UNLOCKING THE GENETIC DIVERSITY OF THE UNDOMESTICATED RICE RELATIVE *ORYZA LONGISTAMINATA*

BY

MARLEE R. LABROO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Crop Sciences
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Master's Committee:

Associate Professor Erik J. Sacks, Chair
Professor Stephen P. Moose
Assistant Professor Alexander E. Lipka

ABSTRACT

Rice (*Oryza sativa*) is a crucial part of the global food supply. Meeting projected increases in demand, which are rising fastest in sub-Saharan Africa, while using resources efficiently and adapting to a changing climate will require new sources of variation for trait improvement. The genetically diverse undomesticated African rice, *Oryza longistaminata*, which is native to a broad range of habitats throughout sub-Saharan Africa, is a valuable resource improving *O. sativa* with genes conferring biotic and abiotic stress tolerances, increased yield, super-ratooning ability, and floral traits conducive to hybrid rice production. The major obstacle to utilization of *O. longistaminata* in breeding programs is a breeding barrier which complicates interspecific hybridization with *O. sativa*, as well as the expense and difficulty of identifying which *O. longistaminata* accessions carrying useful genes. Breeding programs also rely on thorough *ex-situ* conservation of *O. longistaminata*, which enables long-term access to germplasm for breeders and other stakeholders. In this work, three genetic subpopulations were identified in *O. longistaminata* using individuals representing most of the species' range genotyped with densely-spaced, genome-wide molecular markers. Spontaneous interspecific hybrids of *O. longistaminata* and *O. sativa* were discovered in germplasm from the International Rice Research Institute's genebank, which can be used to accelerate introgression from *O. longistaminata* to *O. sativa*. A species distribution model of *O. longistaminata* was generated to identify sampling gaps in *ex-situ* collections, and environmental data was used to predict likely locations of *O. longistaminata* tolerant to abiotic stresses.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Erik J. Sacks, Dr. Lindsay V. Clark, and Shilai Zhang for their significant contributions—intellectual and otherwise—to the creation of this work and to convey my deep appreciation for their collaboration. I would like to further extend my thanks to Nirmal Sharma, Hongxu Dong, Sandeep Sakhale, Colten Maertens, Homayoun Watan, Shilpa Manjunatha, Sara Rothe, Brandi Smith, Michael Douglass, Charlie Burroughs, Dr. Patrick J. Brown, James Whitacre, Getachew Melaku, Chandra Paul, Dr. Joie Ramos, Dr. Duncan A. Vaughan, Dr. Ruaraidh Sackville-Hamilton and the IRRI Genebank team, and the Plant Care Facility staff, particularly Clint Shipley, Andy D'Urso, Michael Maddock, and Nathan Deppe. Graduate committee members Dr. Alexander E. Lipka and Dr. Stephen P. Moose and graduate program coordinator Dr. Jack Juvik also have my warm thanks, as does the Crop Science department and staff, especially Dr. Germán Bollero, Dr. Anthony J. Studer, Dianne Carson, Patricia Sarver, Tracey Malkovich, Anna Tammen, Chinedu Onyejekwe, Jennie Black, and Jordyn Conerty. I also express my gratitude for my graduate colleagues, in particular Esperanza Shenstone, Crystal Sorgini, Liana Acevedo-Siaca, Lenie Quiatchon-Baeza, and Dr. Allie Gardner. It has been a pleasure to work at the University of Illinois.

TABLE OF CONTENTS

CHAPTER 1: *ORYZA LONGISTAMINATA*, A PROMISING GENETIC RESOURCE FOR IMPROVEMENT OF THE DOMESTICATED RICE *O. SATIVA*

Introduction

Asian rice (*Oryza sativa*) currently supplies one-fifth of global per capita calories (Awika, 2011). Global consumption is projected to rise from 439 million tons to 555 million tons between 2010 and 2035 as population increases (Seck et al., 2012), and demand is rising fastest in Sub-Saharan Africa, which currently imports 42% of its rice (Nasrin et al., 2015). Meeting rice production targets with fewer inputs under the additional challenges posed by climate change will require new sources of genes for trait improvement (Mohanty et al., 2013). Rice wild relatives contain genes that are absent from domesticated rice, and thus have great potential as a source of valuable genetic material (Atwell et al., 2014; Kovach & McCouch, 2008). Due to genetic bottlenecks during domestication, *O. sativa* is 80-90% less diverse than its most recent wild ancestor, *O. rufipogon*; by comparison, maize (*Zea mays*) is only 40% less diverse than its wild ancestor *Z. perennis* (Zhu et al., 2007). There has been considerable success introgressing useful traits to domesticated Asian rice from the other 23 species in its genus (Vaughan, 1994). These introgressions include agronomic traits such as yield and cytoplasmic male sterility as well as resistance to abiotic and biotic stresses—some of which are not known to be present in *O. sativa* (Brar & Khush, 1997).

Of the *Oryza* species, eight share the AA genome and form the primary germplasm pool of *O. sativa*, while the others comprise nine distinct genomes that vary in similarity to the AA genome and ploidy (Vaughan, 1994). Phylogenetic analysis suggests that one of the most genetically distinct members of the AA-genome clade relative to *O. sativa* is the undomesticated African species *O. longistaminata* (Wambugu et al., 2015; Zhu et al., 2014). Two other AA-genome species are native to Africa: *O. glaberrima*, African domesticated rice, and its undomesticated progenitor *O. barthii* (Vaughan, 1994). *O. glaberrima*, which was domesticated later than the Asian species *O. sativa* and is predominantly cultivated in West Africa, was used to introgress genes conferring abiotic stress tolerances to *O. sativa* during the development of NERICA cultivars (New Rice for Africa; Sarla & Swamy, 2005). *O. longistaminata* is dispersed through a broad range of environments throughout much of sub-Saharan Africa, and it is expected to have high genetic diversity not only because of its large geographic range but also because it is an obligate outcrosser due to self-incompatibility. Further study of this underutilized wild rice could reveal numerous genes useful for increasing yields, reducing production costs, and increasing biotic and abiotic stress tolerances in domesticated rice.

Morphology

O. longistaminata typically exceeds two meters (IRRI, unpublished raw data) and is unique among the AA-genome species in having numerous long and spreading rhizomes, which confer perenniality; other perennial species in the clade persist vegetatively via tillers and/or stolons (Vaughan, 1994). Via the rhizomes, *O. longistaminata* spreads asexually and often forms dense stands in the wild, which may be composed of single or multiple genotypes (Kiambi et al., 2005). The leaves are green and pubescent, but the basal leaf sheath, nodes, and culms may be purple or green (Causse & Ghesquiere, 1991; IRRI, unpublished raw data).

O. longistaminata also reproduces sexually. In contrast to *O. sativa*, which self-pollinates all but 1-3% of the time on average (Bah et al., 2017), *O. longistaminata* is predominantly self-incompatible (Ghesquiere, 1986). An early survey by Ghesquiere (1986) found that 75% of observed accessions had self-fertilization rates less than 1%; though as many as 14% of the accessions were self-compatible and all were from areas where domesticated rice was cultivated and may not have represented pure *O. longistaminata*. However, Chu & Oka (1969) found that though three *O. longistaminata* accessions observed produced more seeds if cross-pollinated, fertility ranged from 14-36% if self-pollinated. Self-fertilized embryos and endosperms appeared normal, but pollen tube growth was retarded during self-pollination compared to cross-pollination, and pollen tubes usually failed to enter the style or ovule during self-pollination (Chu & Oka, 1969). However, self-pollinated seeds germinated and grew normally (Chu & Oka, 1969). In nature, Ghesquiere (1986) and Bezançon et al. (1977) suggested that there may be variation for self-compatibility among *O. longistaminata* genotypes, with isolated populations possibly more self-compatible than large intermixed populations. Though some populations as large as 2,500 m^2 have been reported to be completely sterile, this may be confounded with self-incompatibility in single-genotype stands (Kiambi et al., 2005).

O. longistaminata is typically photoperiod-sensitive and flowers during short days, although some individuals are photoperiod-insensitive (Ghesquiere, 1986; Melaku, personal communication, 2017; Vaughan, 1994). Its flowers are bisexual with long stamens (as the name indicates) and anthers averaging 6 mm, which produce large quantities of pollen (Causse & Ghesquiere, 1991; IRRI, unpublished raw data). The stigma is dark purple and exserted when mature (IRRI, unpublished raw data). The pistil length (as well as the stigma and style length individually) in *O. longistaminata* is significantly greater than for any other *Oryza* species (Marathi et al., 2015), which presumably facilitates allogamy. The flowers are arranged in a panicle that can extend more than half a meter beyond the flag

leaf sheath (IRRI, unpublished raw data). Seeds shatter at maturity, and timing of maturity can vary even in seeds of a single panicle (Kiambi et al., 2005; IRRI, unpublished raw data). The lemma and palea of the flower are pubescent; the lemma is awned (IRRI, unpublished raw data; Kiambi et al., 2005). The grain has a red or reddish-brown pericarp with a waxy endosperm (IRRI, unpublished raw data; Liu et al., 2009); it is longer than it is wide with an average length of ~9 mm and width of ~2 mm (IRRI, unpublished raw data; Katayama et al., 1987).

Ecology

O. longistaminata occupies a range of habitats throughout sub-Saharan Africa and Madagascar (Bezançon et al., 1978). It is common in permanently wet or seasonally flooded grasslands associated with rivers and ponds, and is often the dominant vegetation, forming stands sometimes interspersed with other perennial grasses such as *Echinochloa* spp. (Bezançon et al., 1978; Dörgeloh, 1999; Scholte, 2007; Scholte et al., 2000; Vaughan, 1994). *O. longistaminata* commonly grows with the undomesticated annual rice species *O. barthii* (Bezançon et al., 1978). Near the coasts *O. longistaminata* also grows in lagoons, which can be salty or acidic (Bezançon et al., 1978; Catarino et al., 2002). While Bezançon et al. (1978) report that *O. longistaminata* habitat is usually sunny, it can be found in forests (e.g. Cameroon) though this is likely less common. *O. longistaminata* commonly occurs along the borders of cultivated rice fields and may invade paddies as a weed, especially when cultivation ceases (Bezançon et al., 1978; Rodenburg & Johnson, 2009).

O. longistaminata has a well-studied relationship with inundation and water availability in seasonally flooded grasslands; it has been used as an indicator species to assess the impacts of dam construction and rehabilitation of affected floodplains (Kleynhans et al., 2007; Scholte et al., 2000). Kleynhans et al. (2007) reported that *O. longistaminata* is most suited to water depths from 0.1 to 0.5 meters, but stands can grow at 1-4 m inundation (Vaughan, 1994). *O. longistaminata* requires a period of around 25 days standing water or longer to flower and set seed, and flooding can facilitate spread of asexual propagules even though it is not required for vegetative spread (Green & El-Moghraby, 2009; Kleynhans et al., 2007). Sudden floods have been shown to decrease overall productivity of floodplains, but a study of controlled flood releases in the formerly dammed Logone floodplain found a positive correlation between *O. longistaminata* biomass and water depth up to 1 m (Scholte, 2007). Conversely, the species' cover greatly decreased with decreased water inputs due to damming (Scholte, 2005), and it is rarely found above the river floodline in unaltered systems (Scholte, 2007). Although water levels are not the only determinant of the species' occurrence, which also shows high unexplained temporal

variability, there is an apparent niche at water levels of 0.5-1 m in floodplains (Green & El-Moghraby, 2009; Higgins et al., 1996; Murray-Hudson et al., 2011; Scholte, 2007).

O. longistaminata experiences seasonal and multi-year variation in water availability (Higgins et al., 1996; Murray-Hudson et al., 2011; Scholte, 2007). For example, on the Nyl River of South Africa the wet season is followed by a five-month dry season from May to September; flooding occurs in only about 4 of 10 years with a typical inundation duration of 50 days (Higgins et al., 1996). In years without inundation the plants are thought to utilize stored resources in the rhizomes, which form dense underground layers. These resources are exhausted after approximately 3 years of drought (Kleynhans et al., 2007), and inability to regenerate from propagules can lead to local extinction (Higgins et al., 1996). During the dry season, aboveground *O. longistaminata* biomass dries. Recession of floods permits easy grazing and natural burning (Dörgeloh, 1999; Petersen & Fohrer, 2010).

O. longistaminata provides key ecosystem functions in floodplains and it forms associations with numerous animal species. It provides habitat for aquatic birds and supports high species diversity; both insectivorous and piscivorous birds use it for shelter and access to open water (Arbeiter & Tegetmeyer, 2011; Higgins et al, 1996). Indigenous people have noted that birds spread *O. longistaminata* seed (Gupta, 2004). It is also a source of fodder for large, nomadic ungulates such as roan antelope, waterbuck, and reedbuck (Arbeiter & Tegetmeyer, 2011; Dörgeloh, 1999; Higgins et al., 1996; Scholte et al., 2000).

In African agroecosystems, *O. longistaminata* is a noxious weed (Gupta, 2004; Johnson et al., 2004; Rodenburg & Johnson, 2009) that invades paddy fields and may reduce yields up to 85% or require abandoning the field entirely (Rodenburg & Johnson, 2009). *O. longistaminata* is also a host of rice yellow mottle virus and may act as a reservoir for the virus during the dry season when domesticated rice, the primary host, is not planted (Abo et al., 2000; Rodenburg & Johnson, 2009). It hosts insect pests that damage domesticated rice such as the African rice gall midge, *Orseolia oryzovora* (Williams et al., 1999), stemborers, and diopsid flies (Ba et al., 2008). The rhizomatous growth of *O. longistaminata* makes control extremely difficult. Mechanical management strategies include deep tillage during the dry season, which desiccates the rhizomes, or “underwater mowing” in which the developed plant stem is cut underwater, precluding regrowth (Rodenburg & Johnson, 2009). Though sometimes effective, these strategies are labor-intensive and require a growing season to take effect. Glyphosate may also be used to control *O. longistaminata* before emergence of domesticated rice seedlings (Rodenburg & Johnson, 2009). It is worth noting that although *O. longistaminata* shares the

red pericarp of weedy red rice, it is only known to affect the African continent currently (Goulart et al, 2014; Labrada, 2002). However, it is a potentially invasive weed of rice agroecosystems globally if introduced, and it also has the potential to facilitate transgene escape from domesticated rice should transgenic rice cultivars be deployed (Kanya, 2010).

O. longistaminata has a relationship with humans in its native range beyond its role as a weed and provision of ecosystem services (Gupta, 2004). It is collected for food by indigenous people, most notably the Bela and Fulani people of Mali. Interspecific hybrids between *O. sativa* and *O. longistaminata* are reportedly consumed in Malawi, albeit rarely (Kiambi et al., 2005). The vegetation of *O. longistaminata* is used as fodder for various livestock including cattle and goats (Gupta, 2004). It has ceremonial use as fencing and in constructing masks and can be used to make baskets (Gupta, 2002). Though *O. longistaminata* has been investigated for its potential to treat wastewater sludge in urban areas, it is not effective using current practices (Kouawa et al., 2015).

Phylogenetics and Genetic Diversity

There are about 23 species in the genus *Oryza*, depending on the classification used, which can be found on all continents except Europe and Antarctica (Vaughan et al., 2008). The genus contains 10 genome types which may be diploid or tetraploid with n=12: AA, BB, CC, BBCC, DD, CCDD, EE, FF, GG, and HHJJ (Vaughan et al., 2008). *O. longistaminata* is diploid with 2n=24 chromosomes and belongs to the AA-genome clade of *Oryza* (Vaughan, 1994). Species belonging to the AA-genome clade occur naturally in Asia, Africa, Oceania, and the Americas. In Africa, the AA-genome species *O. barthii, O. glaberrima,* and *O. sativa* grow sympatrically with *O. longistaminata* (Vaughan, 1994; Wambugu et al., 2015)*; O. barthii* and *O. glaberrima* are also native to Africa, whereas *O. sativa* was introduced from Asia. Three or four other *Oryza* species with other genome types are also native to Africa: *O. brachyantha* (FF genome), *O. eichingeri* (CC genome), and *O. punctata* (BB genome); the tetraploid form of *O. punctata* with a BBCC genome is sometimes classified as a separate species, *O. schweinfurthiana* (Vaughan, 1994; Wambugu et al., 2013). In older literature *O. longistaminata* was part of the *O. perennis* taxon, which included multiple perennial *Oryza* species as well as *O. barthii* (Oka, 1974). However, *O. longistaminata* was defined as a species by Clayton (1968) and this taxonomy has remained accepted to date.

The phylogenetic relationship and origin of the AA-genome clade has long been an area of research. Before the availability of genome-wide markers at high density and low cost, very few studies

had sufficient genomic breadth to confidently resolve the order of divergence within the clade (Wambugu et al., 2015; Zhu et al., 2014). Determining the history of this clade is also complicated by its rapid recent radiation—the AA-genome is thought to have diverged 0.69-9 million years ago (MYA)—which implies that all AA-genome species are relatively closely related to their most recent common ancestor and can lead to polytomy (Wambugu et al., 2015; Zhu et al., 2014; Zhu & Ge, 2005). Recent evidence using whole chloroplast sequences suggests that *O. longistaminata* is basal to the clade; its closest relative may be the perennial South American species *O. glumaepatulata* (Wambugu et al., 2015). The divergence of *O. longistaminata* from the AA-genome clade has been estimated to have occurred around 2 MYA (Vaughan et al., 2008). The apparent early divergence and the unique phenotype of *O. longistaminata* supports the possibility that the AA-genome clade arose in Africa, though species diversity within the AA clade and *Oryza* as a whole is greatest in Asia (Vaughan et al., 2008). A study using SSR and RAPD markers also found *O. longistaminata* to be basal to the other AA-genome species, though other species' relationships in the study were inconsistent with current understanding (Ren et al., 2003).

Other approaches suggest *O. meridionalis*, an Oceanic species, is most basal within the AA-genome clade. Park et al. (2003) drew this conclusion using MITE-AFLP (transposon) markers, as did Zhu et al. (2014) using 69 nuclear sequences containing 762 markers. Another study using chloroplast data of 6 polymorphic markers, 2 gene sequences, and 12 indels suggested *O. rufipogon* as the basal AA-genome species (Yin et al., 2015), but the relationship was not replicated if 8 nuclear SSR markers were used instead. Zhu et al. (2014) acknowledge that further research is needed to determine whether *O. longistaminata* or *O. meridionalis* is the basal species. There is a dearth of studies with adequate genome coverage to determine phylogenetic relationships within the AA-genome. For example, no phylogenetic analysis of NGS nuclear markers or the 6K SNP-chip of the International Rice Research Institute has been done to the author's knowledge. However, a finding common to all studies is that *O. longistaminata* is relatively distant from *O. sativa* (Vaughan et al., 2008).

Counterintuitively, almost all studies have found that *O. longistaminata* is more closely related to Asian, South American, and Australian species than to the African species *O. barthii* and *O. glaberrima*, including those with highest coverage of the nuclear genome (2,876 SNPs and 6,485 indels discovered in 60,821 bp of aligned nuclear sequences; Zhu et al., 2014) and the chloroplast genome (484 SNPs discovered in 143,331 bp of aligned chloroplast sequence; Wambugu et al., 2015) to date. Several studies of smaller genomic scope show a similar pattern (Ren et al., 2003; Duan et al., 2007; Kwon et al.,

2005; Marathi et al., 2015). Few analyses have shown *O. longistaminata* grouping closely with other African AA-genome species (Park et al., 2003; Yin et al. 2015). This suggests that geography alone does not explain the radiation of the AA-genome clade.

O. longistaminata is highly diverse compared to some other rice wild relatives, which is consistent with its outcrossing and perennial habit. In a sample of individuals from Kenyza, Tanzania, Mozambique, Zambia, Zimbabwe, Botswana, Namibia, and Madagascar, 96.5% of 184 AFLP markers were polymorphic (Kiambi et al., 2005). Though the study did not cover the full geographic range of *O. longistaminata*, which includes West Africa, Kiambi et al. found a region in Tanzania to be most diverse (2005). A survey of diversity in a sample of Ethiopian individuals found that all 64 SSRs amplified were polymorphic (Melaku et al., 2013). In *O. longistaminata*, diversity is generally higher within than among subpopulations (Kiambi et al., 2005; Ghesquiere, 1989).

Interspecific Hybridization and Breeding Barriers

Chu & Oka (1969) characterized interspecific breeding barriers between *O. longistaminata* and other AA-genome species as defined at the time of the study: *O. sativa* (modern *O. sativa*), Asian *O. perennis* (*O. rufipogon*), American *O. perennis* (*O. glumaepatulata*), Oceanic *O. perennis* (Australian *O. rufipogon*), *O. glaberrima* (*O. glaberrima*), and *O. breviligulata* (*O. barthii*). *O. longistaminata* was unique among all other AA-genome species screened in that its use as the female parent resulted in very low success rates of 0—5% in obtaining interspecific seeds during hybridization with any other species; in contrast, the range of success rates for all other species used as the female parent (in all combinations) was 22—66% (Chu & Oka, 1969). If used as the male parent with any other species, the rate of success in obtaining interspecific seeds from *O. longistaminata* ranged from 37—66% (Chu & Oka, 1969), though hybridity was not confirmed with molecular markers. Similarly, Lu et al. (2003) found seed set from 1.5% to 16.1% if *O. longistaminata* was used as female and seed set from 10.4% to 25.7% if used as male in crosses with all other AA-genome species. Although the undomesticated African species *O. barthii* is commonly sympatric with *O. longistaminata*, no evidence for interspecific hybridization between these species has been found; interspecific crossability was the lowest of all species combinations observed, with success rates of 0% and 1.5% in the cross *O. longistaminata*/*O. barthii* and success rates of 24.8% and 37% for *O. barthii*/*O. longistaminata* (Chu & Oka, 1969; Lu et al., 2003). Ghesquiere (1988) suggested that selection against introgression from *O. barthii* in *O. longistaminata* has strengthened the breeding barrier between the two species. In contrast, hybridization with *O. glaberrima* (African cultivated rice which was derived from and is closely related to *O. barthii*) ranged from 16.1% seed set

for *O. longistaminata*/*O. glaberrima*—the maximum observed of crosses between any species and female *O. longistaminata*—and 25.7% for *O. glaberrima*/*O. longistaminata* (Lu et al., 2003). Overall, interspecific hybridization with *O. longistaminata*, which is self-incompatible, may be inhibited by unilateral incompatibility with self-compatible species of the AA-genome clade; abnormalities in pollen tube formation may preclude fertilization, leading to a pre-zygotic breeding barrier (Lewis & Crowe, 1958).

However, the capacity of *O. longistaminata* to hybridize with other species has mainly been focused on domesticated Asian rice, *O. sativa*, primarily because *O. longistaminata* is a potential gene donor to *O. sativa*. Concern that transgenes in *O. sativa* may escape to wild relatives has also promoted research on interspecific crossability (Causse & Ghesquiere, 1991; Kanya, 2010; Kilewa, 2014). While hybrids of *O. sativa* and *O. longistaminata* from field and laboratory settings have been documented, there nonetheless exists a breeding barrier between *O. longistaminata* and *O. sativa*. Chu & Oka (1969, 1970) found that the post-zygotic barrier was controlled by a set of complementary dominant genes where interspecific heterozygosity conferred lethality rather than chromosomal or cytoplasmic interspecific incompatibility. Because Chu & Oka (1969) observed that *O. sativa*/*O. longistaminata* crosses produced seed 49% of the time, whereas *O. longistaminata*/*O.sativa* crosses had success rates of only 3%, they hypothesized that the complementary lethal genes influenced endosperm development. Doubled copies of dominant *O. longistaminata* alleles in the endosperm led to faster deterioration than doubled copies of dominant *O. sativa* alleles, resulting in directionality in success of crosses between *O. longistaminata* and *O. sativa* (Chu & Oka, 1970). In *O. longistaminata*/*O. sativa* crosses Chu & Oka (1970) observed necrosis of endosperms and cessation of embryo growth ~3-4 days after pollination (DAP), whereas embryo growth of the *O. sativa*/*O. longistaminata* crosses ceased 10 DAP. By 15 DAP, 97% of the *O. sativa*/*O. longistaminata* embryos showed developmental abnormalities relative to the parental embryos, such as lack of tissue differentiation of the coleoptile and radicle from the plumule primordium (Chu & Oka, 1970). The endosperms of the *O. sativa*/*O. longistaminata* crosses also had multiple abnormalities (low numbers of cells formed, irregularly shaped cells, lack of starch deposit, lack of aleuron layer formation, and excessive nucellus growth around the embryo), ultimately leading to failure of starch deposition, and resulting in shrunken seeds (Chu & Oka, 1970). The *S40* locus on chromosome 1, which is associated with pollen sterility and partial embryo-sac sterility in heterozygous near-isogenic lines derived from an interspecific cross of *O. longistaminata* and *O. sativa* (Chen et al., 2017), is consistent with the locus described by Chu & Oka (1970). De-repression of MADS box genes including *OsMADS87*, *OsMADS81-85*, and *OsMADS99* 0-7 DAP, possibly regulated upstream

by the PCR2 complex, is also thought to partially explain developmental abnormalities of the endosperm observed in crosses of *O. sativa* and *O. longistaminata* (Ishikawa et al., 2011). In general, deterioration of the embryo and the endosperm co-occur, with only two exceptions observed in which either the embryo was normal and the endosperm deteriorated or vice versa (Chu & Oka, 1970).

The F_1 progeny of crosses of *O. longistaminata* and *O. sativa* may also be less viable and fertile than non-hybrid plants. In F_1 plants from crosses of *O. longistaminata* and *O. sativa,* Chu & Oka (1969) observed cessation of growth ~30 days after germination (3.2% of F_1 plants) and embryo sac sterility (13-50% of F_1 plants). Average seed set in BCF_1 progeny of *O. sativa*/*O. longistaminata* was 4% (Sacks et al., 2006), which was less than the 50% seed set predicted by the model of Chu & Oka (1969). Therefore, it is likely that the interspecific breeding barrier is controlled by multiple genes (Sacks et al., 2006). Pollen fertility in an F_1 hybrid of *O. sativa* and *O. longistaminata* was 32.53%, though the male *O. longistaminata* parent had 64.51% pollen fertility (Tao & Sripichitt, 2000), and some F_1 interspecific hybrids have been reported nearly male-sterile, with 13% pollen fertility (Ghesquiere, 1988; Sacks et al., 2006). Chen et al. (2009) identified a QTL from *O. longistaminata* on chromosome 6 associated with pollen (and spikelet) sterility in an interspecific backcross population of the *O. sativa* cultivar RD-23. The gene *S44(t)*, also on chromosome 6, reduced pollen fertility in hybrid heterozygotes as well, and it appeared to interact strongly with environment (Zhao et al., 2012). Another QTL associated with pollen sterility was mapped to chromosome 1 (Taneichi et al., 2005).

