ANNUAL REVIEWS

Annual Review of Plant Biology

The Genomics of *Oryza* Species Provides Insights into Rice Domestication and Heterosis

Erwang Chen,[1,2] Xuehui Huang,[3] Zhixi Tian,[4] Rod A. Wing,[5] and Bin Han[1]

[1]National Center of Plant Gene Research; Shanghai Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences; and CAS Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200233, China; email: bhan@ncgr.ac.cn

[2]University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100049, China

[3]College of Life Sciences, Shanghai Normal University, Shanghai 200234, China; email: xhhuang@shnu.edu.cn

[4]State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

[5]Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA; email: rwing@email.arizona.edu

Annu. Rev. Plant Biol. 2019. 70:639–65

The *Annual Review of Plant Biology* is online at plant.annualreviews.org

https://doi.org/10.1146/annurev-arplant-050718-100320

Keywords

complex traits, domestication, genome sequencing, genotyping, heterosis, hybrid vigor, *Oryza*, plant fitness, rice breeding

Abstract

Here, we review recent progress in genetic and genomic studies of the diversity of *Oryza* species. In recent years, unlocking the genetic diversity of *Oryza* species has provided insights into the genomics of rice domestication, heterosis, and complex traits. Genome sequencing and analysis of numerous wild rice (*Oryza rufipogon*) and Asian cultivated rice (*Oryza sativa*) accessions have enabled the identification of genome-wide signatures of rice domestication and the unlocking of the origin of Asian cultivated rice. Moreover, similar studies on genome variations of African rice (*Oryza glaberrima*) cultivars and their closely related wild progenitor *Oryza barthii* accessions have provided strong evidence to support a theory of independent domestication in African rice. Integrated genomic approaches have efficiently investigated

many heterotic loci in hybrid rice underlying yield heterosis advantages and revealed the genomic architecture of rice heterosis. We conclude that in-depth unlocking of genetic variations among *Oryza* species will further enhance rice breeding.

Contents

1. INTRODUCTION

Asian cultivated rice (*Oryza sativa* L.), which was domesticated from its wild progenitor (*O. rufipogon*), is one of the most important crops in the world. According to the United Nations Food and Agriculture Organization, 70% more food must be produced over the next three decades to feed over 9 billion people by the year 2050 (24). Rapid population growth and the threat of climate change will require an efficient global strategy to ensure sustainable and equitable food security (34). Research on evolution of *Oryza* genomes has been an ongoing focus in the past decades (121, 143). The two processes involved in the evolution of rice, domestication and diversification, are closely related to human civilization and life. The main purpose of this review is to discuss three ways that rice genomics studies have been used to provide insights into rice domestication, rice heterosis, and rice breeding. Approximately 20 genomes of *Oryza* species and several thousands of rice cultivars have been deeply sequenced or resequenced with low-fold coverage (9, 60, 121, 170), which has advanced study on the genetics of rice domestication and heterosis. We discuss the evolutionary relationships and differentiation among the *Oryza* species, the typical AA genome domestication process between wild rice and cultivars, and the progress of recent research based on genome resequencing and functional genomics related to complex agronomical traits.

2. EVOLUTION OF *ORYZA* SPECIES

2.1. Natural Diversity of the Genus *Oryza*

The genus *Oryza* consists of 27 species that have been classified into 11 distinct genome types based on molecular markers and subsequent cytogenetic analysis. Of them, six are diploids ($n = 12$: AA, BB, CC, EE, FF and GG) and five are allotetraploids ($n = 24$: BBCC, CCDD, HHJJ, HHKK, and KKLL) (31, 113, 121, 170). Great efforts have been exerted in estimating the divergence time and ancestral effective population sizes of major lineages in *Oryza* (1, 41, 113), the results suggesting that *Oryza* originated in the middle Miocene (13–15 million years ago) and had two rapid diversifications (171).

