. nivara*	487	1	_	8	_	11																																					
488	IRGC105491	*W4*	*O. rufipogon*	488	1	_	8	_	11																																					
493	IRGC105706	*W5*	*O. nivara*	493	1	_	8	_	11																																					
494	IRGC105711	*W2*	*O. rufipogon*	494	1	_	8	_	11																																					
495	IRGC105717	*Admix W4/ W1*	*O. nivara*	495	1	_	8	_	11																																					
496	IRGC105720	*W1*	*O. rufipogon*	496	1	WT	8	WT? het?	11																																					
498	IRGC105735	*Admix W1/ W4*	*O. rufipogon*	498	1	_	8	_	11																																					
499	IRGC105767	*W1*	*O. rufipogon*	499	1	_	8	_	11																																					
501	IRGC105821	*Admix W4/ W1*	*O. nivara*	501	1	_	8	_	11																																					
505	505_A1_2_S2	*W1*	*O. rufipogon*	505	1	_	4.67	_	11																																					
508	IRGC105890	*Admix W1/ W6*	*O. rufipogon*	508	0.8333333	_	3	_	2																																					
509	IRGC105897	*W1*	*O. rufipogon*	509	0.3333333	_	6.33	_	6																																					
514	IRGC105956	*W1*	*O. rufipogon*	514	1	_	8	_	11																																					
523	523_A1_S2	*Admix W4/ W1*	*O. nivara*	523	1	_	7.33	_	11																																					
549	IRGC81881	*Admix W1/ W6*	*O. rufipogon*	549	1	_	8	_	11																																					
553	IRGC100926	*W1*	*O. rufipogon*	553	0	14bp del	7	22bp del	4																																					
555	555_B1_1_S2	*W2*	*O. rufipogon*	555	1	WT	8	WT	11																																					
568	IRGC106263	*Admix W1/ W4*	*O. rufipogon*	568	1	WT	8	WT	11																																					
600	IRGC100187	*W1*	*O. sat. x O. ruf.*	600	1	_	2.67	_	11																																					
602	IRGC100900	*W2*	*O. nivara*	602	0.8333333	_	7	_	11																																					
605	IRGC100911	*W1*	*O. spontanea*	605	0.1666667	14bp del	3.33	22bp del	7																																					
612	IR64	indica	*O. sativa*	612	_	_	_	_	4																																					
616	RT0034	indica	O. sativa	_	_	_	_	_	4																																					
617	MCR010277	admixed	O. sativa	_	_	_	_	_	3																																					
618	Pecos	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
619	Rosemont	tropical-japonica	O. sativa	_	_	_	_	_	3																																					
620	Jasmine85	indica	O. sativa	_	_	_	_	_	4																																					
621	LaGrue	tropical-japonica	O. sativa	_	_	_	_	_	1																																					
622	Bengal	admixed-japonica	O. sativa	_	_	_	_	_	3																																					
624	Kaybonnet	tropical-japonica	O. sativa	_	_	_	_	_	3																																					

NSE_ID	Variety_name	Subpopulation [1, 2]	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
625	Katy	tropical-japonica	O. sativa	_	_	_	_	_	3																																						
626	C101A51	indica	O. sativa	_	_	_	_	_	4																																						
627	Early	admixed-japonica	O. sativa	_	_	_	_	_	1																																						
628	Jefferson	tropical-japonica	O. sativa	628	_	_	_	_	3																																						
629	Panda	admixed-japonica	O. sativa	_	_	_	_	_	3																																						
630	Saber	tropical-japonica	O. sativa	_	_	_	_	_	3																																						
631	Dragon Eyeball 100	admixed-japonica	O. sativa	_	_	_	_	_	1																																						
632	Francis	tropical-japonica	O. sativa	_	_	_	_	_	3																																						
633	Jing 185-7	indica	O. sativa	_	_	_	_	_	4																																						
634	Rondo (4484-1693)	indica	O. sativa	_	_	_	_	_	4																																						
635	Azucena	tropical-japonica	O. sativa	_	_	_	_	_	3																																						
636	Sadu Cho	indica	O. sativa	_	_	_	_	_	6																																						
638	Moroberekan	tropical-japonica	O. sativa	_	_	_	_	_	5																																						
639	Nipponbare	temperate-japonica	O. sativa	_	_	_	_	_	1																																						
641	Tainung 67	temperate-japonica	O. sativa	_	_	_	_	_	1																																						
642	Zhenshan 97B	indica	O. sativa	_	_	_	_	_	4																																						
643	Minghui 63	indica	O. sativa	_	_	_	_	_	4																																						
644	IR64-21	indica	O. sativa	_	_	_	_	_	4																																						
646	Swarna	indica	O. sativa	_	_	_	_	_	6																																						
647	Cypress	tropical-japonica	O. sativa	_	_	_	_	_	3																																						
651	Dular	aus	O. sativa	_	_	_	_	_	9																																						
652	Li-Jiang-Xin-Tuan-Hei-Gu	temperate-japonica	O. sativa	_	_	_	_	_	1																																						
655	NSFTV655	*W2*	*O. nivara*	655	1	_	8	_	11																																						
657	NSFTV657	*W2*	*O. nivara*	657	1	_	8	_	11																																						
658	NSFTV658	*W2*	*O. niv. x O. ruf.*	658	1	_	8	_	11																																						
665	IRGC100203	*Admix W1/ W4*	*O. ruf. x O. sat.*	665	1	WT	2.33	22bp del	11																																						
666	IRGC100211	*W2*	*O. rufipogon*	666	1	_	6.33	_	11																																						
669	669_C2_3_S2	*W6*	*O. nivara*	669	0	14bp del	8	WT	6																																						
673	IRGC100647	*W6*	*O. rufipogon*	673	1	_	3.44	_	6																																						
676	676_A1_1_S2	*W6*	*O. rufipogon*	676	0	14bp del	3.44	22bp del	1																																						
682	IRGC100904	*W1*	*O. rufipogon*	682	1	_	8	_	11																																						
683	IRGC100918	*Admix W4/ W1*	*O. nivara*	683	0.7777778	_	6.67	_	11																																						
685	IRGC100923	*Admix W4/ W1*	*O. rufipogon*	685	0.2222222	14bp del	3.11	WT	4																																						