Though a significant interspecific breeding barrier exists, several studies indicate *O. longistaminata* can spontaneously hybridize with *O. sativa* in the field (Bezançon et al., 1977; Bezançon et al. 1978; Bolaji & Nwokeocha, 2013; Causse & Ghesquiere, 1991; Ghesquiere, 1989; Kanya, 2010; Kiambi et al., 2005; Kilewa, 2014). To the authors' knowledge, only two of these studies used molecular markers to confirm hybridity (Kanya, 2010; Kilewa, 2014). Interspecific hybrids between *O. sativa* and *O. longistaminata* near rice fields were first reported by Chu & Oka (1970), who called the putative hybrids *obake*—meaning monsters or ghosts. Chu & Oka (1970) did not believe that the hybrid *obake* plants they observed were F_1 plants and hypothesized that wild *O. longistaminata* populations, especially near cultivated fields, may carry cryptic introgression from domesticated rice; they noted that the cross-pollinated species may have been more likely to absorb genes from foreign pollen than the self-pollinated *O. sativa*. However, molecular testing of Chu & Oka's hypotheses was not available at the time. Recently, Kilewa (2014) confirmed interspecific hybridization in Tanzanian populations of *O. longistaminata* collected along the borders of rice fields using phenotype and one SSR marker, and

populations from the wild were used as a control. Kilewa (2014) found that in field conditions *O. longistaminata*/*O. sativa* progeny were nearly 3-fold more frequent than *O. sativa*/*O. longistaminata* progeny, and the *O. longistaminata/O. sativa* hybrids were not phenotypically obvious. Kanya (2010) made controlled bidirectional crosses of *O. sativa* and *O. longistaminata* and concluded that spontaneous hybridization was possible, though progeny were only obtained if *O. sativa* was used as the female. Kiambi et al. (2005) observed *in-situ* hybridization in Malawi between *O. longistaminata* and an unknown landrace, Kalulu, during a collection mission; however, no molecular confirmation was obtained.

Plant breeders have also made controlled crosses between *O. longistaminata* and *O. sativa*; there have been at least 11 crossing studies reporting hybrids to date (Table 1.1). Most previous crosses of *O. sativa* and *O. longistaminata* used *O. sativa* as the female parent and used embryo rescue to obtain progeny. Chu & Oka (1969) crossed *O. longistaminata* accessions of unspecified provenance to the *O. sativa japonica* cultivar Taichung-65 and obtained interspecific progeny. Ghesquiere (1988) crossed an *indica* landrace from Guinea Bissau, BS125, to a Botswanan *O. longistaminata* accession, WL-02, and obtained a single progeny that was nearly male-sterile with short rhizomes. Khush et al. (1991) obtained a hybrid progeny by crossing the *indica* cultivar IR24 and a Malian accession of *O. longistaminata* maintained in India; the *O. longistaminata* accession is now held at the IRRI genebank as accession 110404. Kaushal & Ravi (1998) crossed the *O. sativa* basmati cultivars PR 106 and Pusa Basmati 1 to at least two *O. longistaminata* accessions of unknown provenance from IRRI, BTC-C-I and BTC-L-II; only crosses to PR 106 survived, though more germinated. Maekawa (1996) crossed the *japonica* cultivars Shiokari and later Taichung-65 to a Kenyan *O. longistaminata* accession, Mpunga wa Majani, and obtained fertile progeny (Gichuhi et al., 2016). Tao & Sripichitt (2000) obtained a fertile progeny with long rhizomes from a cross between an *indica* cultivar from Thailand, RD23, and an unknown *O. longistaminata* accession from Niger (Hu et al., 2003). Lu et al. (2003) crossed *O. longistaminata* of unspecified provenance to *O. sativa* without embryo rescue and obtained 16.2% seed set, but their germinability was unconfirmed. Waheed et al. (2012) reported that crosses between *O. longistaminata* and the *O. sativa* cultivar Malakand yielded seeds without confirmed viability. Ramos et al. (2016) obtained an interspecific hybrid progeny from a cross between the Taichung 65 and *O. longistaminata* accession 110404, the same accession previously used by Khush (1991).

Some studies have successfully recovered progeny from the reciprocal cross, using *O. longistaminata* as the female parent and *O. sativa* as male. Progeny were recovered from a cross of *O.*

sativa cultivar Miandry bararata as the male parent to *O. longistaminata* of Madagascar as the female parent without embryo rescue (Rakotomalala, 2001). However, crosses using Kenyan *O. longistaminata* as females and *O. sativa* cultivar Basmati 370 as males (without the use of embryo rescue) resulted in no viable progeny (Kanya, 2010; Kanya et al., 2012). Crosses of female *O. longistaminata* of unknown provenance and male *O. sativa* without embryo rescue resulted in 3.8% seed set; viability was unconfirmed (Lu et al., 2003). Using unknown *O. longistaminata* as female and the *O. sativa* cultivar Taichung 65 as male, Chu & Oka (1970) obtained progeny without embryo rescue; they also reported that germination of F_1 seeds from *O. longistaminata*/*O. sativa* crosses was higher (83%) than *O. sativa*/*O. longistaminata* crosses (4%), though the latter produced ~16-fold as many seeds as the former. The increased viability of *O. longistaminata*/*O. sativa* F_1 seeds many explain Kilewa's observation (2014) of more frequent *O. longistaminata* female parentage in spontaneous F_1 hybrids with *O. sativa* sampled in the field: though seed production from *O. longistaminata*/*O. sativa* may be less than that from *O. sativa*/*O. longistaminata*, concordant with the complementary dominant genes conferring lethality in interspecific hybrids described by Oka & Chu (1970), the seeds produced by *O. longistaminata*/*O. sativa* crosses may be more viable and more likely to survive field conditions

It is worth noting that there may be many more unreported failures than successes in crossing, and the reported rate of success varies. Ramos (personal communication, 2017) stated that well over a thousand crosses were required to obtain a single viable F_1 progeny using four *O. longistaminata* accessions as males and multiple *O. sativa* as females. In contrast, Kaushal & Ravi (1998) obtained viable progeny from around 550 *O. sativa*/*O. longistaminata* spikelet pollinations, a success rate of ~2.5%, and Tao & Sripichitt (2000) obtained a viable progeny from 119 *O. sativa*/*O. longistaminata* spikelet pollinations for a success rate of 0.08%. Kanya (2010) reported a 6% success rate, though a sample of 8 putative hybrids screened with molecular markers revealed 3 false positives, suggesting the true success rate was closer to 2%. However, Chu & Oka (1969) reported 3% success out of 184 trials in recovering hybrid progeny without embryo rescue if *O. longistaminata* was used as the female parent, whereas the recovery rate was 49% in 1,197 trials if *O. sativa* was used as the female.

Putative hybrids of *O. longistaminata* and *O. sativa* cannot always be verified phenotypically, but characteristics have been reported for the few F_1 hybrids that exist. Interspecific F_1 hybrids vary in the number and length of rhizomes produced (Kilewa, 2014; Ghesquiere, 1988), and the F_1 hybrids have intermediate rhizomatousness relative to *O. longistaminata* (Kilewa, 2014; Sacks et al., 2006; Tao & Sripichitt, 2000). Kilewa (2014) reported that the F_1 progeny exceed both *O. sativa* and *O.*

longistaminata in height, number of tillers, and length and width of flag leaves. The heights of BC_7F_2 individuals derived from the *O. sativa* cultivar RD-23 (also used as the parent for backcrossing) and a Nigerian *O. longistaminata* accession showed transgressive segregation both above and below the parent heights as well as intermediate heights relative to the parents. Hybrids of *O. longistaminata* and *O. sativa* backcrossed to *O. sativa* may have the long awns of *O. longistaminata* and many exhibit stigma, collar, or spikelet color (Causse & Ghesquiere, 1991). Kanya et al. (2012) reported that interspecific F_1 hybrids produce more seeds than either the *O. sativa* or *O. longistaminata* parent, though no other study has produced similar results.

Breeding

Use of *O. longistaminata* in breeding domesticated rice has been limited, in part due to the interspecific breeding barrier, but valuable introgressions have resulted in improvement of multiple traits in *O. sativa*. There is also evidence for other useful traits than have not yet been utilized in domesticated rice. A major contribution of *O. longistaminata* to domesticated rice production was the *Xa21* gene, which confers resistance to multiple strains of bacterial blight, *Xanthomonas oryzae* (Khush et al, 1991), an important disease of rice. *Xa21* was later mapped and cloned, and transgenic lines were developed to confirm its function (Ronald et al., 1992; Tu et al., 1998; Wang et al., 1996).

O. longistaminata has also been used to develop super-ratooning perennial rice (Tao & Sripichitt, 2000; Zhang et al., 2017). A single planting of modern lines can produce at least four crops over two years, and ratoon yields are ~75-80% of the main crop yields (Zhang et al., 2017) with potential for considerable environmental and economic benefits (Sacks, 2013). Super-ratooning perennial rice perennates by producing new tillers from below ground rather than rhizomes—rhizome production is not required for perenniality (Sacks et al., 2003). Super-ratooning perennial rice is tolerant of short cut heights at harvest; marker-assisted selection for perennation was possible because selection of quantitative trail loci (QTL) associated with rhizome presence or expression improved recovery of super-ratooning types even when the combination of alleles was insufficient to confer rhizomatous growth (Hu et al., 2003). Rhizome presence in an *O. sativa*/*O. longistaminata* F_2 population was controlled by two dominant complementary loci, *Rhz3* and *Rhz2*, and the degree of rhizome expression (in terms of length, number of nodes, degree of branching, and biomass dry weight) was controlled by many QTLs (Hu et al., 2003; Tao et al., 2001). Some QTLs implicated in degree of expression are located near genes that are differentially expressed in rhizome vs. shoot tissue (Hu et al., 2011). Interestingly, not all genes that increase rhizomatousness in interspecific populations are from *O. longistaminata* because of fixation of

the trait in this species (Hu et al., 2003). The frequency of rhizomatous individuals in a BC_1F_1 population produced from an *O. sativa*/*O. longistaminata* F_1 with long rhizomes was low; only one of 162 plants produced rhizomes (Tao et al., 2001).

O. longistaminata is a source of other potentially useful alleles in domesticated rice breeding. In addition to its resistance to the bacterial blight, it is resistant to other biotic stresses. *O. longistaminata* is allelopathic to barnyard grass, especially under nitrogen- and water-limited conditions (Shen et al., 2015; Xu et al., 2010). Resistance to rice root-knot nematode *Meloidogyne graminicola* was identified in the Botswanan accession WL-02, though the accession DL01-1 from Burundi was susceptible (Soriano et al., 1999). *O. longistaminata* genes can also confer resistance to yellow stem borer in crosses to *O. sativa*, though parental material of other wild species were reported more resistant than *O. longistaminata* (Du, 2008; Panigrahi & Rajamani, 2008). *O. longistaminata* is resistant to rice yellow mottle virus, and resistance was sometimes retained in backcrossed progeny of *O. longistaminata* originating from Madagascar and the recurrent parent *O. sativa* (Thottappilly & Rossel, 1993; Rakotomalala, 2001). Transcriptome data suggests *O. longistaminata* is resistant to rice blast, *Magnaporthe oryzae* (He et al., 2014). Local people in Mali note that it is rare to observe an *O. longistaminata* plant sick or dying of disease, so it may possess more undocumented biotic stress tolerances (Gupta, 2004).

O. longistaminata is also likely to be tolerant of abiotic stress. It possesses traits related to drought adaptation, including high stomatal conductance, membrane stability, leaf thickness, and leaf elongation under stress (Liu et al., 2004; Giulani et al., 2013). Though early perennial lines derived from a cross with *O. longistaminata* were not drought-tolerant, this may be dependent on soil type and the trait may have been unintentionally lost during breeding (Zhang et al., 2013). Climate data suggest that *O. longistaminata* could also be heat tolerant (Atwell et al., 2014).

O. longistaminata possesses traits with agronomic value. Its distinct floral morphology and high pollen production may be useful in producing hybrid rice efficiently (Marathi et al., 2015). Several studies and a patent have also reported that genes from *O. longistaminata* confer nitrogen-use efficiency and yield maintenance when nitrogen is limited, possibly due to its interactions with nitrogen-fixing bacteria (Gichuhi et al., 2016; Reinhold-Hurek et al., 2015). Although it produces few grains, *O. longistaminata* also contains cryptic yield-enhancing alleles (Ramos et al., 2016; Gichuhi et al., 2016; Gichuhi et al., 2016). Brar (2004) reported that the high-yielding variety NSICRC112 released by IRRI contained genes introgressed from *O. longistaminata*.

Conservation and Genetic Resources

Further utilization of *O. longistaminata* in breeding programs and other contexts is dependent on conservation of genetic resources. *O. longistaminata* germplasm is currently held at the International Rice Research Institute in the Philippines, the National Institute of Genetics in Japan, the Africa Rice Centre, the Southern African Development Community Plant Genetic Resources Network (SPGRC) in Zambia, the United States Department of Agriculture, the International Livestock Research Institute in Ethiopia, and the Millenium Seed Bank Project (Wambugu et al., 2013). More research is needed to determine whether there are gaps in current collections (Wambugu et al., 2013). Especially, a comparison of diversity in materials conserved *ex situ* and growing *in situ* is needed to ensure maximal diversity has been captured by genebanks given that African biodiversity is increasingly threatened by land use changes (Wambugu et al., 2013).

No *in situ* conservation programs directly targeting *O. longistaminata* or any African rice species currently exist (Wambugu et al., 2013). However, some protected areas are *O. longistaminata* habitat (Wambugu et al., 2013). The main threat to *O. longistaminata* is habitat destruction. Damming rivers, which decreases inundation, can reduce *O. longistaminata* range or even cause local extinction (Kleynhans et al., 2007). Not only does inundation protect the species from overgrazing, but it is also required to preserve the species' niche (Kleynhans et al., 2007). A study assessing the ecological impacts of damming the Nyl River found that it would reduce the range of *O. longistaminata*, though controlled releases could mitigate the damage (Kleynhans et al., 2007; Higgins et al., 1996). Active rehabilitation of the Logone floodplain in North Cameroon through controlled releases was shown to revitalize the *O. longistaminata* population, which had been reduced by damming (Scholte, 2005). Draining of swamps and wetlands for agriculture also reduces *O. longistaminata* habitat, as does its eradication from farmer's fields as a weed (Kiambi et al., 2005).

Conclusion

The rice wild relative *O. longistaminata* is ecologically important, highly diverged from *O. sativa*, genetically diverse, and an important contributor to past rice breeding efforts. Though challenges exist in conserving and utilizing this species, including a significant breeding barrier, available literature suggests it has potential to make further contributions to rice production.

Table

Table 1.1. Summary of crossing success rates and methods from laboratory efforts to hybridize *O. longistaminata* and *O. sativa*.

O. longistaminata parent	*O. sativa* parent	Florets Evaluated	Germinable Progeny Obtained	Success Rate (Germinable Progeny Obtained / Florets Evaluated *100)	Rescue method	Pollen Fertility of Progeny	Citation
O. longistaminata as male							
WL-02 (Botswana)	BS125, *indica*		1			7.8%	Ghesquiere, 1988
unknown	Taichung-65, *japonica*	1197	44	3.7%	none	20.5%	Chu & Oka, 1969
110404	IR24, *indica*		1				Khush et al., 1991
BTC-L-I	PR 106, basmati	268	18	6.7%	embryo	8%	Kaushal & Ravi, 1998
BTC-L-II	PR 106, basmati	280	6	2.1%	embryo	8%	Kaushal & Ravi, 1998
BTC-C-I	Pusa Basmati 1, basmati	135	5	3.7%	embryo		Kaushal & Ravi, 1998
BTC-L-II	Pusa Basmati 1, basmati	107	2	1.9%	embryo		Kaushal & Ravi, 1998
Mpunga wa Majani	Shiokari, *japonica*						Maekawa, 1996
Mpunga wa Majani	Taichung-65, *japonica*		1		ovary	73.5%	Maekawa, 1996; Gichuhi et al., 2016
unknown (Nigerian)	RD23, *indica*	119	1	3.03%	ovary	32.5%	Tao & Sripichitt, 2000
unknown	Malakand	218	0-4*	0-1.8%*	none		Waheed et al., 2012
110404	Taichung-65, *japonica*		1	**	embryo		Ramos et al., 2016
unknown	unknown	3,595	***	***			Lu et al., 2003
unknown (Kenyan)	Basmati 370	~5,880	330****	~6%****	none		Kanya, 2010
O. longistaminata as female							
unknown (Kenyan)	Basmati 370	~5,880	0	0%	none		Kanya, 2010
unknown (Madagascar)	Miandry bararata				none		Rakotomalala, 2001
unknown	Taichung-65, *japonica*	184	153	83.3%	none	48.1%	Chu & Oka, 1969
							Lu et al., 2003

*Progeny were not tested for germination, therefore the range of possible values was reported from the number of seeds obtained.

**Ramos indicated in personal communication (2017) that well over 1,000 floret pollinations were required to obtain the progeny.

***Though germinability of the progeny was not reported, seed set was 16.2%.

****Kanya (2010) confirmed hybridity with molecular markers in a subset of the germinable progeny, and the true rate of hybrid recovery was likely closer to 2%.

*****Though germinability of the progeny was not reported, seed set was 3.8%.

Literature Cited

Abo, M., Alegbejo, M., Sy, A., & Misari, S. (2000). An Overview of the Mode of Transmission, Host Plants and Methods of Detection of Rice Yellow Mottle Virus. *Journal of Sustainable Agriculture*, *17*(1), 89–111. https://doi.org/10.1300/J064v17n01

Arbeiter, S., & Tegetmeyer, C. (2011). Home Range and Habitat use by Aquatic Warblers Acrocephalus paludicola on their Wintering Grounds in Northwestern Senegal. *Acta Ornithologica*, *46*(2), 117–126. https://doi.org/10.3161/000164511X625883

Atwell, B. J., Wang, H., & Scafaro, A. P. (2014). Could abiotic stress tolerance in wild relatives of rice be used to improve Oryza sativa? *Plant Science*, *215–216*(February), 48–58. https://doi.org/10.1016/j.plantsci.2013.10.007

Awika, J. M. (2011). Major Cereal Grains Production and Use around the World. In S. Awika, J., Piironen, V., Bean (Ed.), *Advances in Cereal Science: Implications to Food Processing and Health Promotion* (pp. 1–13). American Chemical Society. https://doi.org/10.1021/bk-2011-1089.ch001

Ba, N. M., Dakouo, D., Nacro, S., & Karamage, F. (2008). Seasonal abundance of lepidopteran stemborers and diopsid flies in irrigated fields of cultivated (Oryza sativa) and wild rice (Oryza longistaminata) in western Burkina Faso. *International Journal of Tropical Insect Science*, *28*(1), 30. https://doi.org/10.1017/S1742758408930435

Bah, S., van der Merwe, R., & Labuschagne, M. T. (2017). Estimation of outcrossing rates in intraspecific (Oryza sativa) and interspecific (Oryza sativa × Oryza glaberrima) rice under field conditions using agro-morphological markers. *Euphytica*, *213*(4), 81. https://doi.org/10.1007/s10681-017-1872-x

Bezançon, G; Bozza, J; Koffi, G; Second, G. (1977). *Variability of Oryza longistaminata and the Sativa Complex of Oryza in African. Ecological and Evolutive Aspects.* Bouake.

Bezançon, G., Bozza, J., Koffi, G., & Second, G. (1978). Genetic Diversity of Indigenous Rice in Africa. In G. J. Buddenhagen, I.W., Persley (Ed.), *Rice in Africa1* (pp. 7–11). Ibadou: Academic Press.

Bezançon G, Causse M, Ghesquière A, De Kochko A, Pham J, Second G, et al. (1989). Les riz en Afrique : diversité génétique, relations interspécifiques et évolution. *Bull. Soc. Bot. Fr.*, *314*(July), 251–262. https://doi.org/10.1080/01811789.1989.10826979

Bolaji, A. O., & Nwokeocha, C. C. (2013). Issues concerning reproductive isolation in a rice hybrid swarm involving Oryza sativa Linn ., O . longistaminata A . Chev . et Roehr . and Oryza glaberrima Steud . located in Jebba Nigeria, *7*(October), 2040–2049.

Brar, D. S. (2004). Broadening the gene pool of rice through introgression from wild species. *New York*, 2000–2003.

Brar, D. S., & Khush, G. S. (1997). Alien introgression in rice. *Plant Molecular bBology*, *35*, 35–47. https://doi.org/10.1023/A:1005825519998

Catarino, L., Martins, E. S., & Diniz, M. A. (2002). Vegetation structure and ecology of the Cufada Lagoon (Guinea-Bissau). *African Journal of Ecology*, *40*(3), 252–259. https://doi.org/10.1046/j.1365-2028.2002.00367.x

Causse, M. A.; Ghesquiere, A. (1991). Prospective use of Oryza longistaminata for rice breeding. In *Rice Genetics II: Proceedings of the Second International Rice Genetics Symposium* (pp. 81–89). Manila:

International Rice Research Institute.

Chen, Z., Hu, F., Xu, P., Li, J., Deng, X., Zhou, J., … Tao, D. (2009). QTL analysis for hybrid sterility and plant height in interspecific populations derived from a wild rice relative, Oryza longistaminata. *Breeding Science*, *59*(4), 441–445. https://doi.org/10.1270/jsbbs.59.441

Chu, Y., Morishima, H., & Oka, H. (1969). Partial Self-Incompatibility Found in Oryza perennis subsp. barthii. *Japan Journal of Genetics*, *44*(4), 225–229.

Chu, Y., & Oka, H. (1970a). Introgression Across Isolating Barriers in Wild and Cultivated Oryza Species. *Evolution*, *24*(2), 344–355.

Chu, Y., & Oka, H. (1970b). The Genetic Basis of Crossing Barriers Between Oryza perennis subsp. barthii and its Related Taxa. *Evolution*, *24*(March), 135–144.

Dörgeloh, W. G. (1999). Assessment of veld conditions with multivariate techniques in mixed bushveld, South Africa. *African Journal of Ecology*, *37*(2), 194–201. https://doi.org/10.1046/j.1365-2028.1999.00168.x

Du, M. (2008). *Transfer of Resistance to Yellow Stem Borer (Scirpophaga incertalus Walker) from Wild Species into Rice (Oryza sativa L.) and Molecular Characterization of Introgression*. University of the Phillipines Los Banos.

Duan, S., Lu, B., Li, Z., Tong, J., Kong, J., Yao, W., … Zhu, Y. (2007). Phylogenetic analysis of AA-genome Oryza species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochemical Genetics*, *45*(1–2), 113–129. https://doi.org/10.1007/s10528-006-9062-x

Ghesquière, A. (1986). Evolution of Oryza longistaminata. In *Rice Genetics: Proceedings of the International Rice Genetics Symposium* (pp. 15–25). Manila: International Rice Research Institute. https://doi.org/10.1017/CBO9781107415324.004

Gichuhi, E., Himi, E., Ahmed, N., Takahashi, H., & Maekawa, M. (2016). Preliminary QTL detection for improving basmati rice in a F2 population derived from the cross between kernel basmati and pLIA-1 carrying Oryza longistaminata chromosome. *Sabrao Journal of Breeding and Genetics*, *48*(4), 402–415.

Gichuhi, E., Himi, E., Takahashi, H., & Maekawa, M. (2016). Characterization and QTL Analysis of Oryza longistaminata Introgression Line, pLIA-1, derived from a Cross between Oryza longistaminata and Oryza sativa (Taichung 65) under Non-fertilized Conditions. *Rice Research: Open Access*, *4*(4). https://doi.org/10.4172/2375-4338.1000174

Giuliani, R., Koteyeva, N., Voznesenskaya, E., Evans, M. A., Cousins, A. B., & Edwards, G. E. (2013). Coordination of Leaf Photosynthesis, Transpiration, and Structural Traits in Rice and Wild Relatives (Genus Oryza). *Plant Physiology*, *162*(3), 1632–1651. https://doi.org/10.1104/pp.113.217497

Goulart, I. C. G. R., Borba, T. C. O., Menezes, V. G., & Merotto, A. (2014). Distribution of Weedy Red Rice (Oryza sativa) Resistant to Imidazolinone Herbicides and its Relationship to Rice Cultivars and Wild Oryza Species. *Weed Science*, *62*(2), 280–293. https://doi.org/10.1614/WS-D-13-00126.1

Green, J., & El-Moghraby, A. (2009). Swamps of the Upper White Nile. In H. Dumont (Ed.), *The Nile: Origin, Environments, Limnology and Human Use* (pp. 193–204). Springer Science and Business Media. https://doi.org/10.1007/978-1-4020-9726-3_10

Gupta, A. (2004). *WIPO-UNEP Study on the Role of Intellectual Property Rights in the Sharing of Benefits Arising from the Use of Biological Resources and Associated Traditional Knowledge*.

He, R., Salvato, F., Park, J.-J., Kim, M.-J., Nelson, W., Balbuena, T. S., … Gang, D. R. (2014). A systems-wide comparison of red rice (Oryza longistaminata) tissues identifies rhizome specific genes and proteins that are targets for cultivated rice improvement. *BMC Plant Biology*, *14*(1), 46. https://doi.org/10.1186/1471-2229-14-46

Higgins, S. I., Coetzee, M. A. S., Marneweck, G. C., & Rogers, K. H. (1996). The Nyl flooplain, South Africa, as a functional unit of the landscape: A review of current information. *African Journal of Ecology*, *34*, 131–145.

Hu, F., Wang, D., Zhao, X., Zhang, T., Sun, H., Zhu, L., … Li, Z. (2011). Identification of rhizome-specific genes by genome-wide differential expression Analysis in Oryza longistaminata. *BMC Plant Biology*, *11*(1), 18. https://doi.org/10.1186/1471-2229-11-18

Hu, F. Y., Tao, D. Y., Sacks, E., Fu, B. Y., Xu, P., Li, J., … Li, Z.-K. (2003). Convergent evolution of perenniality in rice and sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4050–4054. https://doi.org/10.1073/pnas.0630531100

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N., & Kinoshita, T. (2011). Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant Journal*, *65*(5), 798–806. https://doi.org/10.1111/j.1365-313X.2010.04466.x

Johnson, D. E., Wopereis, M. C. S., Mbodj, D., Diallo, S., Powers, S., & Haefele, S. M. (2004). Timing of weed management and yield losses due to weeds in irrigated rice in the Sahel. *Field Crops Research*, *85*(1), 31–42. https://doi.org/10.1016/S0378-4290(03)00124-2

Kanya, J. I. (2010). *Gene Flow Between Cultivated Rice (Oryza sativa) and the Wild Rice (Oryza longistaminata), and its Potential Ecological Consequences in Kenya*. University of Nairobi.