Previous studies on comparative genomics of fully assembled genomes give us an opportunity to further understand the evolutionary relationships within the genus (60, 121, 132). The FF genome of the wild rice *O. brachyantha* contains a different set of repeat sequences in comparison to other rice genomes, which demonstrates its ancestral state and the underlying mechanisms of evolution within the *Oryza* genome (9). Moreover, comparative analysis of five de novo assembled AA-genome sequences, and lineage-specific expansion or contraction of gene families, demonstrates the evolution of disease resistance genes (121, 164). These *Oryza* genome resources provide insights into rice evolutionary genomics and are valuable for the improvement of rice and conservation of wild rice germplasm.

AA genomes of the genus *Oryza* have provided a natural model to investigate the evolution of plant genes and genomes (115, 132, 164). Phylogenetic analysis of the diploid AA-genome species indicates a close relationship within the genus *Oryza* and disjunctive distribution of these species in Asia, Africa, Australia, and South America. The AA genomes contain two cultivated species, *O. sativa* L. and *O. glaberrima*, and six wild species, *O. rufipogon*, *O. barthii*, *O. nivara*, *O. longistaminata*, *O. meridionalis*, and *O. glumaepatula* (164). The domesticated *O. sativa* and *O. glaberrima* rice cultivars represent two distinct types. *O. sativa* is distributed globally with a high concentration in South and East Asia, while *O. glaberrima* is grown in West Africa. *O. rufipogon* can be found throughout Asia and Oceania, where most of the accessions are perennial, while the *O. nivara* accessions are annual. Based on the neighbor-joining method, *O. rufipogon* can be divided into three types, *Or-I*, *Or-II*, and *Or-III*. Most of the *O. nivara* accessions can be classified into the *Or-I* type (14, 52). *O. barthii* and *O. longistaminata* are African species, with *O. barthii* being mainly distributed in West Africa, while *O. longistaminata* is found throughout Africa. Comparative genomic analysis among the de novo assembled AA-genome sequences for *O. nivara*, *O. glaberrima*, *O. barthii*, *O. glumaepatula*, and *O. meridionalis* may provide a powerful tool to identify specific genes associated with rice adaptation (164).

2.2. Differentiation of *O. sativa* L. ssp. *indica* and ssp. *japonica*

The Asian cultivated rice *O. sativa* can be classified into two major subspecies, *O. sativa* ssp. *japonica* and *O. sativa* ssp. *indica* (70), which are also referred to as subspecies *keng* (*japonica*) and *hsien* (*indica*), or *keng* (*japonica*) and *seng* (*indica*) (12, 18). These two subspecies represent most of Asian cultivated rice and have a global distribution. Other minor groups of Asian cultivars have also been classified. By using 15 polymorphic enzyme loci on 1,688 landraces, six varietal groups (I to VI) have been identified, among which groups I and IV correspond to typical *indica* and *japonica* varieties (69). Asian cultivars can be divided into five distinct groups, *indica*, *aus*, *aromatic*, *temperate japonica*, and *tropical japonica*. The classification is supported by simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs) (7, 32, 52). Similar results are shown in The 3,000 Rice Genomes Project, which indicates that *O. sativa* can be classified into five varietal groups: *indica*, *aus/boro*, *basmati/sadri*, *tropical japonica*, and *temperate japonica* (127, 138).

3. RICE DOMESTICATION

In the process of rice domestication, the elite or favored traits (domestication traits) in cultivated rice populations are selected and retained. Extant variants received beneficial alleles from their progenitors allowing them to adapt to the local environment and meet the desired domestication traits (36, 81, 101). Therefore, uncovering the process of rice domestication can bring a better understanding of the nature of artificial selection and may suggest a model applicable to domestication studies of other crop species (52, 59). Rice cultivation has been deeply influenced by the progress of human civilization (27, 43). Early human ancestors attempted to improve wild cereal crops by selecting for beneficial domestication genes inherited from their progenitors (20). Asian cultivated rice has been domesticated from ancestors of the wild rice species *O. rufipogon* (52). Large-scale sequencing data have shown a significant difference between wild rice and modern cultivars (52, 149).