NSE_ID	Variety_name	Subpopulation [1, 2]	Species	NSF ID_Kim et al, 2016	Pericarp Color Score_Mean	Rc_14bp_InDel	Hull_color_score_Mean	BH4_22bp_InDel	Rc_HapGroup	SNP-7.5656907.	SNP-7.5672330.	SNP-7.5687101.	SNP-7.5688724.	SNP-7.5710815.	SNP-7.5713712.	SNP-7.5714910.	SNP-7.5735405.	SNP-7.5757763.	SNP-7.5819582.	SNP-7.5830906.	SNP-7.5872167.	SNP-7.5890466.	SNP-7.5921296.	SNP-7.5943903.	SNP-7.5956619.	SNP-7.5984835.	SNP-7.5989764.	SNP-7.6022020.	SNP-7.6066391.	SNP-7.6066523.	SNP-7.6089138.	SNP-7.6097581.	SNP-7.6098396.	SNP-7.6115607.	SNP-7.6116546.	SNP-7.6118547.	SNP-7.6125529.	SNP-7.6139845.	SNP-7.6144231.	SNP-7.6159543.	SNP-7.6195493.	SNP-7.6197216.	SNP-7.6198249.	SNP-7.6217264.	SNP-7.6220628.	SNP-7.6227393.	SNP-7.6228434.	SNP-7.6229668.	SNP-7.6232974.
687	IRGC101193	*Admix W6/ W1*	*O. rufipogon*	687	1	_	5.33	_	10																																					
691	IRGC101967	*W4*	*O. nivara*	691	0.8333333	_	6.67	_	8																																					
701	IRGC103813	*Admix W6/ W5*	*O. niv. x O. ruf.*	701	0	14bp del	2.89	22bp del	3																																					
704	IRGC103818	*Admix W6/ W1*	*O. ruf. x O. niv.*	704	0.2222222	14bp del	3.78	WT	3																																					
707	IRGC103835	*W4*	*O. nivara*	707	1	_	8	_	8																																					
708	708_A1_2_S2	*Admix W4/ W1*	*O. nivara*	708	1	_	6.33	_	11																																					
711	IRGC103841	*W4*	*O. nivara*	711	1	WT	8	WT	8																																					
715	715_B2_1_S2	*Admix W4/ W1*	*O. ruf. x O. niv.*	715	1	_	8	_	11																																					
716	716_B2_S2	*W4*	*O. rufipogon*	716	1	_	8	_	11																																					
717	717_B1_2_S2	*W4*	*O. nivara*	717	1	_	8	_	11																																					
719	719_A1	*W5*	*O. nivara*	719	1	WT	8	WT	12																																					
720	IRGC104703	*W5*	*O. nivara*	720	1	_	8	_	12																																					
736	736_B2_1_S2	*W4*	*O. rufipogon*	736	0.6666667	_	8	_	11																																					
738	IRGC105601	*W4*	*O. ruf. x O. niv.*	738	1	_	8	_	11																																					
743	743_C1_2_S2T	*W5*	*O. nivara*	743	1	_	8	_	12																																					
746	IRGC105740	*Admix W4/ W1*	*O. nivara*	746	0.3333333	_	8	_	11																																					
751	IRGC105895	*W6*	*O. nivara*	751	1	WT	2	22bp del	7																																					
759	759_A1_3_S2	*W1*	*O. rufipogon*	759	0.2222222	14bp del	8	WT	7																																					
760	IRGC106345	*Admix W1/ W5*	*O. nivara*	760	1	_	8	_	11																																					
765	W1943	_	*O. rufipogon*	_	_	_	_	_	11																																					
766	W1944	_	*O. rufipogon*	_	_	_	_	_	10																																					
767	W1945	_	*O. rufipogon*	_	_	_	_	_	11																																					
770	Mojito Colorado	tropical-japonica	*O. sativa*	_	_	_	_	_	3																																					