Kanya, J. I., Hauser, T. P., Kinyamario, J. I., & Amugune, N. O. (2012). Hybridization Potential between Cultivated Rice Oryza sativa and African Wild Rice Oryza longistaminata. *International Journal of Agricultural Research*, 1–12. https://doi.org/10.3923/ijar.2012

Katayama, T., Rakotonjanahary, X., Nakagama, A., Sumi, A. (1987). Distribution and some morphological characters of wild rice in Tanzania. *Kagoshima University Research Center Occasional Papers*, *10*, 24–29. Retrieved from http://ir.kagoshima-u.ac.jp/handle/10232/15951

Kaushal, P., & Ravi. (1998). Crossability of wild species of Oryza with O. sativa cvs PR 106 and Pusa Basmati 1 for transfer of bacterial leaf blight resistance through interspecific hybridization. *The Journal of Agricultural Science*, *130*(4), 423–430. https://doi.org/10.1017/S0021859698005383

Khush, G., Bacalangco, E., & Ogawa, T. (1991). A New Gene for Resistance to Bacterial Blight from O. longistaminata. *Rice Genetics Newsletter*, *7*.

Kiambi, D. K., Ford-Lloyd, B. V, Jackson, M. T., Guarino, L., N., M., & Newbury, H. J. (2005). Collection of wild rice (Oryza L.) in east and southern Africa in responsed to genetic erosion. *PGR Newsletter*, *142*(142), 10–20.

Kiambi, D., Newbury, H., Ford-Lloyd, B., & Dawson, I. (2005). Contrasting genetic diversity among Oryza longistaminata (A. Chev et Roehr) populations from different geographic origins using AFLP. *African*

Journal of Biotechnology, *4*(4), 308–317. Retrieved from http://www.academicjournals.org/ajb/pdf/Pdf2005/Apr/Kiambi et al.pdf

Kilewa, R. A. (2014). *Distribution of wild rice species and hybridization between cultivated rice and Oryza longistaminata in Tanzania*. University of Adelaide.

Kleynhans, M. T., James, C. S., & Birkhead, A. L. (2007). Hydrologic and hydraulic modelling of the Nyl River floodplain Part 3: Applications to assess ecological impact. *Water SA*, *33*(1), 21–25. https://doi.org/http://dx.doi.org/10.4314/wsa.v33i1.47867

Kovach, M., & McCouch, S. (2008). Leveraging natural diversity: back through the bottleneck. *Current Opinion in Plant Biology*, *11*(2), 193–200. https://doi.org/10.1016/j.pbi.2007.12.006

Kwon, S.-J., Park, K.-C., Kim, J.-H., Lee, J. K., & Kim, N.-S. (2005). Rim 2/Hipa CACTA transposon display: a new genetic marker technique in Oryza species. *BMC Genetics*, *6*, 15. https://doi.org/10.1186/1471-2156-6-15

Labrada, R. (2002). Major weed problems in rice – red / weedy rice and the Echinochloa complex. In *FAO Rice Information* (pp. 11–17).

Lewis, D., & Crowe, L. K. (1958). Unilateral interspecific incompatibility in flowering plants. *Heredity*, *12*(i), 233–256. https://doi.org/10.1038/hdy.1958.26

Liu, L., Lafitte, R., & Guan, D. (2004). Wild Oryza species as potential sources of drought-adaptive traits. *Euphytica*, *138*(2), 149–161. https://doi.org/10.1023/B:EUPH.0000046801.27042.14

Lu, B., Naredo, M., Juliano, A., & Jackson, M. (2003). Genomic relationships of the AA genome Oryza species. In *Advances in Rice Genetics* (pp. 118–121).

Maekawa, M. (1996). Identifying cold tolerance genes at the booting stage. In G. Khush (Ed.), *Rice Genetics III: Proceedings of the Third International Rice Genetics Symposium* (pp. 428–433). Manila: International Rice Research Institute.

Marathi, B., Ramos, J., Hechanova, S. L., Oane, R. H., & Jena, K. K. (2015). SNP genotyping and characterization of pistil traits revealing a distinct phylogenetic relationship among the species of Oryza. *Euphytica*, *201*(1), 131–148. https://doi.org/10.1007/s10681-014-1213-2

Melaku, G., Haileselassie, T., Feyissa, T., & Kiboi, S. (2013). Genetic diversity of the African wild rice (Oryza longistaminata Chev. et Roehr) from Ethiopia as revealed by SSR markers. *Genetic Resources and Crop Evolution*, *60*(3), 1047–1056. https://doi.org/10.1007/s10722-012-9900-0

Mohanty, S., Wassmann, R., Nelson, A., Moya, P., & Jagadish, S. V. K. (2013). *Rice and climate change: significance for food security and vulnerability* (No. 23). Copenhagen. Retrieved from https://ccafs.cgiar.org/publications/rice-and-climate-change-significance-food-security-and-vulnerability

Murray-Hudson, M., Combs, F., Wolski, P., & Brown, M. T. (2011). A vegetation-based hierarchical classification for seasonally pulsed floodplains in the Okavango Delta, Botswana. *African Journal of Aquatic Science*, *36*(3), 223–234. https://doi.org/10.2989/16085914.2011.636904

Nasrin, S., Bergman Lodin, J., Jirström, M., Holmquist, B., Andersson Djurfeldt, A., & Djurfeldt, G. (2015). Drivers of rice production: evidence from five Sub-Saharan African countries. *Agriculture & Food Security*, *4*(1), 12. https://doi.org/10.1186/s40066-015-0032-6

Oka, H. I. (1974). Experimental studies on the origin of cultivated rice. *Genetics*, *78*(1), 475–486. Retrieved from http://www.genetics.org/content/78/1/475.short

Panigrahi, D., Rajamani, S. (2008). Genetic Evaluation of the Wild Oryza Species for Resistance Against the Yellow Stem Borer, Scirophaga Incertulas Wlk. *Journal of Plant Protection and the Environment*, *5*(1), 26–29.

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. *Theoretical and Applied Genetics*, *107*(2), 203–209. https://doi.org/10.1007/s00122-003-1252-x

Petersen, G., & Fohrer, N. (2010). Flooding and drying mechanisms of the seasonal Sudd flood plains along the Bahr el Jebel in southern Sudan. *Hydrological Sciences Journal*, *55*(1), 4–16. https://doi.org/10.1080/02626660903525278

Rakotomalala, M. (2001). Case History on Development and Deployment of Resistant Varieties to Control RYMV in Madagascar. In B. Hughes, J.; Odu (Ed.), *Plant Virology in Sub-Saharan Africa* (pp. 423–341). Ibadan: International Institute of Tropical Agriculture.

Ramos, J. M., Furuta, T., Uehara, K., Chihiro, N., Angeles-Shim, R. B., Shim, J., … Jena, K. K. (2016). Development of chromosome segment substitution lines (CSSLs) of Oryza longistaminata A. Chev. & R??hr in the background of the elite japonica rice cultivar, Taichung 65 and their evaluation for yield traits. *Euphytica*, *210*(2), 151–163. https://doi.org/10.1007/s10681-016-1685-3

Ren, F., Lu, 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. *TAG Theoretical and Applied Genetics*, *108*(1), 113–120. https://doi.org/10.1007/s00122-003-1414-x

Rodenburg, J., & Johnson, D. E. (2009). Weed Management in Rice-Based Cropping Systems in Africa. In *Advances in Agronomy* (1st ed., Vol. 103, pp. 149–218). Elsevier Inc. https://doi.org/10.1016/S0065-2113(09)03004-1

Ronald, P. C., Albano, B., Tabien, R., Abenes, L., Wu, K. S., McCouch, S., & Tanksley, S. D. (1992). Genetic and physical analysis of the rice bacterial blight disease resistance locus, Xa21. *Molecular & General Genetics : MGG*, *236*(1), 113–120. https://doi.org/10.1007/BF00279649

Sacks, E. J. (2013). Perennial Rice: Challenges and Opportunities. In C. Batello, L. Wade, S. Cox, N. Pogna, A. Bozzini, & J. Choptiany (Eds.), *Perennial Crops for Food Security: Proceedings of the FAO Expert Workshop* (pp. 7–16). Rome.

Sacks, E. J., Dhanapala, M. P., Tao, D. Y., Sta. Cruz, M. T., & Sallan, R. (2006). Breeding for perennial growth and fertility in an Oryza sativa/O. longistaminata population. *Field Crops Research*, *95*(1), 39–48. https://doi.org/10.1016/j.fcr.2005.01.021

Sacks, E. J., Schmit, V., McNally, K. L., & Sta. Cruz, M. T. (2006). Fertility in an interspecific rice population and its effect on selection for rhizome length. *Field Crops Research*, *95*(1), 30–38. https://doi.org/10.1016/j.fcr.2005.01.026

Sarla, N., & Swamy, B. P. M. (2005). Oryza glaberrima: A source for the improvement of Oryza sativa. *Current Science*, *89*(6), 955–963.

Scholte, P. (2005). *Floodplain Rehabilitation and the Future of Conservation & Development*. *Leiden University*. Retrieved from papers://29803081-7a18-4199-a903-38c6fcf8334a/Paper/p12446

Scholte, P. (2007). Maximum flood depth characterizes above-ground biomass in African seasonally shallowly flooded grasslands. *Journal of Tropical Ecology*, *23*(January 2007), 63–72. https://doi.org/10.1017/S026646740600366X

Scholte, P., de Kort, S., & van Weerd, M. (2000). Floodplain rehabilitation in Far North Cameroon: expected impact on bird life. *Ostrich*, *71*(December), 112–117. https://doi.org/10.1080/00306525.2000.9639884

Seck, P. A., Diagne, A., Mohanty, S., & Wopereis, M. C. S. (2012). Crops that feed the world 7: Rice. *Food Security*, *4*(1), 7–24. https://doi.org/10.1007/s12571-012-0168-1

Soriano, I. R., Schmit, V., Brar, D. S., Prot, J.-C., & Reversat, G. (1999). Resistance to rice root-knot nematode Meloidogyne graminicola identified in Oryza longistaminata and O. glaberrima. *Nematology*, *1*(4), 395–398. https://doi.org/10.1163/156854199508397

Tadjouwa Kouawa, Wanko, A., Beck, C., Mose, R., & Maïga, A. H. (2015). Feasibility study of faecal sludge treatment by constructed wetlands in Sahelian context: Experiments with Oryza longistaminata and Sporobolus pyramidalis species in Ouagadougou. *Ecological Engineering*, *84*, 390–397. https://doi.org/10.1016/j.ecoleng.2015.09.021

Taneichi, T., Koide, Y., & Nishimoto, D. (2005). Hybrid Sterility Gene S13 found in a Distantly Related Rice Species, O. longistaminata. *Genes & Genetic Systems*, *80*(6).

Tao, D., Hu, F., Yang, Y., Xu, P., Li, J., Wen, G., … Sripichitt, P. (2001). *Rhizomatous individual was obtained from interspecific BC2F1 progenies between Oryza sativa and Oryza longistaminata*. Retrieved from https://shigen.nig.ac.jp/rice/oryzabase/asset/rgn/vol18/c2.html

Tao, D., & Sripichitt, P. (2000). Preliminary Report on Transfer Traits of Vegetative Propagation from Wild Rice Species to Oryza sativa via Distant Hybridization and Embryo Rescue. *Kasetsart Journal: Natural Science*, *34*(1), 117–124.

Thottappilly, G., & Rossel, H. (1993). Evaluation of Resistance to Rice Yellow Mottle Virus in Oryza species. *Indian Journal of Virology*, *9*(1), 65–73.

Tu, J., Ona, I., Zhang, Q., Mew, T. W., Khush, G. S., & Datta, S. K. (1998). Transgenic rice variety “IR72” with Xa21 is resistant to bacterial blight. *Theoretical and Applied Genetics*, *97*(1–2), 31–36. https://doi.org/10.1007/s001220050863

Vaughan, D. A. (1994). *The Wild Relatives of Rice: A Genetic Resources Handbook* (1st ed.). Manila: International Rice Research Institute.

Waheed, A., Ahmad, H., Abassi, F. M., Hamid, F. S., Shah, A. H., Ali, H., … Ahmad, N. (2013). Interspecific crosses Oryza rufipogon and Oryza longistaminata with Oryza Sativa (Bas-385 and F. Malakand). *Journal of Materials and Environmental Science*, *4*(1), 103–108.

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. *Scientific Reports*, *5*(1), 13957. https://doi.org/10.1038/srep13957

Wambugu, P. W., Furtado, A., Waters, D. Le, Nyamongo, D. O., & Henry, R. J. (2013). Conservation and utilization of African Oryza genetic resources. *Rice (New York, N.Y.)*, *6*(1), 29. https://doi.org/10.1186/1939-8433-6-29

Wang, G. L., Song, W. Y., Ruan, D. L., Sideris, S., & Ronald, P. C. (1996). The cloned gene, Xa21, confers resistance to multiple Xanthomonas oryzae pv. oryzae isolates in transgenic plants. *Mol Plant Microbe Interact*. https://doi.org/10.1094/MPMI-9-0850

Williams, C. T., Okhidievbie, O., Harris, K. M., & Ukwungwu, M. N. (1999). The host range, annual cycle and parasitoids of the African rice gall midge Orseolia oryzivora (Diptera : Cecidomyiidae) in central and southeast Nigeria UR -./documents_pdf/Williams_1999.pdf. *Bulletin of Entomological Research*, *89*(6), 589–597. https://doi.org/10.1017/s0007485399000747

Xu, G., Zhang, F., Li, T., Wu, D., & Zhang, Y. (2010). Induced Effects of Exogenous Phenolic Acids on Allelopathy of a Wild Rice Accession (Oryza longistaminata, S37). *Rice Science*, *17*(2), 135–140.

Yin, H., Akimoto, M., Kaewcheenchai, R., Sotowa, M., Ishii, T., & Ishikawa, R. (2015). Inconsistent diversities between nuclear and plastid genomes of AA genome species in the genus *Oryza*. *Genes & Genetic Systems*, *advpub*, 269–281. https://doi.org/10.1266/ggs.14-00063

Zhang, S., Hu, J., Yang, C., Liu, H., Yang, F., Zhou, J., … Hu, F. (2017). Genotype by environment interactions for grain yield of perennial rice derivatives (Oryza sativa L./Oryza longistaminata) in southern China and Laos. *Field Crops Research*, *207*, 62–70. https://doi.org/10.1016/j.fcr.2017.03.007

Zhao, J., Li, J., Xu, P., Zhou, J., Hu, F., Deng, X., … Tao, D. (2012). A new gene controlling hybrid sterility between Oryza sativa and Oryza longistaminata. *Euphytica*, *187*(3), 339–344. https://doi.org/10.1007/s10681-012-0691-3

Zhu, Q., & Ge, S. (2005). Phylogenetic relationships among A-genome species of the genus Oryza revealed by intron sequences of four nuclear genes. *New Phytologist*, *167*(1), 249–265. https://doi.org/10.1111/j.1469-8137.2005.01406.x

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. *Molecular Biology and Evolution*, *24*(3), 875–888. https://doi.org/10.1093/molbev/msm005

Zhu, T., Xu, P. Z., Liu, J. P., Peng, S., Mo, X. C., & Gao, L. 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. *Molecular Phylogenetics and Evolution*, *70*(1), 348–361. https://doi.org/10.1016/j.ympev.2013.10.00

CHAPTER 2: POPULATION STRUCTURE AND GENETIC DIVERSITY OF THE UNDOMESTICATED RICE RELATIVE *ORYZA LONGISTAMINATA*

Abstract

The undomesticated rice relative *Oryza longistaminata* is a valuable genetic resource for improvement of the domesticated Asian rice, *O. sativa*. Understanding population structure of *O. longistaminata* across its geographic range—which includes most of sub-Saharan Africa—would facilitate germplasm conservation and management. Though past introgressions from *O. longistaminata* have improved biotic stress resistance, ratooning ability, and yield in *O. sativa*, progress has been slowed by a substantial breeding barrier between the species. This study reports the identification of three genetic subpopulations in *O. longistaminata* using genome-wide markers from 189 accessions across the species' geographic range, quantification of overall diversity *O. longistaminata* among and within its genetic subpopulations, and the identification of spontaneous interspecific hybrid individuals between *O. sativa* and *O. longistaminata* in the germplasm sampled.

Introduction

Global consumption of Asian domesticated rice, *Oryza sativa*, is projected to rise from 439 million tons to 555 million tons between 2010 and 2035 as the human population increases (Seck et al., 2012). Producing the requisite amount of rice using fewer inputs and with the additional challenge of climate change will require new sources of variation for trait improvement (Mohanty et al., 2013). The rice wild relative *O. longistaminata* is a valuable genetic resource for breeding Asian domesticated rice, *O. sativa*. A perennial, cross-pollinated species native to all of sub-Saharan Africa, *O. longistaminata* is thought to be highly diverse (Bezançon et al., 1977; Bezançon et al., 1978; Kiambi et al., 2005; Melaku et al., 2013). Of the eight AA-genome species that make up the primary germplasm pool of domesticated rice, it is one of the most genetically distinct from *O. sativa* (Wambugu et al., 2015, Zhu et al., 2014). *O. longistaminata* is distantly related to the two other AA-genome species of rice indigenous to Africa—a domesticated species, *O. glaberrima*, and its wild progenitor, *O. barthii* (Wambugu et al., 2015; Zhu et al., 2014). Though past breeding efforts have introgressed valuable genes from *O. longistaminata* to domesticated Asian rice, including bacterial blight resistance, perennial habit, and yield-enhancing components, the development of genetic resources for this understudied wild species would assist further progress (Atwell et al., 2014; Gichuhi et al., 2016; Ronald et al., 1992; Sanni et al., 2013; Zhang et al., 2017). Understanding *O. longistaminata* population structure and genetic diversity based on high

genome coverage across the full geographic range of this species would facilitate conservation planning, germplasm management, and improvement of domesticated rice (Wambugu et al., 2013).

Because hybridization with *O. sativa* is complicated by a substantial breeding barrier, manipulating the crossability between these two species could accelerate introgression efforts (Causse & Ghesquiere, 1991). The success rate of laboratory efforts at hybridization has varied from less than 1 in 1000 (Ramos, personal communication, 2017) to 1 in 50 (Kaushal & Ravi, 1998) with the outlying exception of the 1 in 2 success rate of Chu & Oka (1970). Recovering interspecific progeny has typically required embryo rescue (Causse & Ghesquiere, 1991; Khush et al., 1991; Ramos et al., 2016; Tao & Sripichitt, 2000). In contrast to the strong breeding barrier observed in the laboratory, however, Bezançon et al. (1977), Chu & Oka (1970), Ghesquiere (1986), Kanya (2010), and Kilewa (2014) reported the occurrence of spontaneous hybrids between *O. sativa* and *O. longistaminata*, especially along the edges of rice fields in Africa. Identification of natural hybrids of *O. sativa* and *O. longistaminata*, including backcross generations, could facilitate introgression efforts if plant breeders can identify individuals that have *O. longistaminata* genes but which are also highly crossable with *O. sativa*. Given the broad geographic range of *O. longistaminata*, the species' great potential as a source of genes for improving domesticated rice, and the need for a continent- and genome-wide population genetics analysis, the present study was conducted to 1) characterize population structure of *O. longistaminata* throughout its native range in sub-Saharan Africa using 189 accessions in the International Rice Research Institute (IRRI) genebank, 2) quantify genetic diversity of *O. longistaminata* overall and among its genetic groups, and 3) determine if previous reports of spontaneous interspecific hybrids between *O. sativa* and *O. longistaminata* in Africa can be confirmed in the IRRI germplasm collection.

Materials and Methods

Plant materials

In total, 190 accessions of *O. longistaminata* collected from much of the species' native geographic range in sub-Saharan Africa were studied; this included all of the available and viable accessions in the genebank of the International Rice Research Institute. On average, accessions in the IRRI genebank have undergone three seed increases in a screenhouse since acquisition by the genebank. In this study, up to two individuals from each *O. longistaminata* accession were sampled, for a total of 369 individuals. Control outgroups of *O. sativa* (n=11), *O. barthii* (n=7), and *O. glaberrima* (n=4) were obtained from the USDA National Plant Germplasm System. Based on molecular marker data, four

accessions labeled *O. glaberrima* were found to be *O. sativa*, and two accessions labeled *O. longistaminata* were found to be *O. barthii*; the above numbers reflect the species identity rather than the label. The control outgroup species were chosen because they comprise, along with *O. longistaminata*, all of the AA-genome *Oryza* species known to exist in Africa. Additionally, a known *O. sativa*/*O. longistaminata* F_1 hybrid from a controlled cross and two of its F_2 progeny were included in the study as controls for comparison with putative interspecific hybrids observed in the *O. longistaminata* accessions. The control interspecific individuals were from a cross between the Thai *O. sativa* ssp. *indica* cultivar RD-23 and a Nigerian *O. longistaminata* accession (Dayun & Sripichitt, 2000; Hu et al., 2003); molecular data from RD-23, the Nigerian *O. longistaminata* accession, and the *O. sativa* cultivar Minghui 63 were also used in the study (Zhang, unpublished raw data). In total, 397 individuals were studied: 370 *O. longistaminata*, 13 *O. sativa*, 7 *O. barthii*, 4 *O. glaberrima*, and 3 known *O. sativa*/*O. longistaminata* hybrids. Seed were aseptically germinated and *O. longistaminata* plants were grown to maturity in a greenhouse at Urbana, IL. Any aberrant phenotypes in the *O. longistaminata* individuals were recorded (e.g. lack of rhizomes, filled grains produced by selfing, and short stature; Vaughan, 1994).

Molecular markers

Leaf tissue was collected at the seedling stage and DNA was extracted using the CTAB (cetyltrimethylammonium bromide) method with minor modifications (Fulton et al., 1995). Restriction site-associated DNA sequencing (RAD-seq) libraries were prepared according to Clark et al. (2014) based on the method of Poland et al. (2012). In brief, DNA from each individual was digested with the restriction enzymes *Pst*I-HF and *Msp*I followed by ligation to barcoded adapters; then, the samples were pooled, and fragments 200-500 bp were selected and amplified by polymerase chain reaction (PCR). Libraries were sequenced on an Illumina Hi-Seq 2000 for 100 bp single-end reads at the Roy J. Carver Biotechnology Center at the University of Illinois. RAD-seq data were analyzed in the TASSEL 5.2.30 GBSv2 pipeline (Bradbury et al., 2007) and 64-kmer sequences from each read were aligned to the Nipponbare reference genome using Bowtie 2.2.4 (Langmead et al., 2009), yielding 305,086 single nucleotide polymorphisms (SNPs) among 391 individuals. Two individuals with more than 70% of sites missing were removed from the study. For each set of individuals used in a given iteration of population structure analysis, the SNPs were filtered to require a minimum minor allele frequency of 0.01 and a minimum site count equal to 75% of the number of individuals in the dataset. Only the two most common alleles of each SNP were retained.

Genetic data analysis

To identify genetic subpopulations and assign individuals to those groups, two complementary analyses of population structure were conducted using the ADMIXTURE model in the software STRUCTURE 2.3.4 (Falush et al., 2003) and discriminant analysis of principal components (DAPC) in the R package adegenet (Jombart et al., 2010; Jombart & Ahmed, 2011). Three replications of ADMIXTURE at each K=1 through K=10 were run with a burn-in of 10,000 MCMC repetitions followed by 50,000 default MCMC repetitions, and the Evanno method as implemented by StructureHarvester was used to identify the optimal number of clusters (Earl & VonHoldt, 2011). For DAPC, principal components analysis was first conducted with the *glPca* function, then the *find.clusters* function was used to make initial groupings with the *n.start* option set to 500 to ensure convergence; *dapc* was used to assign individuals to each cluster. The number of clusters with the minimum Bayesian Information Criterion (BIC) was chosen as optimal. To determine ancestry of putative hybrids, two replications of STRUCTURE were run at K=4 with the USEPOPINFO and PFROMPOPFLAGONLY options set to 1, MIGRPRIOR set to 0, and a burn-in of 10,000 MCMC repetitions followed by 50,000 default MCMC repetitions. Neighbor-joining trees were generated using Archaeopteryx 0.9893 beta (Han & Zmasek, 2009) and the R package ape (Popescu et al., 2012) to observe local topologies. Genetic distance was calculated in TASSEL 5.2.30, and all trees were rooted using *root.phylo* in the R package ape at the *O. barthii* individual from accession 311694, which had the maximum pairwise genetic distance observed. Geographic maps were drawn in ArcGIS 10.3.1 (ESRI).

Spatial principal components analysis (sPCA) was conducted in the R package adegenet to identify spatial patterns in genetic variation. To reduce computation time, SNPs were thinned to a minimum distance of 5000 bp, and only SNPs with a minor allele frequency greater than 0.01 that had been sampled in all individuals were used, yielding 5,980 SNPs. Genotypes were pooled by accession collection site and allele frequency was estimated per collection site before analysis; a connectivity network of the collection sites was generated using the minimum spanning method. The lagged principal scores were interpolated by the natural neighbor method in ArcGIS 10.3.1.

Estimates of F_{ST} (genetic differentiation among subpopulations), F_{IS} (inbreeding coefficient), and D (Nei's genetic diversity or expected heterozygosity) were calculated using a custom R script, adjusted for sample size following Nei & Chesser (1983), and averaged across loci. Pairwise estimates of genetic differentiation among genetic groups was estimated with Jost's D (Jost, 2008) and calculated using the *pairwiseJostDnumeric* function in R (Clark, 2016). To obtain a 99.2% probability under the binomial

distribution of calling heterozygous loci correctly from the RAD-seq reads, 61,882 SNPs with a minimum depth of 7 were used to calculate the population genetics parameters.