The present rice variants and archaeological records can provide insights useful for domestication research. Archaeological evidence has shown that rice domestication began in the Yangtze Valley in China approximately 8,000–9,000 years ago and early cultivation along the Ganges River in India approximately 4,000 years ago (20). From fossil analysis, we can trace the time and place from changes in phenotype (27). Another method is to dissect the molecular mechanisms of standing domestication genes, such as *shattering 4* (*sh4*) and *PROSTRATE GROWTH 1* (*Prog1*) (66, 83, 126), which can give us a better understanding of nucleotide changes between domesticated rice and its progenitor.

Debates around the origin of rice, based on existing genetic and archaeological evidence, remain contentious. The rice single-origin theory, or snowball model, demonstrates that the earlier critical domestication gene was first introduced to other regions of Asia, where introgression occurred between the cultivar and local populations of *O. nivara* (also recognized as the annual type of *O. rufipogon*) and *O. rufipogon*. The multiple-origin model is a combination model wherein there are multiple mutations existing in divergent wild populations, and the key domestication genes between *indica* and *japonica* are formed by the hybridization between the subspecies after their independent domestication (111). Rice domestication initially resulted from the mutations in wild rice. Identifying only a few of the domestication genes is insufficient to reveal the overall process of domestication. Detecting selective signatures from genome variation can provide sufficient evidence to understand the origin of rice and its domestication.

3.1. Features of Domestication Traits

Compared to their progenitors, modern rice cultivars have great physiological and phenotypical differences between them. In a natural variant population, cultivars exhibit a striking difference in some agronomic traits, including plant architecture, hull color, shattering, awn, pericarp color, and grain size (125) (**Figure 1**). Domesticated plants are generally selected by breeders because they possess at least a subset of traits constituting the domestication syndrome (20, 27). In rice, this syndrome includes large seeds to improve the total production, high resource allocation, more determinate growth and apical dominance, and nonshattering seeds, all of which have been favored by our ancestors. Owing to the powerful dominance of the domestication syndrome, some beneficial alleles were first fixed by our ancestors because they conveyed some advantage in terms of cultivation or crop characteristics (110). These alleles have experienced a protracted process of domestication, through expansion and introgression with the local species. This could distinguish the domesticated plants from their progenitors.