1 *O. sativa* in RDP1, McCouch et al., 2016

2 ORSC species names cross-referenced to Table S1

3 Cross-referenced to Table S1

Table S9. Chloroplast sequence

B. Chloroplast sequence

Loci(bp)[1]	Gene ID	Primer_Forward	Primer_Reverse	Reference
7,888-8,804	ORF100	gccgctttagtccactcagccatc	tcaatgccttttttcaatggtctc	Kanno et al. (1993)[2]
12,013-12,913	psbZ	tatttgcttctcctgatggttggt	gagcggagtagagcagtttggtag	Takahashi et al. (2008)[3]
56,210-57,987	ORF133, 106, 36, and 185	1) agaatctggacccatcgt 2) gagatcggaaaagaaa	ttactatttctatctattcgattt cgaatcggtcataaccac	Kawakami et al. (2007)[4]
76,604-76,773	rps8	aaaacgttgtatttttgttt	ttctcgaggtataatgacag	Kawakami et al. (2007)
77,584-78,118	rpl14 and 16	atctgcagcatttaaaagggtctgaggttgatcat	aaagatctagatttcgtaaacaacatagaggaagaa	Nakamura et al. (1998)[5]

1 Sequence position was assigned based on Genbank accessions NC_001320

2 Kanno, A., N. Watanabe, I. Nakamura, and A. Hirai, 1993 Variations in chloroplast DNA from rice (Oryza sativa): differences between deletions mediated by short direct-repeat sequences within a single species. TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik 86: 579–84.

3 Takahashi, H., Y. Sato, and I. Nakamura, 2008 Evolutionary analysis of two plastid DNA sequences in cultivated and wild species of Oryza. Breeding Science 58: 225–233.

4 Kawakami, S., K. Ebana, T. Nishikawa, Y. Sato, D. A. Vaughan *et al.*, 2007 Genetic variation in the chloroplast genome suggests multiple domestication of cultivated Asian rice (Oryza sativa L.). Genome 50: 180–7.

5 Nakamura, I., H. Urairong, N. Kameya, Y. Fukuta, S. Chitrakon *et al.*, 1998 Six different plastid subtypes were found in 0. sativa-0.rufipogon comp. Rice Genetics Newsletter 15: 80–82.

APPENDIX B: OBSERVATIONS ON THE RDP1 WILD PANEL

The following information consist of personal observations, records, and notes

Trends in self pollinated seed production

Obligate outcrossing individuals become more self-sterile and produce fewer seed the more they are forced to be inbred/self-pollinated.

Changes in time of year – cloning in spring and growing out for flowering in early summer may increase panicle/seed production.

Stolon vs. tiller definition

The following are a list of observations and personal decisions on tiller vs. stolon traits and definitions from October 8, 2010. I took it upon myself to define stolons vs. tillers on my *ORSC* panel accessions as part of understanding and evaluating their life and reproductive habits, as well as to disaggregate the compound trait of "plant type" – a combination of stolon absence/presence and tiller angle.

While stolons can be generally defined as a "creeping horizontal stem or runner with the ability or tendency to form new roots or stems at the nodes," grass stolons in general, and that of the *ORSC* in particular have no standard accepted definition. After an email conversation with Elisabeth Kellogg on stolon definition, she confirmed the prior and suggested that I define the particular traits associated with stolons and keep those consistent for the course of my research.

There is a difference between a stolon and a just-flopped-over tiller, though lodged tillers may or may not develop into stolons depending on the genetics of the plant and possibily also environmental cues such as partial submergence in water.

Stolons must have these characteristics: be horizontal/near horizontal, have lateral meristem outgrowth*, have elbowed panicles or upright apical meristem growth

Designation	**Elbow**	**Stem horizontal to main plant axis**	**Vegetative lateral meristem**	**Lateral meristem with panicle**	**Long internode/Exposed node**	**Branches from near bottom**
Stolon	yes	yes	Yes	yes	yes	yes
Stolon	Slight/none	yes	Yes	Not necessarily	yes	Not necessarily
Stolon?	Not necessarily	<90	Yes	Not necessarily	Almost always	No
Tiller	No/slight	No - vertical	No	May or not	May or not	No

TRAIT	Always a tiller?	Always a stolon?
Elbow (kink at node)	Not necessarily	Not necessarily
Horizontal	Not necessarily	Not necessarily, but stolons must be horizontal or near-horizontal
Vegetative lateral meristem	Never	Yes
Reproductive lateral meristem (with panicle)	Ambiguous – only if vert and 2nd P 1 IN down from 1st P	Ambiguous
Long internode/Exposed node	Ambiguous	Ambiguous
Branches from near bottom	Never	Yes
Roots from submerged node	Undecided, but must have exposed node to be rooting	Yes?

*If long enough, stolons do not need to have a lateral meristem to be considered a stolon.

Tillers, however, may have elongated IN and be curved, or even have an elbow, but if relatively upright and close to plant center, are not considered a stolon.