Results

Three O. longistaminata groups and one interspecific hybrid group identified

O. longistaminata was differentiated from the control outgroups *O. sativa*, *O. barthii*, and *O. glaberrima* in STRUCTURE and DAPC analyses that included all individuals except the three interspecific hybrid controls, parents of the hybrid controls, and Minghui 63 (n_{ind} = 389, n_{SNPs} = 86,535; Fig. 2.1). DAPC analysis identified genetic groups with higher resolution, and the optimal number of groups was six: three groups of *O. longistaminata*, one group of *O. sativa*, one group comprising the African species *O. glaberrima* and *O. barthii*, and one group subsequently identified as putative interspecific hybrids (Fig. 2.1). Choosing seven clusters instead of six split the *O. sativa* group into the *indica* and *japonica* subspecies. In contrast, K=2 was optimal in STRUCTURE analysis (Fig. 2.1, Fig. 2.2) with one cluster including all *O. longistaminata* individuals and the other including the all of the outgroup species, *O. sativa*, *O. barthii*, and *O. glaberrima*. A set of 16 individuals that had greater than 15% admixture with the outgroup species in STRUCTURE (Fig. 2.1) comprised a group of putative interspecific hybrids that also clustered together in the neighbor-joining tree between the outgroup species and the other *O. longistaminata* groups (Fig. 2.3; Table 2.1), and were a distinct group in DAPC. Many of the putative interspecific hybrids had phenotypes that were atypical for *O. longistaminata*.

To further investigate population structure of *O. longistaminata* without potential bias from outgroup species and putative interspecific hybrids, a second set of DAPC and STRUCTURE analyses were conducted on only the *O. longistaminata* individuals previously found to have less than 15% interspecific admixture (n_{ind} = 351, n_{SNPs} = 75,371; Fig. 2.1). DAPC again identified the three *O. longistaminata* genetic groups observed in the prior analysis (Fig. 2.1). Geographic maps of individuals in the three DAPC groups showed that they corresponded to Northwestern Africa, Pan-Africa, and Southern Africa (Fig. 2.4). STRUCTURE analysis of just the *O. longistaminata* subset identified K=2 as optimal (Fig. 2.2). Individuals without intraspecific admixture were concentrated in northwestern Africa and southern Africa, but most individuals were admixed between the two *O. longistaminata* groups (Fig. 2.1, Fig. 2.4); at K=3, all but two individuals in the Pan-African group were admixed intraspecifically. At K=3 in STRUCTURE, the group membership of individuals was unchanged from DAPC (Fig. 2.1). The three *O. longistaminata* DAPC groups also formed distinct clades within the neighbor-joining tree (Fig. 2.3).

The three *O. longistaminata* subpopulations did not show further substructure when analyzed separately in DAPC (not shown). An individual's proportion of intraspecific admixture with each *O. longistaminata* group was moderately correlated with latitude (r^2 = 0.37) and longitude (r^2 = 0.44).

Spatial principal components analysis using 5,980 SNPs at 153 geographic sites showed the overall geographic patterns of genetic structure observed with DAPC and STRUCTURE (Fig. 2.5). Two eigenvectors were retained for analysis (Fig. 2.5C). The first eigenvector accounted for 11.9% of the genetic variation between sites and showed sharp differentiation between a group of individuals in the south and all others (Fig. 2.5A). The second eigenvector, represented 3.4% of genetic variation between sites and showed differentiation from east to west; the northwestern and southern groups represented the most extreme values (Fig. 2.5B).

Because all but 16 of the 126 individuals in the Northwestern Africa *O. longistaminata* group were from a 64,000 km^2 region of Mali, it is possible that high geographic sampling density led to distinction of this group (Fig. 2.4). To mitigate potential bias due to uneven geographic sampling, STRUCTURE and DAPC were rerun using genetic data from a set of *O. longistaminata* accessions filtered to a minimum geographic distance of 25 km (n_{ind} = 173, n_{SNPs} = 74,793). STRUCTURE results still showed optimal K=2 with individuals of pure group ancestry falling in Northwestern and Southern groups. However, DAPC indicated optimal K=2, and the formerly distinct Northwestern group merged with the Pan-African group (data not shown). It is possible that similarly dense sampling of the entire species would maintain the Northwestern group and additionally reveal further substructure in other regions; for example, the sPCA suggested that there may be a distinct northeastern group despite sparse sampling from the northeast in the current dataset (Fig. 2.5A).

Genetic diversity was similar among the three *O. longistaminata* DAPC groups, with the Northwestern Africa group having the lowest estimate of D (Table 2.2). Overall, genetic differentiation among the *O. longistaminata* groups, as indicated by F_{ST} and pairwise Jost's D, was low (Table 2.2, Table 2.3). The Southern Africa group was the most diverse, the most differentiated from the other *O. longistaminata* groups, and also the most closely related to the outgroup species (Table 2.3; Fig. 2.3). In contrast to the Southern Africa group's high genetic diversity, the group's inbreeding coefficient was more than 1.5 times greater than that observed for the other *O. longistaminata* groups (Table 2.2). The Pan-Africa subpopulation was the least differentiated from the whole. Among all pairwise comparisons of the *O. longistaminata* groups, the Northwestern Africa and Southern Africa groups were the most genetically differentiated (Table 2.3).

O. sativa/O. longistaminata hybrids and bidirectional introgression

To further investigate the ancestry of the previously identified putative interspecific hybrids (>15% interspecific admixture) and the *O. longistaminata* observed to have a low proportion of interspecific admixture (up to 15%), a STRUCTURE analysis was conducted using USEPOPINFO to pre-define known groups as *O. sativa*, *O. glaberrima*, *O. barthii*, and *O. longistaminata* with zero interspecific admixture and assign other individuals ancestry from known groups. As controls, the analysis included a known *O. sativa*/*O. longistaminata* F_1 hybrid from a controlled cross and two of its F_2 progeny (n_{ind} = 391, n_{SNPs} = 85,064). As expected, the known interspecific hybrids from controlled crosses were correctly identified as hybrids of *O. sativa* and *O. longistaminata* (Table 2.1). The analysis indicated that all individuals previously observed to have more than 15% interspecific admixture had ancestry predominantly from *O. sativa* (Asian) and *O. longistaminata* (African) but negligible ancestry from *O. barthii* (African) and *O. glaberrima* (African; Table 2.1). Ancestry from *O. sativa* was 68% or higher in two individuals (Table 2.1), indicating introgression of *O. longistaminata* genes into Asian domesticated rice; however, higher ancestry from *O. longistaminata* than *O. sativa* was observed in the other 14 interspecific individuals, indicating introgression of genes from the domesticated Asian species into the undomesticated African species. Five individuals had Q-values similar to the F_1 and F_2 controls, and several other individuals had ratios near those expected in the BC_1 and BC_2 generation for introgression into both *O. longistaminata* and *O. sativa* (Table 2.1). The putative early-generation hybrids between *O. sativa* and *O. longistaminata* were likely from recent crossing events and were from areas of sub-Saharan Africa where *O. sativa* is currently cultivated (WARDA, 2008).

According to STRUCTURE with the USEPOPINFO option on, all of the individuals in the Southern Africa *O. longistaminata* group had low-level admixture (<15%) with *O. sativa* (n = 45; mean = 2.8%; range, 0.7% — 5.5%), although not all individuals showing interspecific admixture belonged to the Southern group. *O. sativa* ancestry less than 15% was rare in the Northwestern African individuals (n = 126; mean, 0.1%; range, 0.0% — 8.3%) and Pan-African individuals (n = 180; mean, 0.0%; range, 0.2% — 5.2%) *O. longistaminata* groups. To examine whether grouping was sensitive to the admixture threshold, DAPC was rerun with interspecific admixture cutoffs of 5% and 1%. At the 5% cutoff, some individuals were reassigned from the Pan-African group to the Southern Africa group, which improved group cohesion in the neighbor-joining tree (Fig. 2.3). Group membership was stable between the 5% and 1% cutoff (not shown).

Discussion

Relationship between O. longistaminata and the other AA-genome species in Africa

Our study indicated that the undomesticated African *O. longistaminata* was more closely related to Asian domesticated rice, *O. sativa*, than to the African *O. glaberrima*-*O. barthii* group even when the analyses were conducted with *O. longistaminata* individuals that exhibited less than 1% interspecific admixture (Fig. 2.3). The majority of previously conducted studies have also concluded *that O. longistaminata* is more closely related to *O. sativa* than to *O. barthii* and *O. glaberrima* (Wambugu et al., 2015; Zhu et al., 2013; Ren et al., 2003; Duan et al., 2007; Kwon et al., 2006; Marathi et al., 2015; Cheng et al., 2002); only two studies show *O. longistaminata* to be more closely related to *O. barthii* and *O. glaberrima* than *O. sativa* (Park et al., 2003; Yin et al., 2015.) If *O. longistaminata* is truly more closely related to its Asian relative *O. sativa*, than its African relatives, *O. glaberrima* and *O. barthii*, then this would suggest that Africa has two distinct lineages of native AA-genome rice species, likely associated with independent migration events.

Population structure in O. longistaminata

To the authors' knowledge, this is the first study to evaluate *O. longistaminata* population structure across most of sub-Saharan Africa using densely-spaced genome-wide molecular markers. Three genetic groups of *O. longistaminata* were identified, two of which were associated with distinct geographic regions of Africa (Northwestern and Southern), which will be useful information for conserving germplasm of this species and for using this wild relative to improve domesticated rice. Though differentiation among the three *O. longistaminata* populations was low, geographic distance appeared to be the main factor associated with genetic differentiation; sharp barriers to gene flow were not observed. STRUCTURE, DAPC, sPCA, and Jost's D consistently indicated that the extremes of differentiation in *O. longistaminata* were between the Northwestern and Southern populations (Fig. 2.1, Fig. 2.4, Fig. 2.5, Table 2.3). Similarly, the individuals of the Pan-African group showed a gradient of admixture with each of the other genetic groups roughly according to their geographic proximity to each. These observations were consistent with the species' biology, which includes perennation with dispersal of rhizomes along rivers during floods (Green & El-Moghraby, 2009; Kleynhans et al., 2007), obligate outcrossing due to self-incompatibility (Ghesquiere, 1986), dispersal of seed by birds (Gupta, 2004), and adaptation to tropical environments that were relatively stable during periods of glaciation (Maley, 1996).

The mechanisms of genetic differentiation within *O. longistaminata* likely included drift and perhaps isolation associated with differences in flowering time. Populations at the edges of a species' geographic range, such as the Northwestern and Southern here, can have low population densities that leave individuals with few neighbors with which to outcross; over time, lower effective population size accelerates genetic drift and increases inbreeding. Consistent with this scenario, the Southern African *O. longistaminata* group had substantially greater inbreeding than the other groups, though it also had unexpectedly high genetic diversity. Given that the Northwestern and Southern populations are separated by more than 20 degrees of latitude and the equator, variation in flowering time could also lead to isolation. Variation in flowering time could be due to genetic differences in day-length sensitivity, the timing of the growing season, or a combination of both. *O. longistaminata* has been previously observed to be a short-day plant, though some individuals (most commonly observed near domesticated rice fields) are insensitive to photoperiod (Ghesquiere, 1986). In the Urbana, IL greenhouse (40°N) most accessions flowered only during the short days of late autumn and winter, with substantial differences in flowering time among accessions; however, some flowered during the long days of summer (i.e. were apparently day-neutral). Similarly, *O. longistaminata* accessions collected in Ethiopia (8-12°N) were reported to flower during long days at Jinghong Rice Breeding Station in China (20°N; Melaku, personal communication, 2017).

Origins of the Southern Africa O. longistaminata group

The Southern Africa group was unique among the three *O. longistaminata* groups in that all individuals had low-level estimated admixture with *O. sativa*. Two competing hypotheses are proposed to account for the apparent interspecific admixture that is a defining feature of the Southern Africa group: 1) this group is more similar to the ancestral *O. longistaminata* population than the other groups, and the apparent admixture with *O. sativa* is an artifact that represents alleles in common to the ancestral *O. longistaminata* population and its Asian AA-genome relatives, with many of the alleles subsequently lost in the more derived Pan-African and Northwestern Africa groups, or 2) the low level of admixture with *O. sativa* in the Southern Africa *O. longistaminata* group was the result of ancient interspecific hybridizations and subsequent introgression.

The proximity of the Southern Africa group to the outgroup species in the neighbor-joining tree could be consistent with the first hypothesis of greater similarity to the ancestral *O. longistaminata* population, but would also be expected if the admixture was actually the result of interspecific hybridization (Fig. 2.3). Additionally, the ordered decrease in genetic diversity of the three *O.*

longistaminata groups from Southern Africa to Pan-Africa to Northwestern Africa could indicate that the species radiated north and west from a southern center of diversity—however, introgression of alleles from another species could explain the high diversity observed in the Southern group (Table 2.2).

Bolstering the interspecific hybridization hypothesis are studies by Kanya (2010) and Kilewa (2014) which used molecular markers to independently verify that modern populations of *O. longistaminata* spontaneously hybridized with domesticated Asian rice along the edges of paddies in Kenya and Tanzania respectively. The data presented here also indicate that recent hybridizations between *O. sativa* and *O. longistaminata* and subsequent bidirectional introgressions have occurred in multiple locations spanning the geographic range of *O. longistaminata*. Intriguingly, recent archaeological data indicates that *O. sativa* was introduced to southeastern Africa via Madagascar and the Comoros Islands by farmers who migrated from Southeast Asia as early as ~1,000 years before present (Nayar, 2014), which could account for a low level of ancient interspecific admixture in the current Southern Africa *O. longistaminata* group. The ancient hybridization hypothesis is consistent with both the high genetic diversity and high inbreeding estimates observed for the Southern Africa *O. longistaminata* group because introgression of alleles from another species would be expected to increase genetic diversity, and *O. sativa*/*O. longistaminata* F_1 hybrids are often self-compatible (Ghesquiere, 1986; Hu et al., 2003). High rates of inbreeding would not be typically expected if the population represented a center of origin and diversity. If the ancient hybridization hypothesis is correct, then the Pan-African group is likely the most similar to the ancestral *O. longistaminata* population, and the Southern and Northwestern groups at the edges of the species range are likely derived.

Implications of spontaneous O. sativa/O. longistaminata hybridization and bidirectional introgression for germplasm conservation and breeding

The current study and at least two prior studies (Kanya, 2010; Kilewa, 2014) have produced molecular evidence of recent spontaneous hybridizations between *O. sativa* and *O. longistaminata*, lending credence to studies that documented interspecific hybrids without confirmation with molecular markers (Benzaçon et al., 1977; Chu & Oka, 1970; Ghesquiere, 1986; Kiambi et al., 2005). Spontaneous progeny derived from backcrosses to each of the parent species have also been observed (Table 2.1). In contrast to findings of spontaneous interspecific hybridization, plant breeders have typically regarded the cross between *O. sativa* and *O. longistaminata* as exceptionally difficult; typically, one- to two-week old embryos or ovules were cultured in-vitro to rescue interspecific progeny from the abortive endosperm (Ramos, 2016; Dayun & Sripichitt, 2000). The current study identified 16 early generation

interspecific progeny (>15% admixed) from Côte D'Ivoire, Mali, Nigeria, Tanzania, Mozambique, Malawi, and Zambia out of 369 individuals with *O. longistaminata* ancestry (Table 2.1); this discovery rate of 4.3% was nearly double the ~2.5% maximum reported from controlled crosses thus far (Kaushal & Ravi, 1998) with the exception of the ~49% success rate of Chu & Oka (1969). The present study cannot eliminate the possibility that the recent hybridization events which led to the observed interspecific progeny occurred during seed increases of *O. longistaminata* accessions in the IRRI screenhouse, but even if so, the rate of hybrid recovery was exceptional compared to controlled crosses.

Concurrent with the trends observed here, Kilewa (2014) reported that in Tanzania the spontaneous production of F_1 interspecific hybrids from *O. sativa* fields where wild *O. longistaminata* grew sympatrically averaged 7.3% for *O. longistaminata*/*O. sativa* crosses and 2.6% for *O. sativa*/*O. longistaminata* crosses. Taken together, the data on the frequency of spontaneous interspecific progeny suggest that greater genetic variation for interspecific crossability exists between *O. sativa* and *O. longistaminata* than had been reported previously for controlled crosses. The greater frequency of interspecific progeny backcrossed to *O. longistaminata* than to *O. sativa* observed in this study could be explained by observational bias, because interspecific individuals with greater ancestry from *O. sativa* would be less likely to exhibit phenotypes typical of *O. longistaminata*, and therefore explorers or germplasm curators may have been less likely to sample or retain them. However, it is also possible that backcrosses to *O. sativa* are typically less fit (e.g. exhibit hybrid breakdown) than backcrosses *to O. longistaminata*. Chu & Oka (1969) found that *O. longistaminata*/*O. sativa* produced ~16-fold fewer seeds than the reciprocal cross, *O. sativa*/*O. longistaminata*, and that the fertility of the F_1 *O. longistaminata*/*O. sativa* progeny was 79% higher than that of the F_1 *O. sativa*/*O. longistaminata* progeny.

Introgression of genes from domesticated Asian rice into wild populations of the African *O. longistaminata* may be considered genetic pollution. Data from our study indicates that interspecific hybridization and introgression in northwestern Africa is predominantly recent, whereas in southern Africa both recent and ancient introgressions have occurred. In northwest Africa *O. glaberrima* was domesticated from the wild *O. barthii* ~3,000 years ago (Linares, 2002). Though Asian domesticated rice, *O. sativa*, was likely introduced into West Africa by Europeans as early as the mid-16th century, extensive production of Asian rice in West Africa and its replacement of African rice cultivars primarily occurred in the second half of the 20th century (Linares, 2002). Thus, opportunities for hybridization between *O. sativa* and *O. longistaminata* in West Africa may have been limited until recently. However,

the distribution of *O. glaberrima* was limited to West Africa. Thus, in southeastern Africa, the introduction of *O. sativa* directly from Asia ~1,000 years ago (Nayar, 2014) would not have faced competition for cropping space from a pre-existing African domesticated rice. If these historical differences explain why ancient introgressions of *O. sativa* genes into wild *O. longistaminata* are common in populations from southern Africa but not from West Africa, then in the future the West African populations of *O. longistaminata* will likely accumulate a greater proportion of genes from *O. sativa* than currently observed.

Given that obtaining *O. sativa*/*O. longistaminata* F_1 progeny from controlled crosses has required considerable effort (Causse & Ghesquiere, 1991; Khush et al., 1991; Dayun & Sripichitt, 2000; Ramos, personal communication, 2017; Ramos et al., 2016; Kaushal & Ravi, 1998), the 16 early-generation interspecific progeny identified in the current study are a valuable resource for breeding improved cultivars of domesticated Asian rice. It should also be possible to mine additional genotypes with *O. sativa* introgressions from the *O. longistaminata* germplasm collection. Furthermore, introgression efficiency from controlled crosses can be improved by screening *O. longistaminata* and *O. sativa* germplasm—including close relatives of the observed interspecific hybrids—for interspecific crossability to identify individuals with relatively high general crossability to the other species. Given the high genetic and geographic diversity of *O. longistaminata*, along with its demonstrated biotic stress tolerances (Soriano et al., 1999; Du, 2008; Panigrahi & Rajamani, 2008; Thottappilly & Rossel, 1993; Rakotomalala, 2001; Gupta, 2004), likely abiotic stress tolerances (Liu et al., 2004; Giulani et al., 2013; Atwell et al., 2014), and other known traits of agronomic value (Marathi et al., 2015; Gichuhi et al., 2016; Reinhold-Hurek et al., 2015; Ramos et al., 2016; Brar, 2004), it is probable that selective introgression of *O. longistaminata* genes into *O. sativa* will result in cultivars with great value to farmers.

Figures and Tables

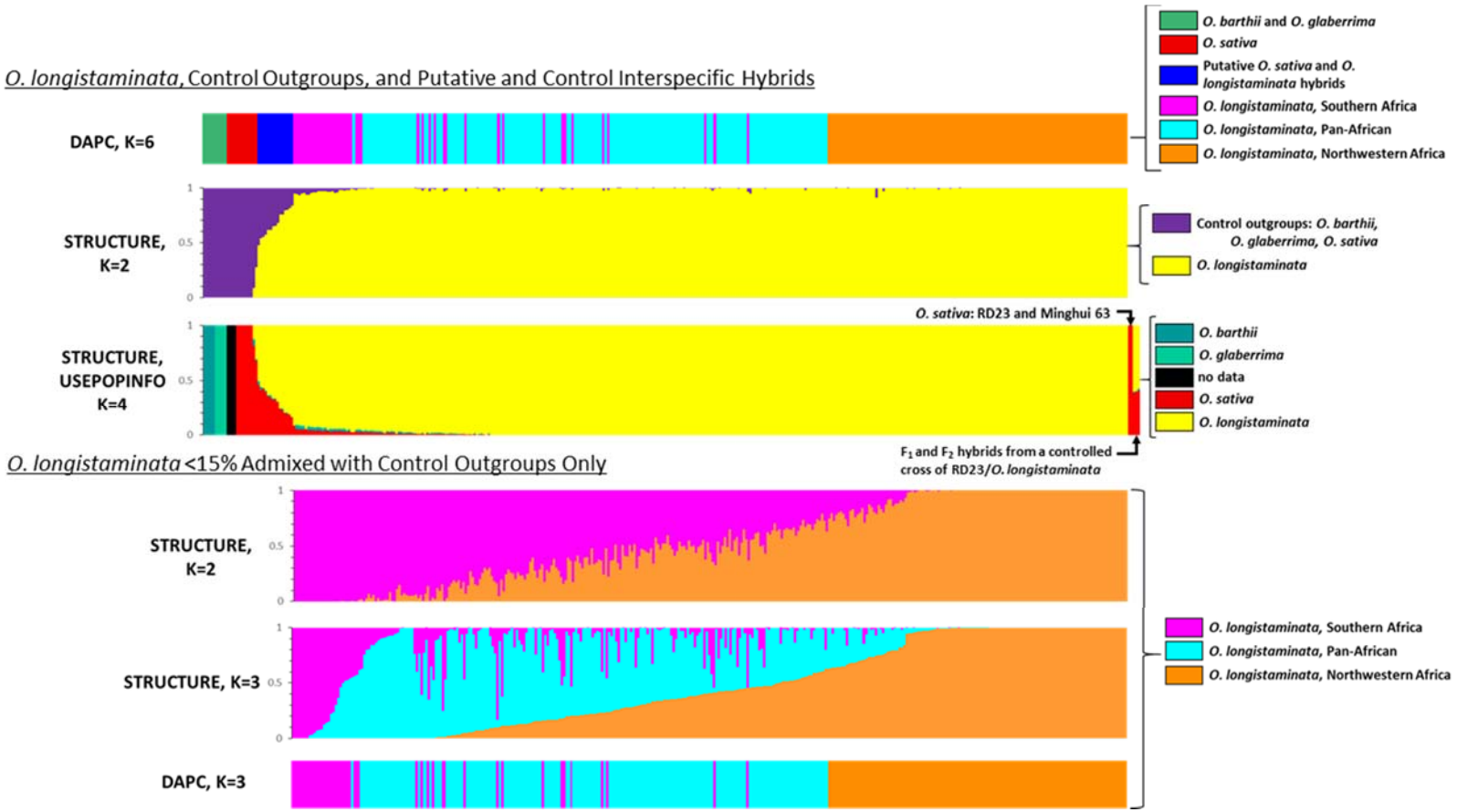
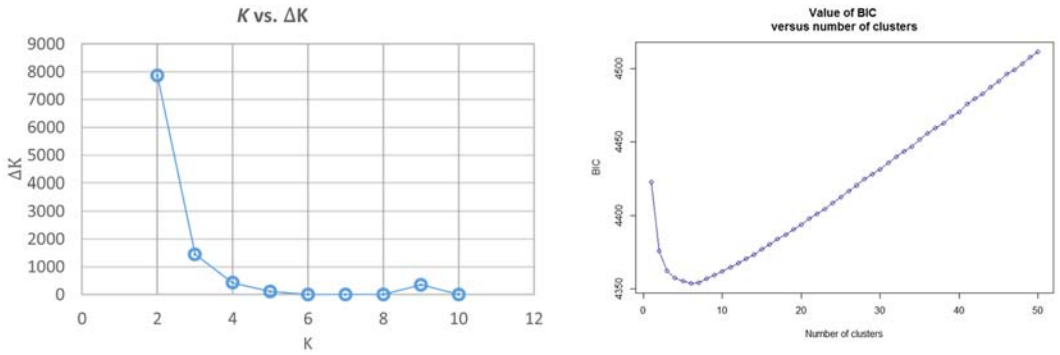


Fig. 2.1 Results for STRUCTURE, DAPC and STRUCTURE with USEPOPINFO for analysis of 397 *Oryza* spp. individuals including *O. longistaminata* (n_{ind} = 370), *O. sativa* (n_{ind} = 13), *O. barthii* (n_{ind} = 7) and *O. glaberrima* (n_{ind} = 4), and known *O. sativa*/*O. longistaminata* hybrids derived from a controlled cross (n_{ind} = 3). Individuals are ordered in the same position on the x-axis for all bar charts; Q-values for STRUCTURE were averaged across three replications, and Q-values for STRUCTURE with the USEPOPINFO option on were averaged across two replications. STRUCTURE and DAPC were first run for all *O. longistaminata* individuals from IRRI germplasm (n_{ind} = 369), *O. sativa* (n_{ind} = 11), *O. barthii* (n_{ind} = 7), and *O. glaberrima* (n_{ind} = 4) using 86,535 SNPs. DAPC showed optimal K = 6: *O. barthii* and *O. glaberrima*, *O. sativa*, putative hybrids of *O. sativa* and *O. longistaminata*, and three groups of *O. longistaminata* corresponding to Southern Africa, Pan-Africa, and Northwestern Africa. Choosing K = 7 for DAPC instead further split the *O. sativa* group into the *indica* and *japonica* subspecies. STRUCTURE results indicated optimal K = 2: one

Fig. 2.1 (cont.) population was comprised of the control outgroups and the other was comprised of *O. longistaminata*, with some individuals admixed between the groups. Upon discovery of putative interspecific hybrids between *O. sativa* and *O. longistaminata*, which clustered in DAPC and had greater than 15% ancestry from the control outgroups, STRUCTURE was run with the USEPOPINFO and PFROMPOPFLAG only options on using 85,064 SNPs in 394 *Oryza* spp. individuals. Prior populations were defined as *O. longistaminata* with no interspecific admixture (n_{ind} = 269), *O. sativa* (n_{ind} = 9), *O. barthii* (n_{ind} = 5), *O. glaberrima* (n_{ind} = 5) and the model was used to estimate ancestry for *O. longistaminata* with interspecific admixture (n_{ind} =103) and known interspecific F_1 and F_2 hybrids from a controlled *O. sativa*/*O. longistaminata* cross (n_{ind} = 3). The known interspecific hybrid F_1 and F_2 controls were correctly identified as hybrids of *O. sativa* and *O. longistaminata,* and interspecific hybrid individuals of *O. sativa* and *O. longistaminata* were also identified in IRRI germplasm. STRUCTURE and DAPC were then run using 75,371 SNPs from only the 351 *O. longistaminata* individuals less than 15% admixed with the control outgroups. In DAPC, K=3 was chosen as optimal; the three *O. longistaminata* groups were again defined geographically as Northwestern Africa, pan-Africa, and Southern Africa. The ΔK plot maximum from STRUCTURE results was at K = 2; individuals with no intraspecific admixture were concentrated in northwest and southern Africa, and most individuals were admixed between the groups roughly according to latitude. At K=3 in STRUCTURE, the intraspecifically admixed individuals corresponded to the Pan-African group and the Northwestern and Southern group were identified; at K = 3, group membership of individuals was consistent in both DAPC and STRUCTURE.