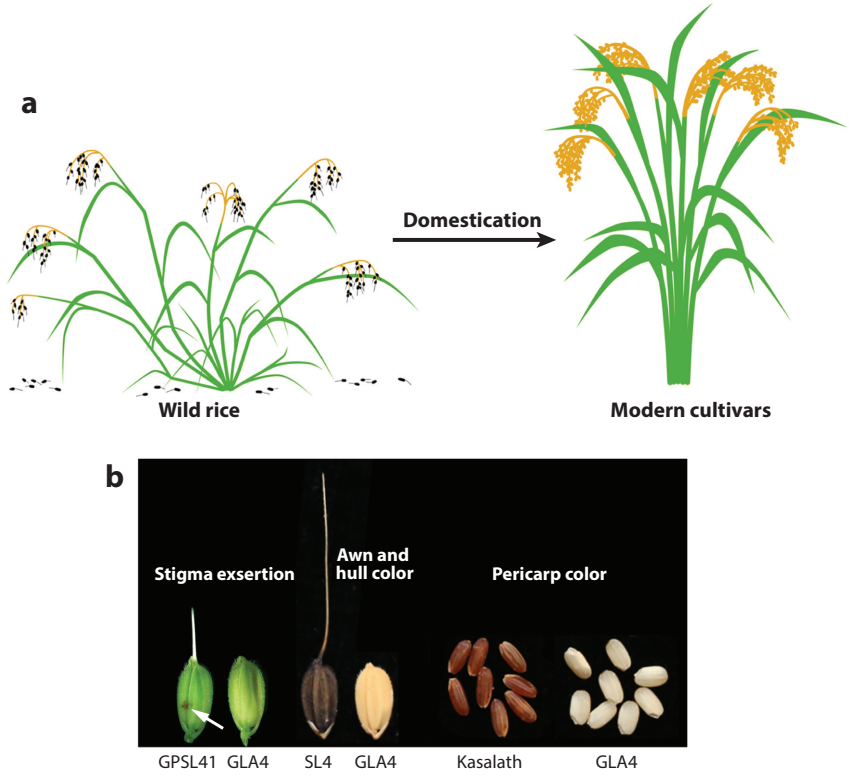

Figure 1

The different phenotypes between wild rice and cultivars. (*a*) Overview of wild rice and cultivars. (*b*) Differences between three standing domestication traits in wild rice and cultivars: stigma exsertion (*white arrow*) in GPSL41 [a chromosome segment substitution line (CSSL) from progeny of *indica* variety Guangluai4 (GLA4) and wild rice W1943] and *Oryza sativa* GLA4, awn and hull color in SL4 (a CSSL from progeny of *indica* GLA4 and wild rice *Oryza rufipogon* W1943) and GLA4, and pericarp color in Kasalath and GLA4.

A few domestication genes relevant to various agronomic traits have been identified and dissected to determine their regulating mechanism. If a gene were relevant to a domestication trait, it might show a decrease in nucleotide diversity, increased linkage disequilibrium, and altered population frequencies of polymorphic nucleotides in the gene and linked regions (7, 52). Past studies linking phenotype to genotype can help us to identify many trait-associated genes through quantitative trait locus (QTL) mapping. QTL mapping is based on the phenotypic differences between parental lines, which represents an effective and accurate way to identify relevant domestication and diversification genes that contribute to phenotypes of interest. *Teosinte glume architecture* (*tga*) is a gene in maize that controls differences in inflorescence architecture (134). In maize, *teosinte branched1* (*tb1*) is the first-identified domestication gene underlying a difference in apical dominance as compared with its progenitor, *teosinte* (21). Higher expression patterns in maize may be have been influenced by human selection (144). The *Q* locus in wheat is a major pleiotropy gene controlling many traits, including the tendency of the spike to shatter and the tenacity of the chaff surrounding the grain. In addition, it is a plant-specific transcriptional regulator in the AP2 family, and a single amino acid change can affect protein dimerization (119).

In rice, shattering is a notable domestication trait that could have been easily selected for by our ancestors and which directly contributes to crop yield. *qSH1* is a major QTL controlling shattering, encoding a homeobox containing a transcription factor, and fixed in *japonica* subspecies. The causative mutation is a single nucleotide in a *cis*-regulatory element, regulating the shattering zone (76).

Long awns in rice are unpopular during harvest and storage, and the trait was artificially selected for during domestication to create the cultivars with short awns or no awns. Population genetic analysis shows a significant reduction in nucleotide diversity of the *An-1* locus in rice cultivars, indicating that the locus is a target for artificial selection. Besides, *An-1* also has pleiotropic effects on grain length, and grain number per panicle in rice (97). Similarly, *An-2/LABA1* is also a domestication gene with a small effect in controlling awn length and grain production (39, 49). *An-2/LABA1* has an additive effect in combination with *An-1* and produces a longer awn in rice development. Another awn gene *GAD1/RAE2* encodes a small secretory signal peptide belonging to the EPIDERMAL PATTERNING FACTOR-LIKE family, and the loss-of-function allele causes the increased number of grains per panicle, shorter grains, and awnless phenotype of cultivated rice (4, 65, 156).

PROG1 is a major QTL regulating tiller angle and the number of tillers in rice. Only one mutation causes an amino acid substitution in *O. sativa*, which is responsible for the phenotype changes. This substitution is located at the functional C terminus of the protein and has a direct influence on transcriptional activation (66, 126). *OsLG1* has been identified in a strong selective sweep with reduced nucleotide diversity at the SPR3 locus between cultivated *O. sativa* and wild *O. rufipogon*. *OsLG1*, which controls a simple morphological change in rice panicle shape and has a large effect on seed shedding and pollinating behaviors, encodes the SQUAMOSA promoter-binding protein (SBP) domain, and controls laminar joint and ligule development. It is very interesting that this gene and its upstream 9.3-kb region are jointly responsible for phenotype change, producing open panicles similar to the wild parent. From complementation tests, the panicle phenotypes directly depend on the expression levels of *OsLG1*; from these results, we can conclude that the 9.3-kb upstream region regulates the expression of *OsLG1* (61).