O. longistaminata, control outgroups, and putative interspecific hybrids



O. longistaminata <15% admixed with control outgroups only

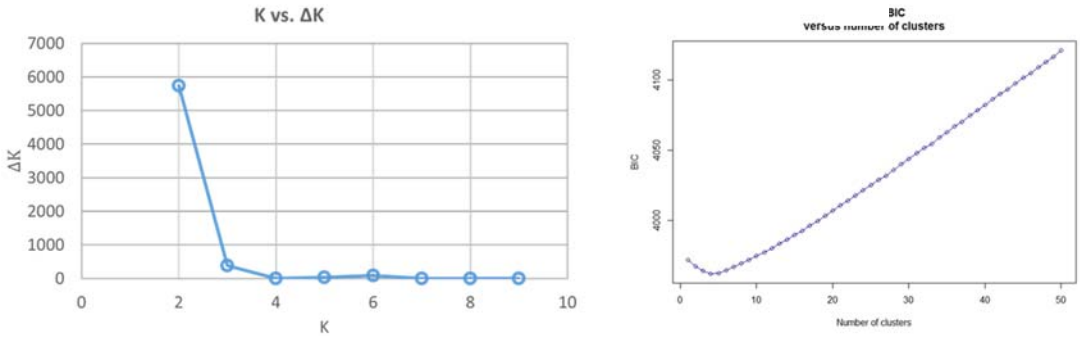


Fig. 2.2 Plots of K vs. ΔK from STRUCTURE and the Bayesian Information Criterion (BIC) curve from DAPC. STRUCTURE analysis of *Oryza longistaminata*, putative interspecific hybrids, and the control outgroups (*O. sativa*, *O. glaberrima*, and *O. barthii*) indicated optimal K=2 by the Evanno method, while the BIC minimum from DAPC indicated optimal K=6. STRUCTURE analysis of *O. longistaminata* <15% admixed with the control groups only indicated optimal K=2 by the Evanno method; while the BIC minimum from DAPC indicated K=4, K = 3 was chosen as optimal because the one cluster contained only a single individual from the southern subpopulation, probably because said individual was >5% admixed with the control outgroups. More stringent admixture cutoffs were used in later analyses to further investigate the relationship of individuals' group assignment and degree of admixture with the control outgroup.

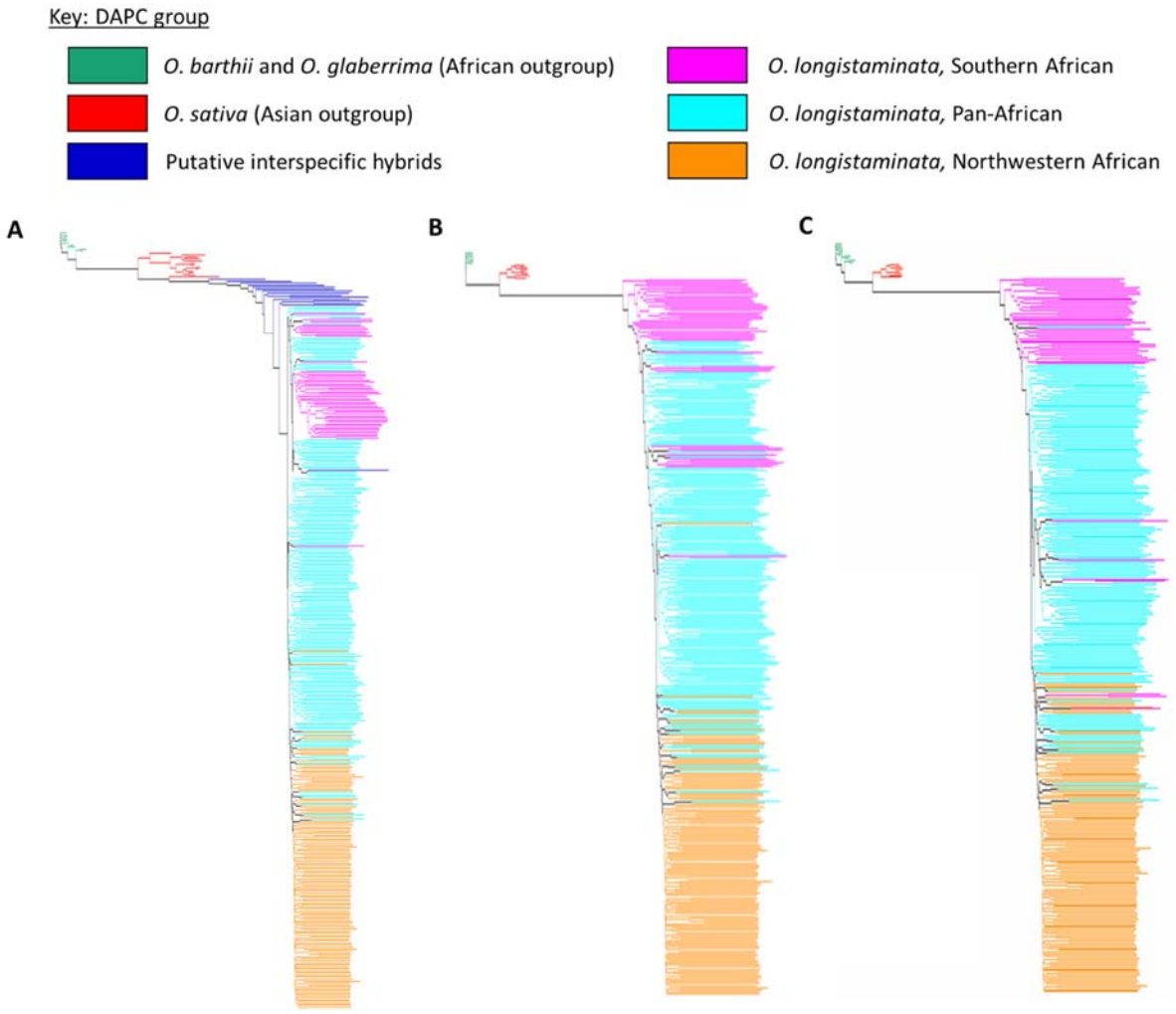


Fig. 2.3 Neighbor-joining trees. Edges are colored according to DAPC groups resulting from the same dataset used to make the tree. A) All *Oryza longistaminata* from IRRI germplasm (n_{ind} = 369), *O. sativa* (n_{ind} = 11), *O. barthii* (n_{ind} = 7), and *O. glaberrima* (n_{ind} = 4) using 86,535 SNPs. Putative interspecific hybrids cluster between *O. sativa* and *O. longistaminata*. B) Only *O. longistaminata* <15% admixed with the control outgroups, *O. sativa, O. glaberrima*, and *O. barthii* (n_{ind} = 351; n_{SNPs} = 75,371). C) Only *O. longistaminata* <5% admixed with the control outgroups (n_{ind} = 344; n_{SNPs} = 74,793). Some individuals are reassigned from the Pan-African group to the Southern African group when admixture cutoff is decreased from 15% to 5%. Group membership is stable if the admixture cutoff is further decreased to 1% (not shown).

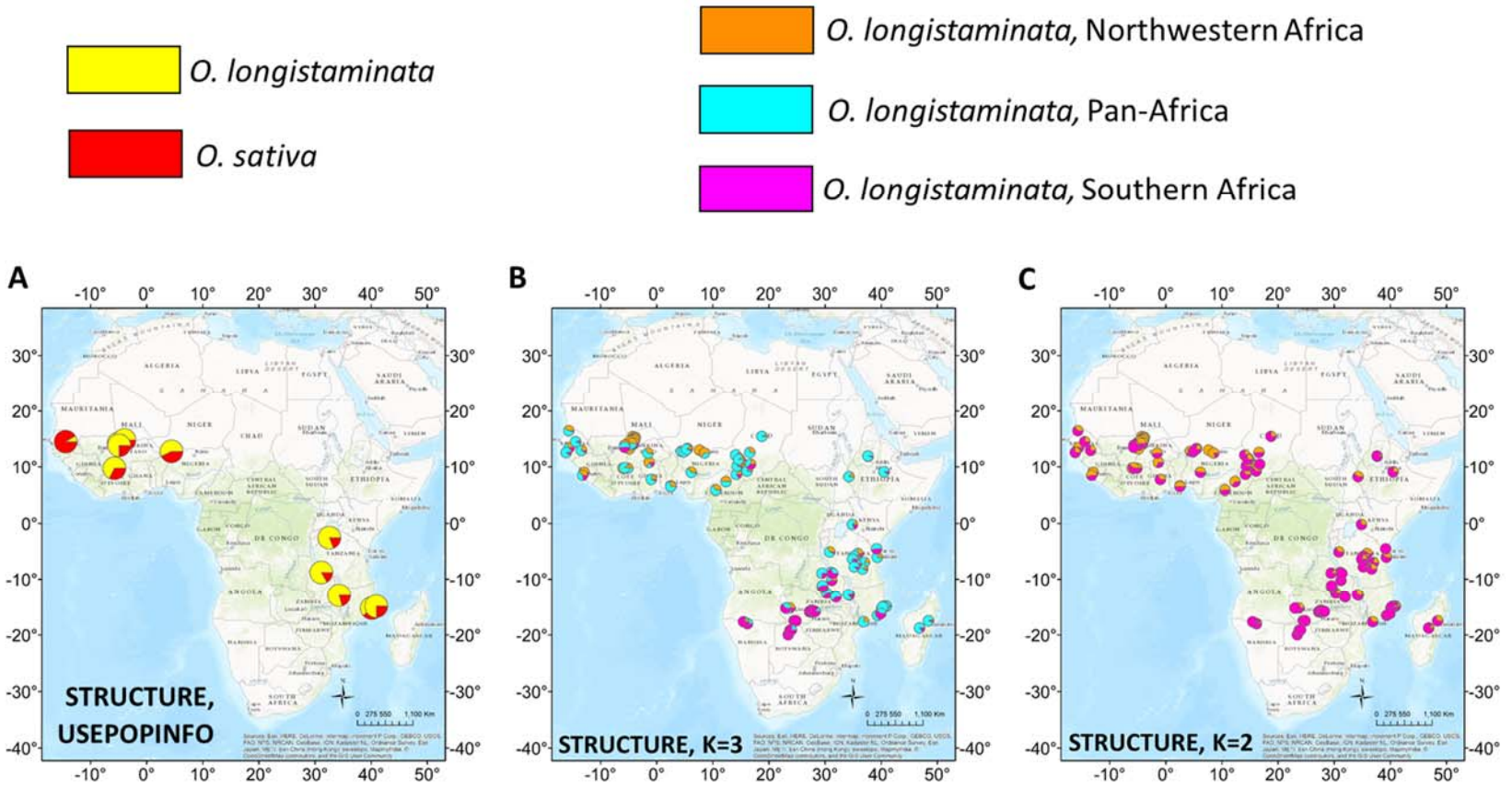


Fig. 2.4 Maps of accession origin locations and STRUCTURE results for *Oryza longistaminata* (World Topo Map, ESRI/DeLorme/HERE/USGS/ Intermap/iPC/NRCAN/ESRI Japan/METI/ESRI China (Hong Kong)/ESRI (Thailand)/MapmyIndia/TomTom). Ancestry of each accession is represented by a pie chart showing average Q for each genetic group for a given STRUCTURE run. A) Individuals with >15% *O. sativa* ancestry as determined by STRUCTURE with the USEPOPINFO option on. B) Ancestry of *O. longistaminata* individuals <15% admixed with the control outgroups, *O. sativa*, *O. barthii*, and *O. glaberrima* at K = 3. C) Ancestry of *O. longistaminata* individuals <15% admixed with the control outgroups, *O. sativa*, *O. barthii*, and *O. glaberrima* at K = 2.

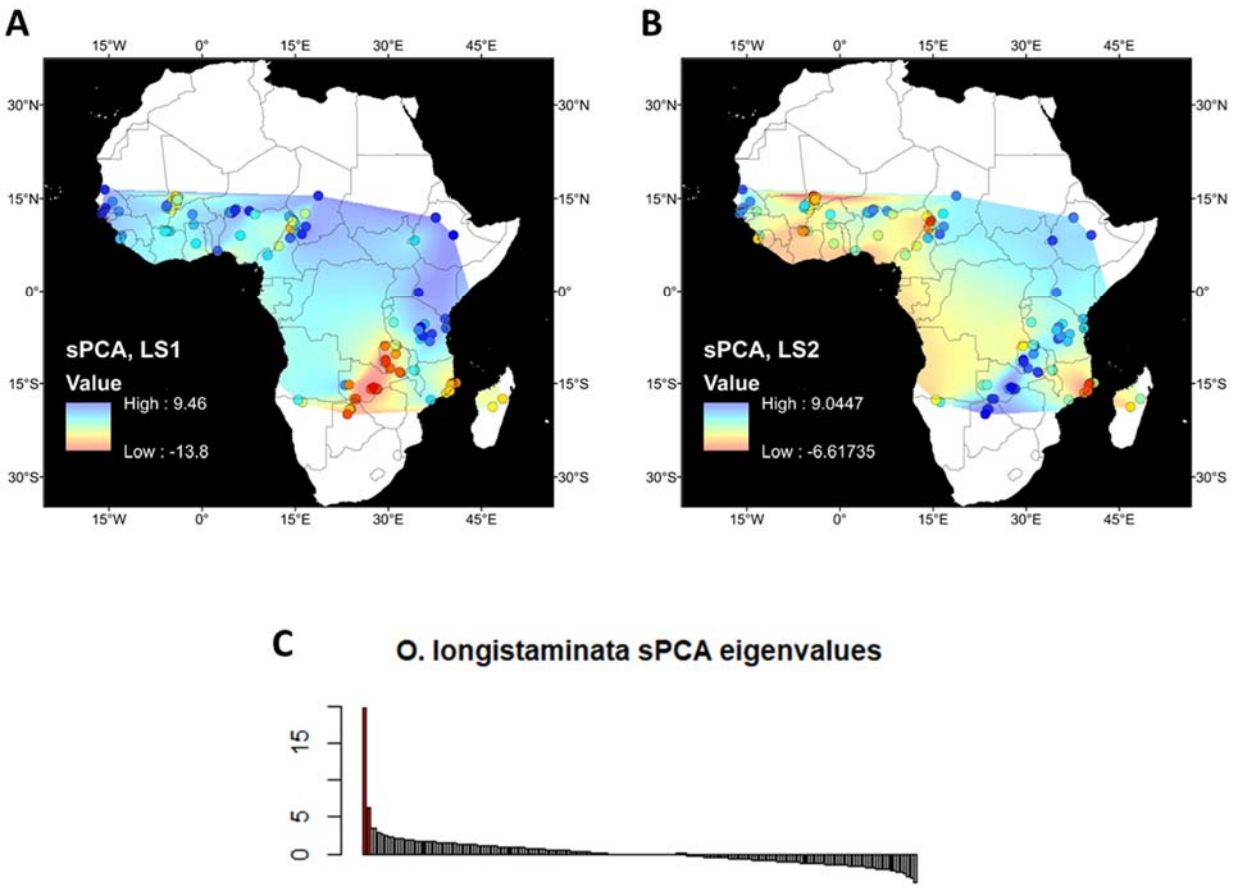


Fig. 2.5 Results of spatial principal components analysis (sPCA) for 5,980 genome-wide SNPs in 351 *Oryza longistaminata* individuals that were <15% admixed with the control outgroups *O. sativa*, *O. glaberrima*, and *O. barthii*. Lagged principal scores from the two retained principal components were plotted for each collection site (circles outlined in black) and values were interpolated between sites by the natural neighbor method. A) Map of lagged principal scores for the first principal component; the southern population (red) is highly differentiated from its neighbors, and some differentiation between individuals in the east (dark blue) and all others is evident. B) Map of lagged principal scores for the second principal component; the northwestern population is more differentiated than all others, resulting in a sharper east-west cline than in A. C) Barplot of eigenvalues from spatial principal components analysis (sPCA) of 5,980 genome-wide markers. The first two global structures, highlighted in red, were retained.

T**able 2.1** Putative *Oryza sativa*/*O. longistaminata* progeny with >15% *O. sativa* ancestry and known F_1 and F_2 control interspecific hybrids derived from a controlled cross. STRUCTURE was run at K=4 with the USEPOPINFO and PFROMPFLAGONLY options and MIGRPRIOR=0. Q values are averaged over two replications. Predefined populations were *O. sativa, O. barthii*, *O. glaberrima*, and the *O. longistaminata* with no interspecific ancestry.

	Q values				
Entry	***O. sativa***	***O. longistaminata***	***O. glaberrima***	***O. barthii***	**Origin**
		Known *O. sativa*/*O. longistaminata* progeny from crosses			
Bt135	0.39*	0.60	0.01	0.00	F_1 control
Bt136	0.40*	0.59	0.00	0.00	F_2 control
Bt137	0.41*	0.57	0.01	0.01	F_2 control
		Putative hybrids: *O. longistaminata* with >15% *O. sativa* ancestry from IRRI genebank			
101741.002	0.82	0.12	0.02	0.05	Senegal
101211.001	0.68	0.30	0.01	0.01	Côte D'Ivoire
104300.002	0.46	0.50	0.01	0.03	Malawi
101222.002	0.43	0.57	0.00	0.00	Mali
110404.002	0.42	0.56	0.01	0.02	India (Mali)
105075.002	0.39	0.59	0.01	0.01	Nigeria
92650.002	0.36	0.63	0.00	0.01	Mali
89159.002	0.35	0.63	0.01	0.01	Mozambique
103886.001	0.33	0.66	0.00	0.00	Tanzania
106456.002	0.33	0.66	0.01	0.01	Mali
101211.002	0.31	0.69	0.00	0.00	Côte D'Ivoire
101222.001	0.24	0.76	0.00	0.00	Mali
89155.001	0.23	0.76	0.01	0.01	Mozambique
86480.002	0.20	0.79	0.00	0.00	Zambia
104300.001	0.19	0.80	0.00	0.01	Malawi
103886.002	0.19	0.81	0.00	0.00	Tanzania
86480.001	0.16	0.84	0.00	0.00	Zambia

*Expected percent ancestry from *O. sativa* for the known *O. sativa*/*O. longistaminata* F_1 and F_2 progeny is 50%, but the limited sampling of *O. sativa* relative to *O. longistaminata* accessions in this study likely resulted in a modest bias in calling alleles as *O. longistaminata* by STRUCTURE.

Table 2.2 Diversity statistics for *Oryza longistaminata* genetic groups identified by discriminant analysis of principal components (DAPC) based on 351 individuals that were <15% admixed with *O. sativa*. Statistics were calculated using 61,882 biallelic SNPs with depth > 7, MAF > 0.01, and site presence in at least 75% of individuals. The mean and standard error of each value is given across loci.

Group	n_{SNPs}	Number of Individuals	*D*	F_{ST}	F_{IS}
Northwestern Africa	49,014	126	0.1331 ± 0.0007	0.0202 ± 0.0001	0.2539 ± 0.0011
Pan-Africa	60,749	180	0.1527 ± 0.0006	0.0116 ± 0.0001	0.2603 ± 0.0011
Southern Africa	49,553	45	0.1617 ± 0.0007	0.0212 ± 0.0002	0.4013 ± 0.0009
Africa, total	61,882	351	0.1568 ± 0.0006		

D, diversity (expected heterozygosity); F_{ST}, subpopulation differentiation from the total population; F_{IS}, inbreeding coefficient.

Table 2.3 Pairwise Jost's *D* statistic showing differentiation between *Oryza longistaminata* genetic groups identified with discriminant analysis of principal components (DAPC) based on 351 individuals that were <15% admixed with *O. sativa* and/or *O. glaberrima*. Statistics were calculated using 61,882 biallelic SNPs with depth > 7, MAF > 0.01, and site presence in at least 75% of individuals. The mean and standard error of each value is given across loci.

Group	Northwestern Africa	Pan-Africa	Southern Africa
Northwestern Africa		0.0131 ± 0.0001	0.0352 ± 0.0003
Pan-Africa			0.0178 ± 0.0002
Southern Africa			

Literature Cited

Atwell, B. J., Wang, H., & Scafaro, A. P. (2014). Could abiotic stress tolerance in wild relatives of rice be used to improve Oryza sativa? *Plant Science*, *215–216*(February), 48–58. https://doi.org/10.1016/j.plantsci.2013.10.007

Bezançon, G; Bozza, J; Koffi, G; Second, G. (1977). *Variability of Oryza longistaminata and the Sativa Complex of Oryza in African. Ecological and Evolutive Aspects.* Bouake.

Bezançon, G., Bozza, J., Koffi, G., & Second, G. (1978). Genetic Diversity of Indigenous Rice in Africa. In G. J. Buddenhagen, I.W., Persley (Ed.), *Rice in Africa1* (pp. 7–11). Ibadou: Academic Press.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Brar, D. S. (2004). Broadening the gene pool of rice through introgression from wild species. *New York*, 2000–2003.

Causse, M. A.; Ghesquiere, A. (1991). Prospective use of Oryza longistaminata for rice breeding. In *Rice Genetics II: Proceedings of the Second International Rice Genetics Symposium* (pp. 81–89). Manila: International Rice Research Institute.

Cheng, C., Tsuchimoto, S., Ohtsubo, H., & Ohtsubo, E. (2002). Evolutionary relationships among rice species with AA genome based on SINE insertion analysis, 323–334.

Chu, Y., & Oka, H. (1970). Introgression Across Isolating Barriers in Wild and Cultivated Oryza Species. *Evolution*, *24*(2), 344–355.

Clark, L. V. (2016). pairwiseJostDnumeric function. Retrieved from https://github.com/lvclark/R_genetics_conv/blob/master/pairwiseJostDnumeric.R

Clark, L. V., Brummer, J. E., Głowacka, K., Hall, M. C., Heo, K., Peng, J., … Sacks, E. J. (2014). A footprint of past climate change on the diversity and population structure of Miscanthus sinensis. *Annals of Botany*, *114*(1), 97–107. https://doi.org/10.1093/aob/mcu084

Du, M. (2008). *Transfer of Resistance to Yellow Stem Borer (Scirpophaga incertalus Walker) from Wild Species into Rice (Oryza sativa L.) and Molecular Characterization of Introgression*. University of the Phillipines Los Banos.

Duan, S., Lu, B., Li, Z., Tong, J., Kong, J., Yao, W., … Zhu, Y. (2007). Phylogenetic analysis of AA-genome Oryza species (Poaceae) based on chloroplast, mitochondrial, and nuclear DNA sequences. *Biochemical Genetics*, *45*(1–2), 113–129. https://doi.org/10.1007/s10528-006-9062-x

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567–1587. https://doi.org/10.1111/j.1471-8286.2007.01758.x

Fulton, T. M., Chunwongse, J., & Tanksley, S. D. (1995). Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter*, *13*(3), 207–209.

https://doi.org/10.1007/BF02670897

Ghesquière, A. (1986). Evolution of Oryza longistaminata. In *Rice Genetics: Proceedings of the International Rice Genetics Symposium* (pp. 15–25). Manila: International Rice Research Institute. https://doi.org/10.1017/CBO9781107415324.004

Gichuhi, E., Himi, E., Takahashi, H., & Maekawa, M. (2016). Characterization and QTL Analysis of Oryza longistaminata Introgression Line, pLIA-1, derived from a Cross between Oryza longistaminata and Oryza sativa (Taichung 65) under Non-fertilized Conditions. *Rice Research: Open Access*, *4*(4). https://doi.org/10.4172/2375-4338.1000174

Giuliani, R., Koteyeva, N., Voznesenskaya, E., Evans, M. A., Cousins, A. B., & Edwards, G. E. (2013). Coordination of Leaf Photosynthesis, Transpiration, and Structural Traits in Rice and Wild Relatives (Genus Oryza). *Plant Physiology*, *162*(3), 1632–1651. https://doi.org/10.1104/pp.113.217497

Green, J., & El-Moghraby, A. (2009). Swamps of the Upper White Nile. In H. Dumont (Ed.), *The Nile: Origin, Environments, Limnology and Human Use* (pp. 193–204). Springer Science and Business Media. https://doi.org/10.1007/978-1-4020-9726-3_10

Gupta, A. (2004). *WIPO-UNEP Study on the Role of Intellectual Property Rights in the Sharing of Benefits Arising from the Use of Biological Resources and Associated Traditional Knowledge*.

Han, M. V, & Zmasek, C. M. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, *10*(1), 356. https://doi.org/10.1186/1471-2105-10-356

Hu, F. Y., Tao, D. Y., Sacks, E., Fu, B. Y., Xu, P., Li, J., … Li, Z.-K. (2003). Convergent evolution of perenniality in rice and sorghum. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 4050–4054. https://doi.org/10.1073/pnas.0630531100

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jombart, T., Devillard, S., Balloux, F., Falush, D., Stephens, M., Pritchard, J., … Nei, M. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, *11*(1), 94. https://doi.org/10.1186/1471-2156-11-94

Kanya, J. I. (2010). *Gene Flow Between Cultivated Rice (Oryza sativa) and the Wild Rice (Oryza longistaminata), and its Potential Ecological Consequences in Kenya*. University of Nairobi.

Kaushal, P., & Ravi. (1998). Crossability of wild species of Oryza with O. sativa cvs PR 106 and Pusa Basmati 1 for transfer of bacterial leaf blight resistance through interspecific hybridization. *The Journal of Agricultural Science*, *130*(4), 423–430. https://doi.org/10.1017/S0021859698005383

Khush, G., Bacalangco, E., & Ogawa, T. (1991). A New Gene for Resistance to Bacterial Blight from O. longistaminata. *Rice Genetics Newsletter*, *7*.

Kiambi, D. K., Ford-Lloyd, B. V, Jackson, M. T., Guarino, L., N., M., & Newbury, H. J. (2005). Collection of wild rice (Oryza L.) in east and southern Africa in responsed to genetic erosion. *PGR Newsletter*, *142*(142), 10–20.

Kiambi, D., Newbury, H., Ford-Lloyd, B., & Dawson, I. (2005). Contrasting genetic diversity among Oryza longistaminata (A. Chev et Roehr) populations from different geographic origins using AFLP. *African Journal of Biotechnology*, *4*(4), 308–317. Retrieved from

http://www.academicjournals.org/ajb/pdf/Pdf2005/Apr/Kiambi et al.pdf

Kilewa, R. A. (2014). *Distribution of wild rice species and hybridization between cultivated rice and Oryza longistaminata in Tanzania*. University of Adelaide.

Kleynhans, M. T., James, C. S., & Birkhead, A. L. (2007). Hydrologic and hydraulic modelling of the Nyl River floodplain Part 3: Applications to assess ecological impact. *Water SA*, *33*(1), 21–25. https://doi.org/http://dx.doi.org/10.4314/wsa.v33i1.47867

Kwon, S.-J., Park, K.-C., Kim, J.-H., Lee, J. K., & Kim, N.-S. (2005). Rim 2/Hipa CACTA transposon display: a new genetic marker technique in Oryza species. *BMC Genetics*, *6*, 15. https://doi.org/10.1186/1471-2156-6-15

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Linares, O. F. (2002). African rice (Oryza glaberrima): History and future potential. *Proceedings of the National Academy of Sciences*, *99*(25), 16360–16365. https://doi.org/10.1073/pnas.252604599

Liu, L., Lafitte, R., & Guan, D. (2004). Wild Oryza species as potential sources of drought-adaptive traits. *Euphytica*, *138*(2), 149–161. https://doi.org/10.1023/B:EUPH.0000046801.27042.14

Maley, J. (1996). The African rain forest - main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. In R. Alexander, I., Swaine, M., Watling (Ed.), *Essays on the Ecology of the Guinea-Congo Rainforest* (pp. 31–73). Proceedings of the Royal Society of Edinburgh.

Marathi, B., Ramos, J., Hechanova, S. L., Oane, R. H., & Jena, K. K. (2015). SNP genotyping and characterization of pistil traits revealing a distinct phylogenetic relationship among the species of Oryza. *Euphytica*, *201*(1), 131–148. https://doi.org/10.1007/s10681-014-1213-2

Melaku, G., Haileselassie, T., Feyissa, T., & Kiboi, S. (2013). Genetic diversity of the African wild rice (Oryza longistaminata Chev. et Roehr) from Ethiopia as revealed by SSR markers. *Genetic Resources and Crop Evolution*, *60*(3), 1047–1056. https://doi.org/10.1007/s10722-012-9900-0

Mohanty, S., Wassmann, R., Nelson, A., Moya, P., & Jagadish, S. V. K. (2013). *Rice and climate change: significance for food security and vulnerability* (No. 23). Copenhagen. Retrieved from https://ccafs.cgiar.org/publications/rice-and-climate-change-significance-food-security-and-vulnerability

Nayar, N. M. (2014). The Origin of African Rice. In *Origin and Phylogeny of Rices* (pp. 117–168). Elsevier. https://doi.org/10.1016/B978-0-12-417177-0.00005-X

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, *47*(3), 253–259. https://doi.org/10.1111/j.1469-1809.1983.tb00993.x

Panigrahi, D., Rajamani, S. (2008). Genetic Evaluation of the Wild Oryza Species for Resistance Against the Yellow Stem Borer, Scirophaga Incertulas Wlk. *Journal of Plant Protection and the Environment*, *5*(1), 26–29.

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. *Theoretical and Applied Genetics*, *107*(2), 203–209.

https://doi.org/10.1007/s00122-003-1252-x

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE*, *7*(2). https://doi.org/10.1371/journal.pone.0032253

Popescu, A. A., Huber, K. T., & Paradis, E. (2012). Ape 3.0: New tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics*, *28*(11), 1536–1537. https://doi.org/10.1093/bioinformatics/bts184

Rakotomalala, M. (2001). Case History on Development and Deployment of Resistant Varieties to Control RYMV in Madagascar. In B. Hughes, J.; Odu (Ed.), *Plant Virology in Sub-Saharan Africa* (pp. 423–341). Ibadan: International Institute of Tropical Agriculture.

Ramos, J. M., Furuta, T., Uehara, K., Chihiro, N., Angeles-Shim, R. B., Shim, J., … Jena, K. K. (2016). Development of chromosome segment substitution lines (CSSLs) of Oryza longistaminata A. Chev. & R??hr in the background of the elite japonica rice cultivar, Taichung 65 and their evaluation for yield traits. *Euphytica*, *210*(2), 151–163. https://doi.org/10.1007/s10681-016-1685-3

Reinhold-Hurek, B., Hurek, T., Hu, L., Wang, Q., & Yang, H. (2014). Transgenic plants for nitrogen fixation.

Ren, F., Lu, 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. *TAG Theoretical and Applied Genetics*, *108*(1), 113–120. https://doi.org/10.1007/s00122-003-1414-x

Ronald, P. C., Albano, B., Tabien, R., Abenes, L., Wu, K. S., McCouch, S., & Tanksley, S. D. (1992). Genetic and physical analysis of the rice bacterial blight disease resistance locus, Xa21. *Molecular & General Genetics : MGG*, *236*(1), 113–120. https://doi.org/10.1007/BF00279649

Sanni, K. A., Tia, D. D., Ojo, D. K., Ogunbayo, A. S., Sikirou, M., & Hamilton, N. R. S. (2013). Diversity of Rice and Related Wild Species in Africa, 87–94.

Seck, P. A., Diagne, A., Mohanty, S., & Wopereis, M. C. S. (2012). Crops that feed the world 7: Rice. *Food Security*, *4*(1), 7–24. https://doi.org/10.1007/s12571-012-0168-1

Soriano, I. R., Schmit, V., Brar, D. S., Prot, J.-C., & Reversat, G. (1999). Resistance to rice root-knot nematode Meloidogyne graminicola identified in Oryza longistaminata and O. glaberrima. *Nematology*, *1*(4), 395–398. https://doi.org/10.1163/156854199508397

Tao, D., & Sripichitt, P. (2000). Preliminary Report on Transfer Traits of Vegetative Propagation from Wild Rice Species to Oryza sativa via Distant Hybridization and Embryo Rescue. *Kasetsart Journal: Natural Science*, *34*(1), 117–124.

Thottappilly, G., & Rossel, H. (1993). Evaluation of Resistance to Rice Yellow Mottle Virus in Oryza species. *Indian Journal of Virology*, *9*(1), 65–73.

Vaughan, D. A. (1994). *The Wild Relatives of Rice: A Genetic Resources Handbook* (1st ed.). Manila: International Rice Research Institute.

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. *Scientific Reports*, *5*(1), 13957. https://doi.org/10.1038/srep13957

Wambugu, P. W., Furtado, A., Waters, D. Le, Nyamongo, D. O., & Henry, R. J. (2013). Conservation and

utilization of African Oryza genetic resources. *Rice (New York, N.Y.)*, *6*(1), 29. https://doi.org/10.1186/1939-8433-6-29

Yin, H., Akimoto, M., Kaewcheenchai, R., Sotowa, M., Ishii, T., & Ishikawa, R. (2015). Inconsistent diversities between nuclear and plastid genomes of AA genome species in the genus *Oryza*. *Genes & Genetic Systems*, *advpub*, 269–281. https://doi.org/10.1266/ggs.14-00063

Zhang, S., Hu, J., Yang, C., Liu, H., Yang, F., Zhou, J., … Hu, F. (2017). Genotype by environment interactions for grain yield of perennial rice derivatives (Oryza sativa L./Oryza longistaminata) in southern China and Laos. *Field Crops Research*, *207*, 62–70. https://doi.org/10.1016/j.fcr.2017.03.007

Zhu, T., Xu, P. Z., Liu, J. P., Peng, S., Mo, X. C., & Gao, L. 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. *Molecular Phylogenetics and Evolution*, *70*(1), 348–361. https://doi.org/10.1016/j.ympev.2013.10.008

CHAPTER 3: SPECIES DISTRIBUTION MODEL AND GAP ANALYSIS OF THE UNDOMESTICATED RICE RELATIVE *ORYZA LONGISTAMINATA*

Abstract

Ensuring *ex-situ* conservation of crop wild relatives and their availability for basic research and applied breeding is fundamental to maintaining genetic resources used to improve domesticated species. The undomesticated rice relative *Oryza longistaminata* is highly diverged from domesticated Asian rice, *Oryza sativa*, and its untapped genetic diversity has potential to greatly improve *O. sativa*. This study reports the use of a species distribution model and the Focused Identification of Germplasm Strategy (FIGS) to identify collection gaps in *O. longistaminata* as well as to predict locations of germplasm that are likely tolerant to agronomically-relevant abiotic stresses. Though collection of *O. longistaminata* has a strong foundation, gaps were identified in the species range, including locations where stress-tolerant germplasm is likely present. Closing the collection gaps and increasing the geographic range of available germplasm for *O. longistaminata* will provide a valuable resource for future efforts to breed improved *O. sativa*.

Introduction

Genebanks provide the invaluable services of conserving crop genetic resources and making them available for basic research and applied breeding (Khoury et al., 2010; Li & Pritchard, 2009). Accessions of domesticated crop landraces and populations of crop wild relatives conserved by genebanks are increasingly the only source of these fundamental gene sources. In farmer's fields, traditional cultivars have been widely replaced by modern high-yielding varieties (Coromaldi et al., 2015; Pascual & Perrings, 2006), and the available habitat for many crop wild relatives has been substantially reduced by human use of land and water resources (Brummitt et al., 2015; Castañeda-Álvarez et al., 2016). Germplasm from genebanks has been essential in improving crop yield potential and increasing abiotic and biotic stress tolerance (Maxted et al., 2016). Crop wild relatives have distinct evolutionary histories that do not include the genetic bottlenecks typically associated with domestication. Because crop wild relatives contain many useful but unexploited genes absent from their domesticated relatives, they are expected to contribute to future gains in crop productivity (Maxted et al., 2016; Vaughan, 1994).

Genebanks, however, face the dual conflicting challenges of 1) maximizing the genetic diversity conserved for a given crop or species, while 2) providing plant breeders with objective criteria for

choosing a subset of accessions that are likely to have genes of interest for a specific trait from a collection that is too large to phenotype economically in its entirety. Species distribution models, in combination with the Focused Identification of Germplasm Strategy (FIGS), offer solutions to these challenges. Species distribution models can be used to characterize the full range of environments a species likely occupies and to identify spatial collection gaps (Castañeda-Álvarez et al., 2016), while FIGS uses environmental information at collection sites to identify germplasm that is more likely to be adapted to specific abiotic or biotic stresses than a random sample, helping to efficiently direct plant breeders to useful material in large germplasm collections (Mackay & Street, 2004). Empirical studies have shown FIGS to effectively predict tolerance for both biotic and abiotic stresses (Khazaei et al., 2013; Bari et al., 2012). By combining FIGS with a species distribution model, it is further possible to identify previously unsampled sites where valuable stress-tolerant germplasm is likely present.

Asian rice (*Oryza sativa*) currently supplies one-fifth of global calories consumed by people (Awika, 2011). During domestication, *O. sativa* underwent a genetic bottleneck and is now 80-90% less diverse than its most recent wild ancestor, *O. rufipogon*, which is endemic to tropical Asia and much of Oceania (Zhu et al., 2007; Vaughan, 1994). Of the seven wild relatives of *O. sativa* which comprise its primary genepool (the AA-genome clade), *O. longistaminata* is one of the most genetically distinct from *O. sativa*, along with the Australian species *O. meridionalis* (Vaughan et al., 2008; Wambugu et al., 2015; Zhu et al., 2014). *O. longistaminata* is also the most broadly distributed *Oryza* species in Africa; it is endemic to much of sub-Saharan Africa including Madagascar, where it is commonly found in riparian environments with full sunlight (Bezançon et al., 1977; Bezançon et al., 1978; Vaughan, 1994). The other two AA-genome species indigenous to Africa, *O. glaberrima* (domesticated African rice) and *O. barthii* (its undomesticated progenitor) occupy a lesser geographic range (Vaughan, 1994). High genetic diversity in *O. longistaminata* is likely the result of large effective population size, self-incompatibility, perennial rhizomatous growth, and adaptation to diverse environments (Bezançon et al., 1977; Bezançon et al., 1978; Ghesquiere, 1986, Vaughan, 1994). Ecologically, *O. longistaminata* is a keystone species in seasonally flooded riparian grasslands and provides habitat for birds and insects as well as fodder for large ungulates (Scholte, 2000; Dörgeloh, 1999; Arbeiter & Tegetmeyer, 2011). In recent years, *O. longistaminata* populations have been under pressure from conversion of wetlands to agricultural use and overgrazing (Kiambi et al., 2005; Kleynhans et al., 2007).

Though there is a strong breeding barrier between *O. sativa* and *O. longistaminata*, *O. longistaminata* has made important contributions to improvement of *O. sativa*. *O. longistaminata* was

the source of the *Xa21* gene conferring resistance to bacterial blight, a major disease of domesticated rice (Ronald et al., 1992; Tu et al., 1998; Wang et al., 1996). Recently, *O. longistaminata* has also been used as the source of genes for improved ratooning ability and increased ratoon yields (Tao et al., 2001; Zhang et al., 2017). Additionally, *O. longistaminata* introgressions in *O. sativa* have been shown to confer increased yield (Brar, 2004; Gichuhi et al., 2016; Gichuhi et al., 2016; Ramos et al., 2016). *O. longistaminata* is resistant to multiple biotic stresses, including root knot nematodes (*Meloidogyne graminicola*; Soriano et al., 1999), yellow stem borer (Du, 2008; Panigrahi & Rajamani, 2008), and rice yellow mottle virus (Thottappilly & Rossel, 1993; Rakotomalala, 2001) and rice blast (*Magnaporthe oryzae*; He et al., 2014). *O. longistaminata* is also predicted to be drought-tolerant (Liu et al., 2004; Giulani et al., 2013), heat-tolerant (Atwell et al., 2014), and to have high nitrogen-use efficiency (Gichuhi et al., 2016; Reinhold-Hurek et al., 2015).As a self-incompatible species, *O. longistaminata* has floral traits not present in the other AA-genome species, such as long anthers and long exserted stigmas, which could be advantageous in economical production of hybrid rice (Marathi et al., 2015).

Though the geographic distribution of *O. longistaminata* has largely been discovered via occurrence records and the collection of germplasm and herbaria specimens, and some locations of potential interest to plant breeders have been noted (e.g. the salt lagoons of the Casamance delta; Bezançon et al., 1977), there has been no systematic study to determine which environmental variables best predict occurrence of this species throughout it entire sampled range in Africa. Nor has a comparison of available germplasm with predicted occurrence to identify collection gaps been conducted. Additionally, locations of all prior or potential future collections likely to have *O. longistaminata* populations with tolerance to key abiotic stresses (i.e., drought, flooding, excess salinity, soil pH extremes, heat, and cold) have not been identified. This study was conducted to address these gaps in knowledge.

Materials and Methods

To facilitate conservation and utilization of *O. longistaminata*, collection gaps were identified using a species distribution model. Maxent version 3.4.1 (Phillips et al., 2017) at default settings was used to predict the species distribution. ArcGIS version 10.3.1 (ESRI) was used to prepare the data for input into Maxent and project the output raster. Environmental variables for precipitation and temperature (BIO1-BIO19) were downloaded at a resolution of 2.5 minutes (~20 km^2) from the WorldClim 1.4 BIOCLIM set for current conditions (~1960-1990; Table 3.1; Hijmans et al., 2005). Soil data were obtained from the Harmonized World Soil Database (HWSD, version 1.2; FAO/IIASA/ISSCAS/JRC,

2012) and resampled from 30 arc-seconds to 2.5 minutes. All variables used the GCS World Geodetic System 1984 geographic coordinate system, had a cell size of 0.041666667 decimal degrees, and were clipped to the extent of Africa. To avoid model instability due to multicollinearity, Pearson's correlation coefficient (*r*) was allowed to range from -0.70 to +0.70 between any two variables (Table 3.2), yielding a final set of 11 environmental variables that were input into the model used to predict *O. longistaminata* occurrence (Table 3.1).

Occurrence records were consolidated from the Crop Wild Relatives Global Occurrence Database, a metadatabase including records from herbaria, major genebanks, and published literature (Global Crop Diversity Trust). Additional records were taken from published literature (Vaughan, 1994; Bezançon et al., 1977). In total, 486 occurrence geographic sites were analyzed. To correct sampling bias and avoid model overfitting, the sites were filtered to a minimum distance of 20 km using the SDMToolbox version 1.1c which resulted in a final set of 296 occurrence sites that were input into the model (Brown, 2014; Fourcade et al., 2014; Kramer-Schadt et al., 2013).

Conservative and liberal presence/absence thresholds were defined for the Maxent predicted distribution. The conservative threshold was determined by maximizing training sensitivity plus specificity, and the liberal threshold was obtained by balancing training omission, predicted areas, and threshold value (Liu, 2016). For both presence/absence thresholds, collection gaps were identified, and FIGS was used to determine which accessions and potential collection sites were likely to contain germplasm tolerant to key abiotic stresses. Seven abiotic stresses that are important constraints on rice production were studied: drought, submergence, high salinity, high and low soil pH, and high and low temperature, as represented by annual precipitation (BIO12), mean precipitation of the wettest quarter (BIO16), topsoil electrical conductivity (t_ece), topsoil pH (topsoil_pH), mean temperature of the warmest quarter (BIO10), and mean temperature of the coldest quarter (BIO11). Threshold values defining stress were determined as those considered stressful for domesticated rice (*O. sativa*) in the literature for high salinity, high and low soil pH, and low temperature (Table 3.3; Gregorio et al., 1999; Li et al., 2016; Zhu et al., 2007). For drought, submergence, and high temperature, the top or bottom 10% of values in the variable range were defined as stressful (Table 3.3). For each variable, the mean, standard deviation, and range was calculated across the predicted presence ranges established by both liberal and conservation thresholds (Table 3.4). Additionally, statistics for these variables were calculated for the entire set of 486 occurrence sites of *O. longistaminata* and only the occurrence sites for germplasm in the IRRI genebank (i.e. available germplasm for breeding; Table 3.4).

Results

Species distribution model and gap analysis

The species distribution model for *O. longistaminata* predicted broad presence across much of sub-Saharan Africa as expected. Presence was predicted for a total area of was 4,850,154 km^2 at the conservative threshold and 13,035,353 km^2 at the liberal threshold (Fig. 3.1A). The distribution of *O. longistaminata* is constrained by the Sahara Desert to the north, the arid lowlands in the Horn of Africa to the east, and the Kalahari and Namib deserts to the south. Notably, the model predicted a gap in *O. longistaminata* presence in the central Congo basin.

Both conservative and liberal presence thresholds for the model indicated that *O. longistaminata* inhabits more geographic area than was sampled during previous collecting expeditions (Fig. 3.1B). Under the conservative presence threshold, *O. longistaminata* was predicted to occur throughout most of Madagascar and inland Mozambique, much of west Central Africa (including Angola, the Democratic Republic of the Congo, Gabon, Cameroon, and Equatorial Guinea), along the entire Niger River and delta, and along the upper Nile River in Sudan and South Sudan. *O. longistaminata* may also occur farther inland in the Congo Delta than previous collections indicate, despite the overall lack of sampling in the region. The liberal presence threshold further indicated that *O. longistaminata* also likely inhabits broad swathes of central and northeast sub-Saharan Africa, and its distribution may extend farther into southwest Africa (i.e. South Africa, Botswana, Zambia) than previous occurrence records show.

Though the publicly available *O. longistaminata* germplasm in the IRRI genebank was sampled from large areas of the species' range, the distribution model showed that there are likely collection gaps—large geographic areas where the species is likely present but from which seed has not been collected and conserved for future use (Fig. 3.1C). Most notable was the absence of any samples from parts of central Africa, northeast Africa, southwest and southeast Africa, and southern Madagascar. Additionally, the IRRI genebank *O. longistaminata* may not represent the full range of environments that the species encounters (Table 3.4). The observed range of Mean Temperature of the Warmest Quarter is 20.2—33.2 °C for IRRI germplasm, but the range increases to 17.4—33.5 °C for all occurrence records; at the liberal presence threshold, the range is 11.8—34.9 °C (Table 3.4). The observed range of Precipitation of the Driest Month also increases greatly between the IRRI germplasm (0—61 mm) and the liberal presence threshold (0—151 mm; Table 3.4). For Mean Temperature of the Wettest Quarter,

the IRRI germplasm range is 18.6—29.4 °C, while the range using the liberal presence threshold is 6.6—32.4 °C; even at the conservative presence threshold, the range is 11.1—30.6 °C, a large increase relative to the IRRI germplasm range (Table 3.4). The range in Annual Mean Temperature widens from 18.8—29.3 °C in the IRRI germplasm to 15.2—29.4 °C for all occurrence records, and further to 7.3—30.7 °C at the liberal presence threshold (Table 3.4).

The species distribution model had a training AUC of 0.89 and a regularized training gain of 0.91. Jackknifing showed that the variable that conferred the highest regularized training gain if used alone in the model was precipitation of the wettest month (BIO13), and its exclusion from the model also caused the greatest reduction in model performance (followed closely by temperature seasonality; Fig. 3.2). Precipitation of the wettest month was collinear with three excluded variables: precipitation of the wettest quarter (BIO16), annual precipitation (BIO12), and precipitation of the warmest quarter (BIO17).

FIGS for abiotic stress tolerance and associated collection gaps

For all occurrence records, sites likely to contain stress tolerant germplasm were identified for all seven stresses studied (drought, submergence, high salinity, high and low soil pH, and high and low temperature; Fig. 3.3, Table 3.5). For the subset of IRRI germplasm, accessions likely to be stress-tolerant were identified for six of the seven stresses studied; however, for low temperature no IRRI germplasm met the threshold of 15 °C average temperature in the coldest quarter although occurrence records did (Fig. 3.3). In all cases, the range of values of the predictor variable of the abiotic stress was greater for the set of all occurrence records than the subset of available IRRI germplasm, suggesting that greater stress tolerance is available in germplasm at sites from which germplasm is not publicly available.

Discussion

Potential sources of O. longistaminata germplasm for rice improvement in the IRRI genebank

The distribution model for *O. longistaminata* indicates that past efforts to collect specimens and seeds over the entire species range have been mostly successful. Given that the range spans an entire continent, this has been a significant achievement. The great diversity of environments in which *O. longistaminata* occurs suggest that many useful genes from this species are likely present in the IRRI genebank collection. This study identified accessions and potential collection areas that are likely sources of genes for tolerance to abiotic stresses for all stresses studied (Fig. 3.3; Table 3.5). Sites predicted to contain drought-tolerant germplasm were abundant throughout the species distribution,

but the most extreme values were in northern Mali and Botswana. Flood-tolerant germplasm was predicted along the northwestern coast of Africa. Germplasm tolerant of salt stress was predicted to occur in the Casamance delta of Senegal as noted by Bezançon (1977) and in Namibia. Germplasm tolerant of acidic soils was concentrated in the Congo Basin, for which no IRRI germplasm is available, and germplasm tolerant of alkaline soils was found in Namibia. No IRRI germplasm predicted to be cold-tolerant was identified though occurrence records in South Africa showed where likely cold-tolerant *O. longistaminata* grows. However, the overall prediction accuracy is limited due to undetected variation in the microenvironmental habitat of *O. longistaminata* and the univariate approach used. The predictive model also does not account for cryptic genetic variation in *O. longistaminata* that may confer abiotic stress tolerance in an *O. sativa* background, nor unique physiological mechanisms in *O. longistaminata* (e.g. rhizomatousness) that may play an important role in adaptation to abiotic stress tolerance but are not present in *O. sativa.*

Precipitation is a primary determinant of O. longistaminata *distribution*

Precipitation appears to be a key factor in determining *O. longistaminata* distribution on a continental scale given that precipitation of the wettest month (BIO13) conferred the highest model gain if used alone and reduced the model gain most if excluded from the model (Fig. 3.2). Extreme values for precipitation of the wettest month were obtained in desert and rainforest regions where the model did not predict *O. longistaminata* occurrence. Because these regions comprise a large geographic area, can be characterized by rainfall levels, and had no *O. longistaminata* occurrence records, model prediction of *O. longistaminata* absence from these regions based on precipitation resulted in high training gain. In eastern Africa, Kiambi et al. (2008) found that annual precipitation levels showed a curvilinear relationship with genetic diversity (as measured by 176 AFLP markers); maximum genetic diversity occurred at moderate precipitation levels, with declines in genetic diversity at the extremes in rainfall values.

Collection gaps and recommended conservation priorities

Gaps exist in the IRRI genebank collection of *O. longistaminata*, which is the largest and most accessible collection of this species. Most of these gaps exist at the geographic edges of the projected species range, as in northeastern, southeastern, and southwestern Africa. Though edge populations are sometimes less genetically diverse than more central populations due to lower effective population size (Gao & Gao, 2016; Pandey & Rajora, 2012; Tollefsrud et al., 2009; Yakimowski & Eckert, 2008), drift in

edge populations may allow increases in frequency of alleles that are deleterious in other parts of the species range (Alleaume-Benharira et al., 2006), and sampling of edge populations may increase capture of these alleles. Edge populations may also inhabit environments not usually encountered by the species and carry unique traits that have evolved as adaptations to these environments (Bridle & Vines, 2006). In this study, populations predicted to be stress-tolerant sometimes occurred in areas where no IRRI germplasm was available, and thus are a sampling priority. Populations in southeastern Africa were predicted likely to be cold-tolerant, and populations in central Africa were predicted to be tolerant of acidic soils.

Confirming *O. longistaminata* at the spatial extremes of the distribution predicted by the model would further understanding of where this species occurs. Notably, no germplasm is currently available from much of central Africa with the exceptions of Chad and Cameroon to the north, and just two samples from Botswana in the south (Fig. 3.1C). *O. longistaminata* may be absent from much of Central Africa and represent another edge of the *O. longistaminata* species distribution, but it is also possible that prior sampling of the Congo Basin (a difficult-to-access region) is insufficient to draw negative conclusions (Vaughan, personal communication, 2016). Because much of the Congo Basin represents a unique combination of environmental variables in the species distribution model not found elsewhere in Africa, a sampling gap in the region would lead to exclusion of this environment from the model. Though the Congolese rainforests of the Congo Basin are densely forested and perhaps lack the sunlight *O. longistaminata* likely requires in abundance (Bezançon et al., 1977), the rainforests are interrupted by grassy meadows near rivers; in these meadows, *Echinochloa* spp. grow (Sautte & Pourtier, 2014) and these are commonly sympatric with *O. longistaminata* in other flooded grasslands (Scholte et al., 2007). This suggests *O. longistaminata* may grow in the Congo Basin as well, though it may be rarer in central Africa than in other regions. Thus, the absence of *O. longistaminata* from the Congo Basin is inconclusive and may be a sampling artifact and confirmation is needed. Additionally, confirmation of occurrence would be desirable for southern Africa (i.e. Angola, Mozambique, Madagascar, and South Africa), and northeastern Africa (i.e. Ethiopia, the Sudan, and South Sudan).