3.2. The Origin of Asian Cultivated Rice

Despite many rice domestication traits having been explored, the origin of the Asian cultivated rice *Oryza sativa* L. spp. *indica* and spp. *japonica* has long been an unsolved mystery with many unresolved questions. What is the geographical origin of cultivated rice? Which types of *O. rufipogon* accessions are the direct wild progenitors of cultivated rice? Did the two subspecies of cultivated rice, *indica* and *japonica*, derive from a single or multiple domestication processes? A comprehensive genome variation study will be an efficient approach to answer these questions. A large number of representative diverse accessions of the wild rice *O. rufipogon* with geographically broad distributions and *indica* and *japonica* varieties should be collected and sequenced to investigate the genome variations and relationships among the complex of *O. sativa* L. and *O. rufipogon*. A practicable strategy for the analysis of rice genome variations has been proposed: (*a*) collection of domesticated rice and its wild progenitors, (*b*) whole genome profiling of sequence variation, (*c*) detection of selective sweeps from genomic screening, and (*d*) annotation of domestication loci and inference of the origin of rice domestication (42, 52). Owing to its relatively lower cost and more developed technology, the strategy of sequencing has been successfully applied in several cereal species (19, 52).

In total, 446 geographically diverse accessions of wild rice species *O. rufipogon* and 1,083 cultivated *indica* and *japonica* varieties have been collected and sequenced. The comparative genome

analysis shows that subspecies *indica* and *japonica* descend from different subpopulations of wild rice, namely, *Or-I* and *Or-IIIa*, respectively (51, 52). By screening the selective signatures in the whole rice genome, 55 selective sweeps have been identified. A few of them contain characterized domestication genes (**Table 1**), such as *sh4* (83), *Bh4* (169), *qSH1* (76), *Prog1* (66, 126), *LG1* (61), *An-1* (97), and *An-2* (39, 49). The phylogenetic tree derived from the genomic data of all 55 domestication loci showed that unlike the genome-wide pattern, the two subspecies are often clustered together at the domestication loci. Based on the analysis of approximately 8 million SNP sites from 1,529 genomes and integrating all the data, Huang et al. (52) proposed a demographic scenario in which *japonica* was first domesticated from *Or-IIIa*, whereas *indica* was subsequently developed from *Or-I* with the adoption of many domestication alleles from *japonica*. The *O. sativa* ssp. *japonica* group was initially domesticated from wild rice in southern China (52). The *O. sativa* ssp. *indica* group was created subsequently when *japonica* rice spread into Southeast Asia and South Asia and was crossed with local wild rice (52).

Thus, the rice domestication process can be proposed as a single-origin model with multiple introgressions (**Figure 2**). Explained simply, some useful mutations may have randomly occurred in a population of wild rice species of *Or-III* and were then selected and fixed, generating the *sinica* cultivars or *proto-japonica* varieties. The *sinica* cultivars or *proto-japonica* varieties were further spread to other places in Asia. The *indica* varieties were subsequently generated through crosses between the *proto-japonica* varieties and the *Or-I* accessions after many cycles of crosses and selections. The favored mutations fixed with their flanking regions in cultivated rice provide strong evidence to explain their origins (51, 52). A recent study confirmed that there is significant gene flow from *japonica* to both *indica* (17%) and *aus* (15%), which led to the transfer of domestication alleles from early domesticated *japonica* to *proto-indica* and *proto-aus populations* (14). The results support a model in which different rice subspecies have separate origins, but de novo domestication occurred only once, in *O. sativa* ssp. *japonica*, and introgressive hybridization from early *japonica* to *proto-indica* and *proto-aus* led to domesticated *indica* and *aus* rice (14). Although more candidate domestication genes in the most selective sweep regions are still to be functionally characterized, the role of the genes associated with key domestication traits has been well explained in previous studies, which may inspire us to trace the time and geographical origins of specific traits.