O. longistaminata habitat has been under increasing pressure in recent years from human activities, including dam construction and draining of wetlands and their repurposing (Kiambi et al., 2005). Because *O. longistaminata* is a weed of domesticated rice in Africa, efforts to eradicate it near farmlands are common (Rodenburg & Johnson, 2009). Given these anthropogenic pressures, reduction in the available wild stands of *O. longistaminata* is a risk. Although *ex-situ* conservation of this species

has a strong foundation, further sampling is needed to capture the full genetic diversity of *O. longistaminata* and the valuable traits—including abiotic stress tolerances—that it holds.

Figures and Tables

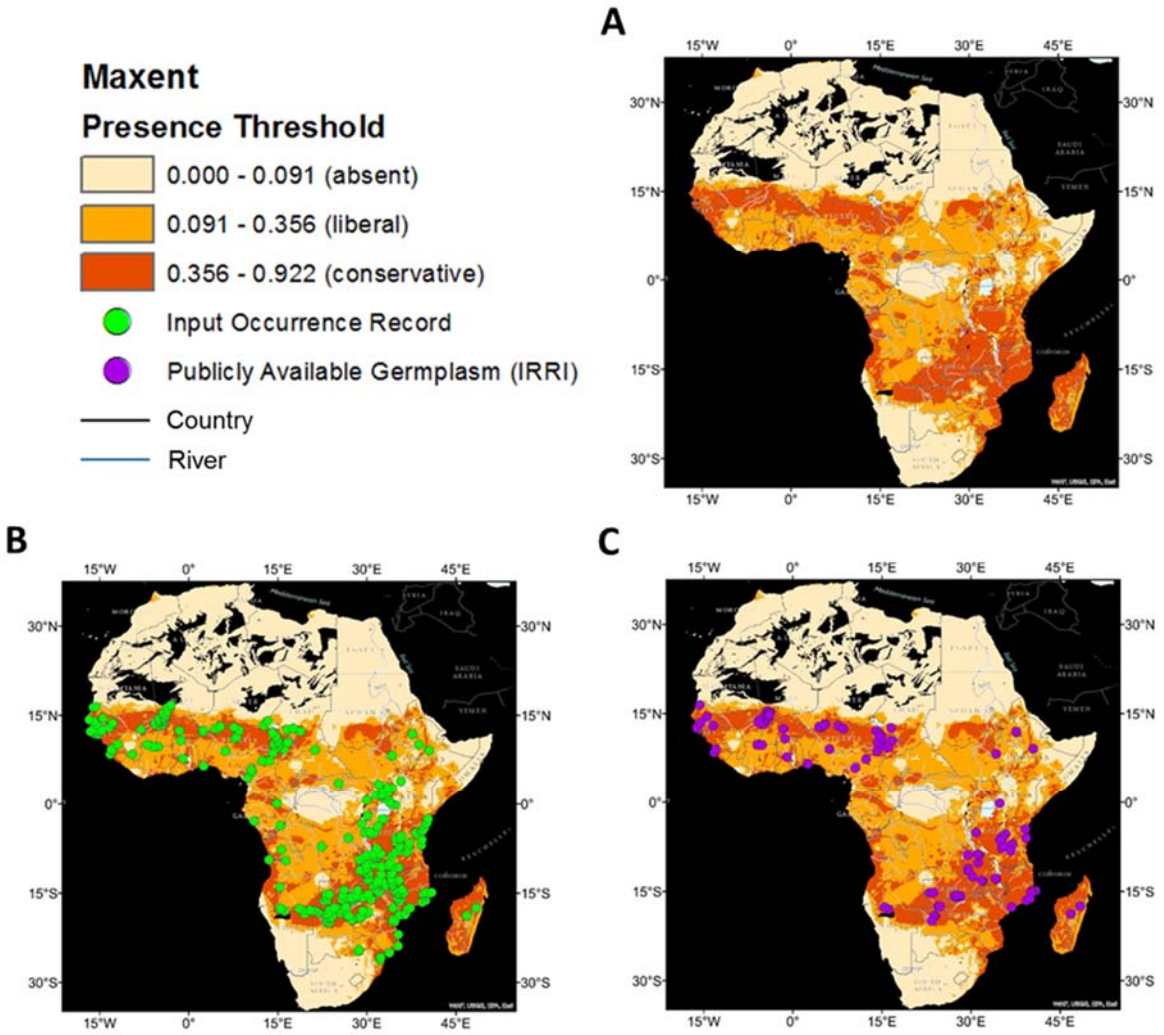


Fig. 3.1 Maxent species distribution model for *Oryza longistaminata*. A liberal presence threshold was established by balancing training omission, predicted area, and threshold values (light orange). A conservative threshold was established by maximizing training sensitivity plus specificity (dark orange). A) The predicted distribution of *O. longistaminata*. The species' range was limited by the Sahara Desert to the north, the arid lowlands in the Horn of Africa to the east, and the Kalahari and Namib Deserts to the south. There was an apparent absence from part of the Congo Basin that may have been due to the environmental conditions of the Congolese rainforests or lack of sampling of this environment (which has uniquely high extreme precipitation and low topsoil pH values) leading to underrepresentation of the environment in the model. B) The species distribution model overlaid with the occurrence records used to generate it. Novel sites where *O. longistaminata* is predicted to be present included Sudan and South Sudan to the northeast, inland Mozambique and southern Madagascar to the south, and the Congo Basin and much of Angola (especially the coastal region) in central Africa. C) Publicly available *O. longistaminata* germplasm in the IRRI genebank. Little to no material from central, southeast, and northeast Africa is available despite occurrence records.

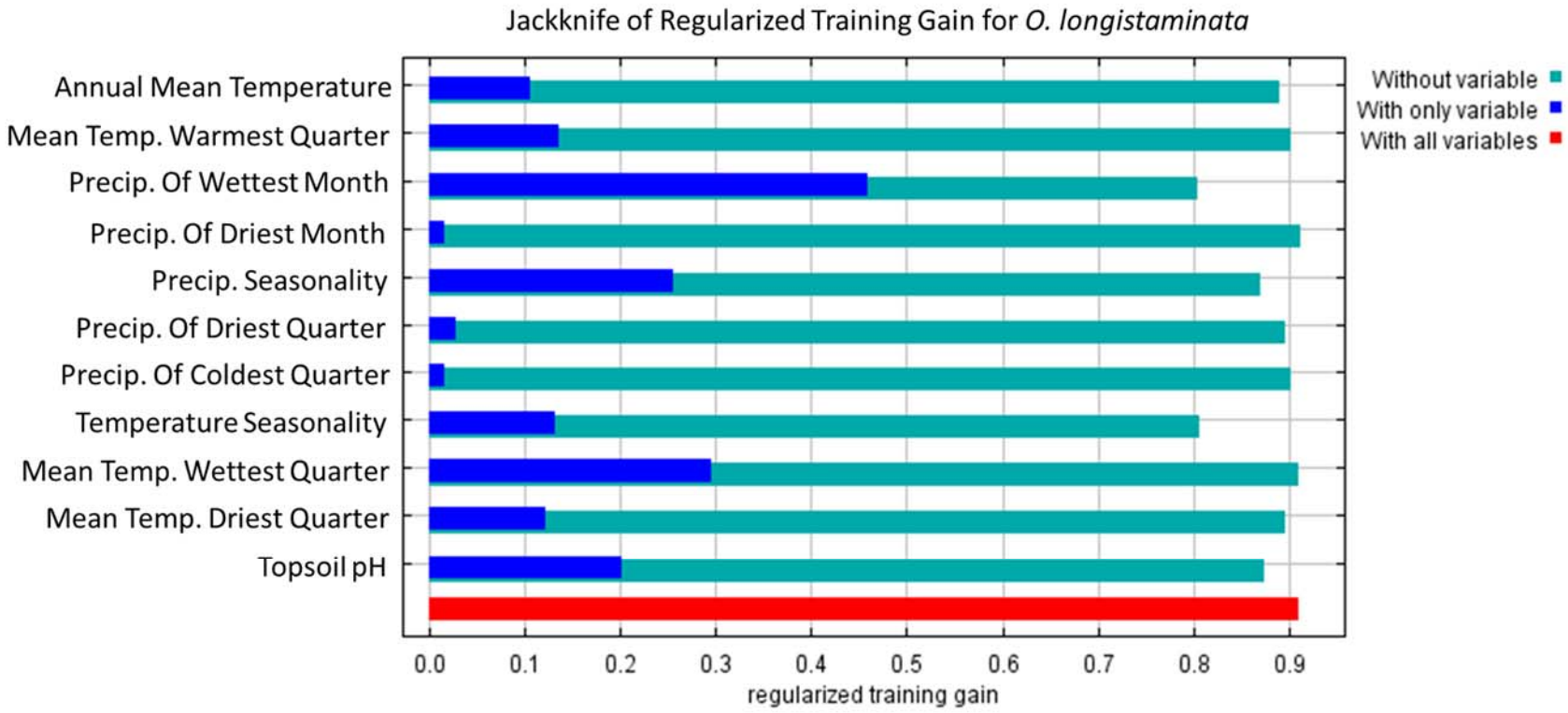


Fig. 3.2 Jackknife of regularized training gain for *Oryza longistaminata* showing change in regularized model training gain in area if each variable is excluded (teal), used alone (blue), or included with all others (red). Precipitation of the wettest month (BIO13) had the highest gain when used alone, and its exclusion caused the most reduction in model performance followed closely by temperature seasonality. Precipitation of the wettest month had extreme values in desert and rainforest regions, which were most of the area that the model predicted *O. longistaminata* is unlikely to occur.

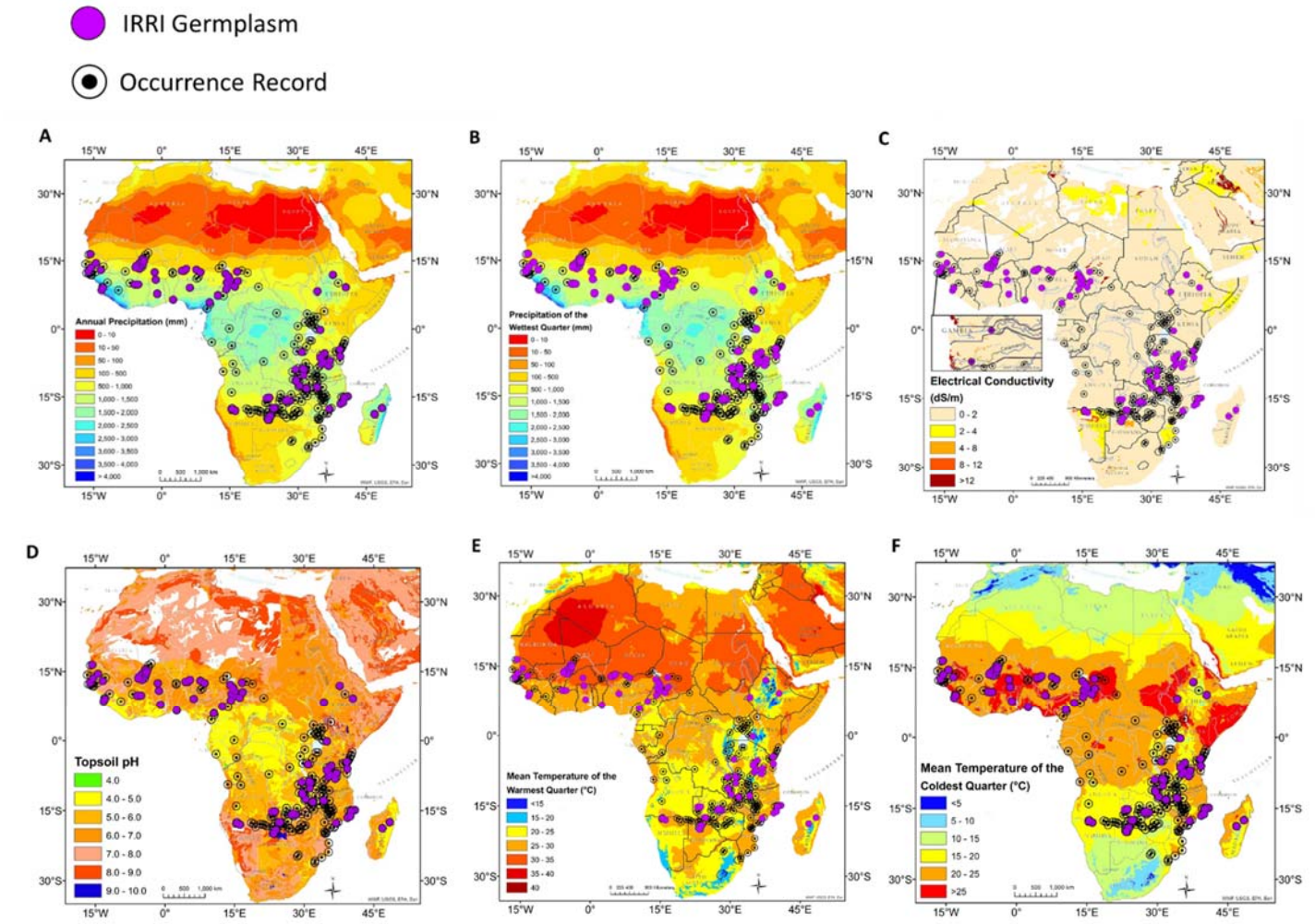


Fig. 3.3 Maps of Africa for environmental predictors of abiotic stress tolerance in *Oryza longistaminata* germplasm. *O. longistaminata* occurrence is indicated by black bullseyes, and sites of IRRI germplasm are highlighted by purple circles. A) Annual precipitation, the predictor of drought tolerance. Occurrence sites and IRRI germplasm with values less than ~500 mm were considered likely to be drought-tolerant and were

Fig. 3.3 (cont.) common throughout sub-Saharan Africa excepting the Congo Basin and Madagascar. Accessions near deserts in northern Mali and northern Namibia and Botswana had most extreme values for this variable. B) Precipitation of the wettest quarter, the predictor of flood tolerance. Occurrence sites and IRRI germplasm with values greater than ~2,000 mm were considered likely to be flood-tolerant and were common along the northwest coast of sub-Saharan Africa. C) Electrical conductivity of topsoil, the predictor of salt tolerance. Occurrence sites and IRRI germplasm with values greater than 12 dS/m were considered likely to be salt-tolerant and were observed in the Casamance delta of Senegal, as described by Bezançon (1970); sites with likely salt-tolerant germplasm were also found in northern Namibia and coastal Mozambique. D) Topsoil pH, the predictor of tolerance to pH extremes. Occurrence sites and IRRI germplasm with values less than 5.5 were considered likely to be tolerant to acid soils and were observed throughout the Congo Basin, near the southern coast of northwest Africa, and in Madagascar. Occurrence sites and IRRI germplasm with values greater than 8.5 were considered likely to be tolerant to alkaline soils and were observed in Namibia and Botswana. E) Mean temperature of the warmest quarter, the predictor of heat tolerance. Occurrence sites and IRRI germplasm with values greater than ~32 °C were considered likely to be heat-tolerant and were common in northwest Africa. F) Mean temperature of the coldest quarter, the predictor of cold tolerance. Occurrence sites and IRRI germplasm with values less than 15°C were considered likely to be stress-tolerant, and occurrence sites where stress-tolerant germplasm is likely present was observed in South Africa.

Table 3.1 Predictor variables from the WorldClim BIOCLIM set and the Harmonized World Soil Database evaluated in this study; variables input to Maxent to generate the species distribution model are listed first, and variables excluded due to multicollinearity with others in the selected set are listed below.

Predictor Variable	**ID**
Annual Mean Temperature	BIO1
Mean Diurnal Range	BIO2
Mean Temperature of Wettest Quarter	BIO8
Mean Temperature of Driest Quarter	BIO9
Mean Temperature of Warmest Quarter	BIO10
Precipitation of Wettest Month	BIO13
Precipitation of Driest Month	BIO14
Precipitation Seasonality	BIO15
Precipitation of Driest Quarter	BIO17
Precipitation of Coldest Quarter	BIO19
Topsoil pH	T_PH_H20
*Excluded due to multicollinearity**	
Isothermality	BIO3
Temperature Seasonality	BIO4
Max. Temperature of Warmest Month	BIO5
Min. Temperature of Coldest Month	BIO6
Temperature Annual Range	BIO7
Mean Temperature of Coldest Quarter	BIO11
Annual Precipitation	BIO12
Precipitation of Wettest Quarter	BIO16
Precipitation of Warmest Quarter	BIO18

*To reduce multicollinearity, pairwise *r* was allowed to range from -0.70 to 0.70.

Table 3.2 Pairwise *r* values for environmental predictor variables used in Maxent to predict distribution of *O. longistaminata*. To reduce multicollinearity, pairwise *r* of variables included in the model was allowed to range from -0.70 to 0.70; values outside the range are shown in italics, as are the names and IDs of predictor variables excluded from the model.

ID	BIO1	BIO10	BIO11	BIO12	BIO13	BIO14	BIO15	BIO16	BIO17	BIO18	BIO19	BIO2	BIO3	BIO4	BIO5	BIO6	BIO7	BIO8	BIO9	topsoil_ph
BIO1	—																			
BIO10	0.69	—																		
BIO11	0.62	0.20	—																	
BIO12	-0.02	-0.50	0.45	—																
BIO13	-0.07	-0.43	0.51	*0.93*	—															
BIO14	0.06	-0.29	0.20	0.58	0.35	—														
BIO15	0.45	0.33	0.29	-0.24	-0.01	-0.40	—													
BIO16	-0.03	-0.45	0.48	*0.94*	*0.99*	0.37	-0.05	—												
BIO17	-0.06	-0.32	0.23	0.64	0.40	*0.98*	-0.43	0.42	—											
BIO18	-0.14	-0.54	0.34	*0.82*	*0.74*	0.54	-0.21	*0.75*	0.58	—										
BIO19	0.08	-0.18	0.24	0.61	0.53	0.42	-0.24	0.54	0.45	0.25	—									
BIO2	0.03	0.33	-0.32	-0.56	-0.46	-0.42	0.27	-0.46	-0.46	-0.49	-0.39	—								
BIO3	-0.12	-0.45	*0.71*	*0.73*	0.67	0.51	-0.07	0.66	0.56	0.67	0.37	-0.39	—							
BIO4	-0.12	0.46	*-0.78*	*-0.73*	*-0.74*	-0.37	-0.05	*-0.73*	-0.42	-0.65	-0.34	0.50	*-0.93*	—						
BIO5	0.60	*0.96*	0.06	-0.57	-0.48	-0.37	0.33	-0.51	-0.40	-0.62	-0.23	-0.54	-0.53	-0.57	—					
BIO6	-0.57	0.14	*0.96*	0.52	0.52	0.31	0.16	0.50	0.35	0.40	0.34	-0.54	*0.72*	-0.78	0.04	—				
BIO7	0.07	0.48	*-0.70*	*-0.75*	-0.70	-0.46	0.08	-0.69	-0.52	-0.68	-0.40	*0.74*	*-0.87*	*0.94*	0.64	*-0.79*	—			
BIO8	0.57	0.37	*0.75*	0.08	0.15	0.02	0.43	0.13	0.03	0.15	-0.06	-0.04	0.39	-0.43	0.28	0.65	-0.33	—		
BIO9	-0.47	0.61	0.15	-0.13	-0.14	-0.05	0.08	-0.15	-0.04	-0.31	0.13	-0.13	-0.22	0.25	0.55	0.22	0.16	-0.10	—	
topsoil_ph	0.05	0.30	-0.29	-0.64	-0.56	-0.40	0.09	-0.58	-0.44	-0.56	-0.30	0.37	-0.47	0.46	0.36	-0.33	0.48	0.04	0.04	—

Table 3.3 Abiotic stresses studied, the environmental predictors used to predict abiotic stress tolerance, and the variable values of the environmental predictors defined as stressful.

Abiotic Stress	**Environmental Predictor**	**Variable Values Defined as Stressful**	Citation
Drought	Annual Precipitation (BIO12)	Bottom 10% of observed range; < ~500 mm	
Submergence	Precipitation of the Wettest Quarter (BIO16)	Top 10% of observed range; > ~2,000 mm	
High salinity	Topsoil Electrical Conductivity (t_ece)	>12 dS/m	Gregorio et al., 1999
Extreme soil pH	Topsoil pH (t_ph_h20)	Acid soils, <5.5; Alkaline soils, >8.5	Kochian et al., 2004; Li et al., 2016
Heat	Mean Temperature of the Warmest Quarter (BIO10)	Top 10% of observed range; ~32 °C	
Cold	Mean Temperature of the Coldest Quarter (BIO11)	<15 °C	Zhu et al., 2007

Table 3.4 Mean, standard deviation (SD), and range of the 11 environmental variables used to generate the Maxent *Oryza longistaminata* distribution model. Statistics are shown at the liberal presence threshold (0.091), which was established by balancing training omission, predicted areas, and threshold value, and at the conservative presence threshold (0.356), which was established by maximizing training sensitivity plus sensitivity. Statistics are also shown for all *O. longistaminata* occurrence records in the model, and for the IRRI germplasm subset only. Use of the liberal presence threshold does not always increase observed range of the environmental variables relative to use of the conservative threshold.

Environmental Variable	Liberal Presence Threshold	Conservative Presence Threshold	All Occurrence Records	IRRI Germplasm Only
Topsoil pH				
Range	4.0—10.6	4.0—10.6	4.3—8.9	4.4—8.9
Mean ± SD	6.0 ± 1.0	6.2 ± 0.9	6.1 ± 1.5	6.2 ± 0.8
Mean Temperature of Warmest Quarter (^{o}C)				
Range	11.8—34.9	17.1—34.1	17.4—33.5	20.2—33.2
Mean ± SD	26.4 ± 3.2	26.7 ± 3.0	26.7 ± 3.0	28.4 ± 3.4
Precipitation of Driest Quarter (mm)				
Range	0—506	0—217	0—217	0—217
Mean ± SD	30 ± 48	17 ± 35	21 ± 38	10 ± 27
Mean Precipitation of Coldest Quarter (mm)				
Range	0 — 2975	0 — 2975	0—2546	0—2546
Mean ± SD	170 ± 287	79 ± 233	117 ± 17	168 ± 29
Precipitation Seasonality				
Range	20 — 187	26 — 183	26—173	39 — 150
Mean ± SD	93 ± 28	105 ± 23	103 ± 2	117 ± 2
Precipitation of the Driest Month (mm)				
Range	0 — 151	0 — 64	0—61	0 — 61
Mean ± SD	6 ± 11	3 ± 8	5 ± 1	2 ± 1
Precipitation of the Wettest Month (mm)				
Range	37 — 1157	67 — 1157	69—965	86—965
Mean ± SD	220 ± 86	227 ± 96	229 ± 0.6	230 ± 8
Mean Temperature of the Driest Quarter (^{o}C)				
Range	1.9 — 32.0	11.7 — 30.8	12.7—29.4	16.0 — 28.3
Mean ± SD	22.3 ± 3.8	21.7 ± 3.8	21.7 ± 2.0	23.4 ± 3.0
Mean Temperature of the Wettest Quarter (^{o}C)				
Range	6.6 — 32.4	11.1 — 30.6	16.4—30.8	18.6 — 29.4
Mean ± SD	24.6 ± 2.5	24.9 ± 2.1	24.8 ± 1.0	25.7 ± 1.2
Mean Diurnal Range (^{o}C)				
Range	4.0 — 18.4	4.0 — 17.1	5.9 – 16.5	5.9 — 16.5
Mean ± SD	12.7 ± 2.2	12.9 ± 2.1	12.8 ± 0.1	13.3 ± 0.1
Annual Mean Temperature (^{o}C)				
Range	7.3 — 30.7	14.8 —30.4	15.2—29.4	18.8 — 29.3
Mean ± SD	243 ± 28	242 ± 28	24.1 ± 0.2	256 ± 2

Table 3.5 List of IRRI germplasm records and occurrence records from the Global Crop Diversity Trust (GCDT) for which abiotic stress tolerance was predicted.

Source	Accession	Latitude	Longitude	Value of Predictor Variable
Drought; bottom 10% of range for Annual Precipitation (~500 mm)				
IRRI	101754	16.456	-15.683	227
IRRI	101763	16.456	-15.683	227
IRRI	92640	15.417	-4.283	345
IRRI	92626	15.267	-3.850	364
IRRI	92627	15.267	-3.850	364
IRRI	92644	15.133	-4.267	386
IRRI	92639	15.133	-4.250	386
IRRI	92643	15.117	-3.900	395
IRRI	92638	14.967	-4.267	409
IRRI	92607	14.917	-4.317	411
IRRI	92609	14.833	-4.350	422
IRRI	103560	14.817	-4.266	434
IRRI	92636	14.833	-3.967	448
IRRI	106456	14.817	-3.967	450
IRRI	92617	14.750	-4.200	455
IRRI	104128	12.583	16.583	456
IRRI	92637	14.767	-4.000	457
IRRI	101223	14.500	-4.833	458
IRRI	101225	14.500	-4.833	458
IRRI	104100	-19.967	23.417	459
IRRI	92645	14.750	-4.033	459
IRRI	92618	14.717	-4.183	463
IRRI	106455	14.717	-4.100	464
IRRI	81968	-19.119	23.988	469
IRRI	117261	-18.003	16.159	471

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
IRRI	86485	-19.174	23.751	472
IRRI	92612	14.450	-4.917	476
IRRI	103916	-7.300	35.533	478
IRRI	117263	-17.658	15.466	483
IRRI	92633	14.617	-4.167	488
IRRI	92647	14.517	-4.117	491
IRRI	92649	14.433	-4.283	493
IRRI	92631	14.400	-4.133	494
IRRI	92628	14.467	-4.150	494
IRRI	92634	14.483	-4.167	495
IRRI	92635	14.483	-4.167	495
IRRI	92641	14.500	-4.167	495
IRRI	92642	14.517	-4.150	495
IRRI	92619	14.567	-4.217	496
IRRI	92625	14.417	-4.200	496
IRRI	92632	14.517	-4.200	497
IRRI	92658	14.517	-4.183	497
IRRI	92624	14.350	-4.267	498
IRRI	92654	14.367	-4.200	498
IRRI	92657	14.533	-4.183	498
IRRI	92615	14.317	-4.217	499
IRRI	92655	14.317	-4.100	499
GCDT		14.317	-4.100	499
GCDT		14.533	-4.183	498
GCDT		14.350	-4.267	498
GCDT		14.567	-4.217	496
GCDT		14.500	-4.167	495
GCDT		14.483	-4.167	495
GCDT		14.517	-4.150	495

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		14.467	-4.150	494
GCDT		14.400	-4.133	494
GCDT		14.417	-4.133	494
GCDT		14.433	-4.283	493
GCDT		14.590	-4.200	493
GCDT		14.617	-4.167	488
GCDT		-18.833	24.166	484
GCDT		-17.658	15.466	483
GCDT		-17.655	15.460	483
GCDT		-17.654	15.376	477
GCDT		14.450	-4.917	476
GCDT		-7.300	35.500	475
GCDT		-19.167	23.417	474
GCDT		-17.787	15.608	473
GCDT		-19.183	23.266	472
GCDT		-19.174	23.751	472
GCDT		-18.003	16.159	471
GCDT		-18.002	16.156	471
GCDT		-19.119	23.988	469
GCDT		-19.117	23.980	469
GCDT		-19.136	23.974	469
GCDT		14.717	-4.100	464
GCDT		-19.750	23.250	459
GCDT		-19.850	23.417	459
GCDT		-19.967	23.417	459
GCDT		-19.967	23.417	459
GCDT		-20.000	23.417	458
GCDT		-19.966	23.450	458
GCDT		-17.788	15.354	458

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		-17.788	15.342	458
GCDT		14.767	-4.000	457
GCDT		-19.817	23.566	456
GCDT		14.756	-4.150	456
GCDT		12.583	16.583	456
GCDT		14.750	-4.200	455
GCDT		14.751	-4.197	455
GCDT		14.817	-3.967	450
GCDT		-19.950	23.650	448
GCDT		14.817	-4.267	434
GCDT		14.583	-4.800	432
GCDT		14.833	-4.350	422
GCDT		14.917	-4.317	411
GCDT		14.967	-4.267	409
GCDT		12.833	17.583	407
GCDT		15.117	-4.300	388
GCDT		15.133	-4.267	386
GCDT		15.267	-3.850	364
GCDT		15.417	-4.283	345
GCDT		13.310	12.600	306
GCDT		13.317	30.209	301
GCDT		15.783	-3.617	288
GCDT		3.917	35.732	265
GCDT		16.456	-15.683	227
GCDT		16.456	-15.683	227
GCDT		16.460	-15.680	227
GCDT		16.466	-15.850	223
GCDT		16.320	-3.582	221
GCDT		14.044	13.330	217

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		16.360	-16.080	217
GCDT		16.737	-2.995	180
Submergence; top 10% of range for Mean Precipitation of the Wettest Quarter (< ~2,000 mm)				
IRRI	105057	8.485	-13.237	2546
GCDT		-13.237	8.485	2546
GCDT		-13.237	8.485	2546
GCDT		-13.080	9.380	2254
GCDT		-13.080	9.430	2234
High salinity; topsoil electrical conductivity > 12 dS/m				
IRRI	117261	-18.003	16.159	248.2
IRRI	117263	-17.658	15.466	17.2
GCDT		-18.003	16.159	248.2
GCDT		-18.002	16.156	248.2
GCDT		-17.788	15.354	17.2
GCDT		-17.788	15.342	17.2
GCDT		-17.787	15.608	17.2
GCDT		-17.658	15.466	17.2
GCDT		-17.655	15.460	17.2
GCDT		-17.654	15.376	17.2
GCDT		-17.505	15.460	17.2
GCDT		-17.472	15.526	17.2
GCDT		-17.472	15.533	17.2
GCDT		-18.368	36.571	16.8
GCDT		-18.617	36.400	16.8
GCDT		-18.585	36.439	16.8
GCDT		16.360	-16.080	14.6
Acidic soils; topsoil pH < 5.5				
IRRI	101436	-17.489	48.426	4.4

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
IRRI	115136	-15.210	23.686	4.4
IRRI	101207	9.750	-5.167	4.8
IRRI	105067	9.750	-5.167	4.8
IRRI	105078	7.365	12.34333333	4.9
GCDT		-19.183	23.266	4.3
GCDT		-2.925	10.997	4.4
GCDT		-17.489	48.426	4.4
GCDT		-17.489	48.426	4.4
GCDT		-15.210	23.686	4.4
GCDT		11.860	-15.600	4.5
GCDT		7.328	13.575	4.6
GCDT		3.552	25.269	4.8
GCDT		-15.333	26.000	4.8
GCDT		10.083	-7.000	4.8
GCDT		11.780	-15.400	4.9
GCDT		12.400	-15.790	4.9
GCDT		8.470	-9.520	4.9
GCDT		-12.000	34.050	4.9
GCDT		7.365	12.343	4.9
GCDT		7.365	12.343	4.9
GCDT		-7.201	22.398	4.9
GCDT		12.300	-16.560	4.9
GCDT		1.717	33.616	4.9
GCDT		1.000	32.833	4.9
GCDT		1.317	32.467	4.9
GCDT		-9.539	16.341	4.9
GCDT		-8.483	30.100	4.9

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		0.092	14.878	4.9
GCDT		1.700	32.833	4.9
GCDT		1.468	31.540	4.9
GCDT		-5.800	34.917	4.9
GCDT		9.380	-13.080	5.0
GCDT		9.430	-13.080	5.0
GCDT		9.270	-13.020	5.0
GCDT		9.017	-12.950	5.0
GCDT		9.011	-12.952	5.0
GCDT		9.011	-12.952	5.0
GCDT		4.495	9.974	5.0
GCDT		-18.777	46.831	5.0
GCDT		-18.777	46.831	5.0
GCDT		-11.167	29.533	5.0
GCDT		-11.268	29.536	5.0
GCDT		-11.267	29.535	5.0
GCDT		-11.250	29.535	5.0
GCDT		-11.350	29.550	5.0
GCDT		-10.822	29.939	5.0
GCDT		-11.416	29.500	5.0
GCDT		-10.250	29.900	5.0
GCDT		-12.966	34.300	5.0
GCDT		-10.200	31.200	5.0
GCDT		-9.383	30.100	5.0
GCDT		-9.367	30.100	5.0
GCDT		-9.250	30.917	5.0
GCDT		-8.950	31.367	5.0

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		-8.983	31.374	5.0
GCDT		-9.166	31.400	5.0
GCDT		-10.333	33.666	5.0
GCDT		-11.166	28.867	5.0
GCDT		-12.550	30.283	5.0
GCDT		-12.793	30.110	5.0
GCDT		-12.809	30.100	5.0
GCDT		-14.333	31.500	5.0
GCDT		-14.486	28.475	5.0
GCDT		-13.666	33.266	5.0
GCDT		-17.333	26.483	5.0
GCDT		11.500	14.683	5.0
GCDT		13.696	-12.850	5.1
GCDT		-5.727	26.870	5.1
GCDT		-24.700	28.557	5.1
GCDT		-10.533	33.917	5.1
GCDT		12.937	-13.411	5.1
GCDT		12.937	-13.411	5.1
GCDT		-15.338	28.239	5.2
GCDT		-8.917	31.750	5.2
GCDT		9.800	-5.750	5.2
GCDT		9.833	-5.750	5.2
GCDT		-1.970	30.115	5.2
GCDT		-8.900	29.450	5.2
GCDT		2.460	32.670	5.3
GCDT		1.583	33.433	5.3
GCDT		-5.150	38.466	5.3

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		-4.983	31.200	5.3
GCDT		9.750	-5.667	5.3
GCDT		-7.013	35.500	5.3
GCDT		-2.550	32.550	5.4
GCDT		-0.056	34.175	5.4
GCDT		-5.100	30.833	5.4
GCDT		-3.833	31.833	5.4
GCDT		-6.368	34.885	5.4
GCDT		-6.368	34.885	5.4
GCDT		-0.055	34.192	5.4
GCDT		-12.830	34.150	5.4
GCDT		-8.800	31.500	5.4
Alkaline soils; topsoil pH > 8.5				
IRRI	104100	-19.967	23.417	8.9
IRRI	117263	-17.658	15.466	8.5
GCDT		-17.472	15.533	8.5
GCDT		-17.472	15.526	8.5
GCDT		-17.505	15.460	8.5
GCDT		-17.658	15.466	8.5
GCDT		-17.655	15.460	8.5
GCDT		-17.654	15.376	8.5
GCDT		-17.787	15.608	8.5
GCDT		-17.788	15.354	8.5
GCDT		-17.788	15.342	8.5
GCDT		-18.269	21.769	8.7
GCDT		-18.441	23.216	8.9
GCDT		-19.967	23.417	8.9

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		-19.967	23.417	8.9
GCDT		-20.000	23.417	8.9
GCDT		-19.966	23.450	8.9
GCDT		-19.950	23.650	8.9
Extreme cold; Mean Temperature of Coldest Quarter < 15				
GCDT		-10.533	33.917	12.5
GCDT		-24.803	28.490	12.7
GCDT		-24.700	28.557	12.9
GCDT		-24.508	28.725	13.5
GCDT		-19.013	29.147	14.8
GCDT		-19.220	28.752	14.8
GCDT		-19.216	28.750	14.8
Extreme heat; top 10% of Mean Temperature of Warmest Quarter (°C)				
IRRI	104128	12.583	16.583	33.2
IRRI	92628	14.467	-4.150	32.0
IRRI	92634	14.483	-4.167	32.0
IRRI	92635	14.483	-4.167	32.0
IRRI	92641	14.500	-4.167	32.0
IRRI	92642	14.517	-4.150	32.0
IRRI	92647	14.517	-4.117	32.0
IRRI	104079	13.283	5.467	32.0
IRRI	105075	12.767	4.417	32.0
IRRI	92631	14.400	-4.133	31.9
IRRI	92632	14.517	-4.200	31.9
IRRI	92657	14.533	-4.183	31.9
IRRI	92658	14.517	-4.183	31.9
IRRI	105076	12.800	4.400	31.9
IRRI	105077	13.150	5.333	31.9

Table 3.5 (cont.)

Source	Accession	Latitude	Longitude	Value of Predictor Variable
GCDT		12.833	17.583	33.5
GCDT		12.583	16.583	33.2
GCDT		16.737	-2.995	32.8
GCDT		12.333	18.833	32.6
GCDT		12.467	2.417	32.2
GCDT		12.517	2.433	32.2
GCDT		16.320	-3.582	32.2
GCDT		9.430	-13.080	32.0
GCDT		12.200	2.383	32.0
GCDT		12.767	4.417	32.0
GCDT		13.283	5.467	32.0
GCDT		14.500	-4.167	32.0
GCDT		14.483	-4.167	32.0
GCDT		14.517	-4.150	32.0
GCDT		14.467	-4.150	32.0
GCDT		9.380	-13.080	31.9
GCDT		12.800	4.400	31.9
GCDT		13.150	5.333	31.9
GCDT		14.533	-4.183	31.9
GCDT		14.400	-4.133	31.9
GCDT		14.417	-4.133	31.9

Literature Cited

Alleaume-Benharira, M., Pen, I. R., & Ronce, O. (2006). Geographical patterns of adaptation within a species' range: Interactions between drift and gene flow. *Journal of Evolutionary Biology*, *19*(1), 203–215. https://doi.org/10.1111/j.1420-9101.2005.00976.x

Arbeiter, S., & Tegetmeyer, C. (2011). Home Range and Habitat use by Aquatic Warblers Acrocephalus paludicola on their Wintering Grounds in Northwestern Senegal. *Acta Ornithologica*, *46*(2), 117–126. https://doi.org/10.3161/000164511X625883

Awika, J. M. (2011). Major Cereal Grains Production and Use around the World. In S. Awika, J., Piironen, V., Bean (Ed.), *Advances in Cereal Science: Implications to Food Processing and Health Promotion* (pp. 1–13). American Chemical Society. https://doi.org/10.1021/bk-2011-1089.ch001

Bari, A., Street, K., Mackay, M., Endresen, D. T. F., de Pauw, E., & Amri, A. (2012). Focused identification of germplasm strategy (FIGS) detects wheat stem rust resistance linked to environmental variables. *Genetic Resources and Crop Evolution*, *59*(7), 1465–1481. https://doi.org/10.1007/s10722-011-9775-5

Bezançon, G; Bozza, J; Koffi, G; Second, G. (1977). *Variability of Oryza longistaminata and the Sativa Complex of Oryza in African. Ecological and Evolutive Aspects.* Bouake.

Bezançon, G., Bozza, J., Koffi, G., & Second, G. (1978). Genetic Diversity of Indigenous Rice in Africa. In G. J. Buddenhagen, I.W., Persley (Ed.), *Rice in Africa1* (pp. 7–11). Ibadou: Academic Press.

Brar, D. S. (2004). Broadening the gene pool of rice through introgression from wild species. *New York*, 2000–2003.

Bridle, J. R., & Vines, T. H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in Ecology and Evolution*, *22*(3), 140–147. https://doi.org/10.1016/j.tree.2006.11.002

Brown, J. L. (2014). SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution*, *5*(7), 694–700. https://doi.org/10.1111/2041-210X.12200

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., … Lughadha, E. M. N. (2015). Green plants in the red: A baseline global assessment for the IUCN Sampled Red List Index for Plants. *PLoS ONE*, *10*(8), 1–22. https://doi.org/10.1371/journal.pone.0135152

Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H., Eastwood, R. J., … Toll, J. (2016). Global conservation priorities for crop wild relatives. *Nature Plants*, *2*(4), 16022. https://doi.org/10.1038/nplants.2016.22

Coromaldi, M., Pallante, G., & Savastano, S. (2015). Adoption of modern varieties, farmers' welfare and crop biodiversity: Evidence from Uganda. *Ecological Economics*, *119*, 346–358. https://doi.org/10.1016/j.ecolecon.2015.09.004

Dörgeloh, W. G. (1999). Assessment of veld conditions with multivariate techniques in mixed bushveld, South Africa. *African Journal of Ecology*, *37*(2), 194–201. https://doi.org/10.1046/j.1365-2028.1999.00168.x

Du, M. (2008). *Transfer of Resistance to Yellow Stem Borer (Scirpophaga incertalus Walker) from Wild Species into Rice (Oryza sativa L.) and Molecular Characterization of Introgression*. University of the

Phillipines Los Banos.

Fourcade, Y., Engler, J. O., R??dder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: A performance assessment of methods for correcting sampling bias. *PLoS ONE*, *9*(5), 1–13. https://doi.org/10.1371/journal.pone.0097122

Gao, L. Z., & Gao, C. W. (2016). Lowered diversity and increased inbreeding depression within peripheral populations of wild rice oryza rufipogon. *PLoS ONE*, *11*(3), 1–18. https://doi.org/10.1371/journal.pone.0150468

Ghesquière, A. (1986). Evolution of Oryza longistaminata. In *Rice Genetics: Proceedings of the International Rice Genetics Symposium* (pp. 15–25). Manila: International Rice Research Institute. https://doi.org/10.1017/CBO9781107415324.004

Gichuhi, E., Himi, E., Ahmed, N., Takahashi, H., & Maekawa, M. (2016). Preliminary QTL detection for improving basmati rice in a F2 population derived from the cross between kernel basmati and pLIA-1 carrying Oryza longistaminata chromosome. *Sabrao Journal of Breeding and Genetics*, *48*(4), 402–415.

Gichuhi, E., Himi, E., Takahashi, H., & Maekawa, M. (2016). Characterization and QTL Analysis of Oryza longistaminata Introgression Line, pLIA-1, derived from a Cross between Oryza longistaminata and Oryza sativa (Taichung 65) under Non-fertilized Conditions. *Rice Research: Open Access*, *4*(4). https://doi.org/10.4172/2375-4338.1000174

Giuliani, R., Koteyeva, N., Voznesenskaya, E., Evans, M. A., Cousins, A. B., & Edwards, G. E. (2013). Coordination of Leaf Photosynthesis, Transpiration, and Structural Traits in Rice and Wild Relatives (Genus Oryza). *Plant Physiology*, *162*(3), 1632–1651. https://doi.org/10.1104/pp.113.217497

Gregorio, G. B., Senadhira, D., & Mendoza, R. D. (1997). Screening Rice for Salinity Tolerance. *IRRI DISCUSSION PAPER SERIES*, *22*(22), 30.

He, R., Salvato, F., Park, J.-J., Kim, M.-J., Nelson, W., Balbuena, T. S., … Gang, D. R. (2014). A systems-wide comparison of red rice (Oryza longistaminata) tissues identifies rhizome specific genes and proteins that are targets for cultivated rice improvement. *BMC Plant Biology*, *14*(1), 46. https://doi.org/10.1186/1471-2229-14-46

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, *25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Khazaei, H., Street, K., Bari, A., Mackay, M., & Stoddard, F. L. (2013). The FIGS (Focused Identification of Germplasm Strategy) Approach Identifies Traits Related to Drought Adaptation in Vicia faba Genetic Resources. *PLoS ONE*, *8*(5). https://doi.org/10.1371/journal.pone.0063107

Khoury, C., Laliberté, B., & Guarino, L. (2010). Trends in ex situ conservation of plant genetic resources: A review of global crop and regional conservation strategies. *Genetic Resources and Crop Evolution*, *57*(4), 625–639. https://doi.org/10.1007/s10722-010-9534-z

Kiambi, D. K., Ford-Lloyd, B. V, Jackson, M. T., Guarino, L., N., M., & Newbury, H. J. (2005). Collection of wild rice (Oryza L.) in east and southern Africa in responsed to genetic erosion. *PGR Newsletter*, *142*(142), 10–20.

Kiambi, D. K., Newbury, H. J., Maxted, N., & Ford-Lloyd, B. (2008). Molecular genetic variation in the

African wild rice Oryza longistaminata A . Chev . et Roehr . and its association with environmental variables. *African Journal of Biotechnology*, *7*(10), 1446–1460. Retrieved from http://www.academicjournals.org/AJB

Kleynhans, M. T., James, C. S., & Birkhead, A. L. (2007). Hydrologic and hydraulic modelling of the Nyl River floodplain Part 3: Applications to assess ecological impact. *Water SA*, *33*(1), 21–25. https://doi.org/http://dx.doi.org/10.4314/wsa.v33i1.47867

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schr??der, B., Lindenborn, J., Reinfelder, V., … Wilting, A. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, *19*(11), 1366–1379. https://doi.org/10.1111/ddi.12096

Li, D. Z., & Pritchard, H. W. (2009). The science and economics of ex situ plant conservation. *Trends in Plant Science*, *14*(11), 614–621. https://doi.org/10.1016/j.tplants.2009.09.005

Li, Q., Yang, A., & Zhang, W.-H. (2016). Efficient acquisition of iron confers greater tolerance to saline-alkaline stress in rice (*Oryza sativa* L.). *Journal of Experimental Botany*, *67*(22), 6431–6444. https://doi.org/10.1093/jxb/erw407

Liu, C., Newell, G., & White, M. (2016). On the selection of thresholds for predicting species occurrence with presence-only data. *Ecology and Evolution*, *6*(1), 337–348. https://doi.org/10.1002/ece3.1878

Liu, L., Lafitte, R., & Guan, D. (2004). Wild Oryza species as potential sources of drought-adaptive traits. *Euphytica*, *138*(2), 149–161. https://doi.org/10.1023/B:EUPH.0000046801.27042.14

Mackay, M., & Street, K. (2004). Focused identification of germplasm strategy- FIGS. In *Proceedings of the 54th Australian Cereal Chemistry Conference and the 11th Wheat Breeders' Assembly* (pp. 138–141). Melbourne: Cereal Chemistry Division, Royal Australian Chemical Institute.

Marathi, B., Ramos, J., Hechanova, S. L., Oane, R. H., & Jena, K. K. (2015). SNP genotyping and characterization of pistil traits revealing a distinct phylogenetic relationship among the species of Oryza. *Euphytica*, *201*(1), 131–148. https://doi.org/10.1007/s10681-014-1213-2

Maxted, N., Dulloo, M., & Ford-Lloyd, B. (Eds.). (2016). *Enhancing Crop Genepool Use: Capturing Wild Relative and Landrace Diversity for Crop Improvement*. Oxfordshire: CAB International.

Pandey, M., & Rajora, O. P. (2012). Genetic diversity and differentiation of core vs. peripheral populations of eastern white cedar, Thuja occidentalis (cupressaceae). *American Journal of Botany*, *99*(4), 690–699. https://doi.org/10.3732/ajb.1100116

Panigrahi, D., Rajamani, S. (2008). Genetic Evaluation of the Wild Oryza Species for Resistance Against the Yellow Stem Borer, Scirophaga Incertulas Wlk. *Journal of Plant Protection and the Environment*, *5*(1), 26–29.

Pascual, U., & Perrings, C. (2007). Developing incentives and economic mechanisms for in situ biodiversity conservation in agricultural landscapes. *Agriculture, Ecosystems and Environment*, *121*(3), 256–268. https://doi.org/10.1016/j.agee.2006.12.025

Pourtier, R., & Sautter, G. (2014). Congo River. In *Encyclopaedia Brittanica*. Encyclopaedia Brittanica, Inc. Retrieved from https://www.britannica.com/place/Congo-River/Plant-life

Rakotomalala, M. (2001). Case History on Development and Deployment of Resistant Varieties to Control RYMV in Madagascar. In B. Hughes, J.; Odu (Ed.), *Plant Virology in Sub-Saharan Africa* (pp.

423–341). Ibadan: International Institute of Tropical Agriculture.

Ramos, J. M., Furuta, T., Uehara, K., Chihiro, N., Angeles-Shim, R. B., Shim, J., … Jena, K. K. (2016). Development of chromosome segment substitution lines (CSSLs) of Oryza longistaminata A. Chev. & R??hr in the background of the elite japonica rice cultivar, Taichung 65 and their evaluation for yield traits. *Euphytica*, *210*(2), 151–163. https://doi.org/10.1007/s10681-016-1685-3

Reinhold-Hurek, B., Hurek, T., Hu, L., Wang, Q., & Yang, H. (2014). Transgenic plants for nitrogen fixation.

Rodenburg, J., & Johnson, D. E. (2009). Weed Management in Rice-Based Cropping Systems in Africa. In *Advances in Agronomy* (1st ed., Vol. 103, pp. 149–218). Elsevier Inc. https://doi.org/10.1016/S0065-2113(09)03004-1

Ronald, P. C., Albano, B., Tabien, R., Abenes, L., Wu, K. S., McCouch, S., & Tanksley, S. D. (1992). Genetic and physical analysis of the rice bacterial blight disease resistance locus, Xa21. *Molecular & General Genetics : MGG*, *236*(1), 113–120. https://doi.org/10.1007/BF00279649

Scholte, P., de Kort, S., & van Weerd, M. (2000). Floodplain rehabilitation in Far North Cameroon: expected impact on bird life. *Ostrich*, *71*(December), 112–117. https://doi.org/10.1080/00306525.2000.9639884

Soriano, I. R., Schmit, V., Brar, D. S., Prot, J.-C., & Reversat, G. (1999). Resistance to rice root-knot nematode Meloidogyne graminicola identified in Oryza longistaminata and O. glaberrima. *Nematology*, *1*(4), 395–398. https://doi.org/10.1163/156854199508397

Tao, D., Hu, F., Yang, Y., Xu, P., Li, J., Wen, G., … Sripichitt, P. (2001). *Rhizomatous individual was obtained from interspecific BC2F1 progenies between Oryza sativa and Oryza longistaminata*. Retrieved from https://shigen.nig.ac.jp/rice/oryzabase/asset/rgn/vol18/c2.html

Thottappilly, G., & Rossel, H. (1993). Evaluation of Resistance to Rice Yellow Mottle Virus in Oryza species. *Indian Journal of Virology*, *9*(1), 65–73.

Tollefsrud, M. M., Sønstebø, J. H., Brochmann, C., Johnsen, Ø., Skrøppa, T., & Vendramin, G. G. (2009). Combined analysis of nuclear and mitochondrial markers provide new insight into the genetic structure of North European Picea abies. *Heredity*, *102*(6), 549–62. https://doi.org/10.1038/hdy.2009.16

Tu, J., Ona, I., Zhang, Q., Mew, T. W., Khush, G. S., & Datta, S. K. (1998). Transgenic rice variety “IR72” with Xa21 is resistant to bacterial blight. *Theoretical and Applied Genetics*, *97*(1–2), 31–36. https://doi.org/10.1007/s001220050863

Vaughan, D. A. (1994). *The Wild Relatives of Rice: A Genetic Resources Handbook* (1st ed.). Manila: International Rice Research Institute.

Vaughan, D. A., Song, G., Akito, K., & Norihiko, T. (2008). Phylogeny and Biogeography of the Genus Oryza. In H. Hirano, A. Hirai, Y. Sano, & T. Sasaki (Eds.), *Rice Biology in the Genomics Era* (pp. 219–234). Springer.

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. *Scientific Reports*, *5*(1), 13957. https://doi.org/10.1038/srep13957

Wang, G. L., Song, W. Y., Ruan, D. L., Sideris, S., & Ronald, P. C. (1996). The cloned gene, Xa21, confers

resistance to multiple Xanthomonas oryzae pv. oryzae isolates in transgenic plants. *Mol Plant Microbe Interact*. https://doi.org/10.1094/MPMI-9-0850

Zhang, S., Hu, J., Yang, C., Liu, H., Yang, F., Zhou, J., … Hu, F. (2017). Genotype by environment interactions for grain yield of perennial rice derivatives (Oryza sativa L./Oryza longistaminata) in southern China and Laos. *Field Crops Research*, *207*, 62–70. https://doi.org/10.1016/j.fcr.2017.03.007

Zhu, J., Dong, C. H., & Zhu, J. K. (2007). Interplay between cold-responsive gene regulation, metabolism and RNA processing during plant cold acclimation. *Current Opinion in Plant Biology*, *10*(3), 290–295. https://doi.org/10.1016/j.pbi.2007.04.010

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. *Molecular Biology and Evolution*, *24*(3), 875–888. https://doi.org/10.1093/molbev/msm005

Zhu, T., Xu, P. Z., Liu, J. P., Peng, S., Mo, X. C., & Gao, L. 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. *Molecular Phylogenetics and Evolution*, *70*(1), 348–361. https://doi.org/10.1016/j.ympev.2013.10.008