3.3. Independent Domestication of African Rice

Unlike the widespread cultivation of Asian rice, African rice has had a more limited geography of cultivation (130). *O. glaberrima* has been cultivated in West Africa for approximately 3,000 years (91). Analysis of the molecular data of isozyme studies, as well as the simple sequence repeat (SSR) and SNP data, indicates that African rice is closely related to *O. barthii* (57, 125, 135). Compared to Asian cultivated rice, African rice has some undesirable features, such as a seed that shatters easily, brittle grain, and lower yields. Currently, *O. glaberrima* is being replaced everywhere in West Africa by the Asian species, which were introduced into the continent by the Portuguese as early as the mid-sixteenth century (125, 130).

Population genomic analyses of 20 *O. glaberrima* and 94 *O. barthii* accessions demonstrate the evolutionary history of domestication and artificial selection in African rice (*O. glaberrima*). This study provides evidence to support the proposition that *O. glaberrima* was domesticated from the *O. barthii* subgroup and independently domesticated from *O. sativa* (135). Similar work combining the resequencing data of *O. sativa*, *O. rufipogon*, *O. glaberrima*, and *O. barthii* demonstrates a strong genetic differentiation between Asian and African rice (56). Furthermore, the resequencing of 93 traditional *O. glaberrima* landraces from across the species range in West and Central sub-Saharan Africa illustrates the domestication history and geographical adaptation. Based on the SNP map,

Table 1 Selective signatures in the whole rice genome

Quantitative trait locus	Trait	Gene ID	Encoding protein	Biological function	Reference(s)
Domestication genes					
Seed shattering in chromosome 1 (*qSH1*)	Shattering	Os01g0848400	BEL1-type homeobox	Absence of abscission layer formation	76
Shattering 4 (*sh4*)	Shattering	Os04g0670900	Transcription factor	Absence of abscission layer formation	83
Awn-1 (*An-1*)	Awn length	Os04g0350700	Helix-loop-helix protein	Long awn and larger grains by regulating cell division	97
Awn-2 (*An-2*)/*LONG AND BARBED AWN 1* (*LABA1*)	Awn length	Os04g0518800	Cytokinin-activating enzyme	Reducing the cytokinin concentration	39, 49
GRAIN NUMBER, GRAIN LENGTH AND AWN DEVELOPMENT 1 (*GAD1*)/*REGULATOR OF AWN ELONGATION 2* (*RAE2*)	Awn length	Os08g0485500	Secretory signal peptide	Peptide signal	4, 65, 156
Black hull 4 (*Bh4*)	Hull color	Os04g0460000	Amino acid transporter	Controlling seed hull color	169
Red pericarp (*Rc*)	Pericarp color	Os07g0211500	Basic helix-loop-helix protein	Positive regulator of proanthocyanidin synthesis	38
Liguleless gene 1 (*OsLG1*)	Panicle shape and ligule development	Os04g0656500	SQUAMOSA promoter-binding protein (SBP)	Morphological change of panicle shape	61
QTL for grain width and weight on chromosome 5 (*GW5*)/*GRAIN SIZE ON CHROMOSOME 5* (*GSE5*)/*QTL for seed width on chromosome 5* (*qSW5*)	Grain width	Os05g0187500	Plasma membrane–associated protein with IQ domains	Interacts with *OsGSK2* in the brassinosteroid signaling pathway	92, 117, 142
PROSTRATE GROWTH 1 (*Prog1*)	Tiller angle	Os07g0153600	Zinc-finger nuclear transcription factor	Regulates tiller angle and number of tillers	66, 126
Improvement and diversification genes					
Ideal Plant Architecture 1 (*IPA1*)/*SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 14* (*OsSPL14*)/*WEALTHY FARMER'S PANICLE* (*WFP*)	Plant architecture	Os08g0509600	Transcription factor containing SBP-box	Controlling ideal plant architecture	64, 103, 162

(*Continued*)

Table 1 (*Continued*)

Quantitative trait locus	Trait	Gene ID	Encoding protein	Biological function	Reference(s)
DENSE AND ERECT PANICLE 1 (*DEP1*)	Dense and erect panicle	Os09g0441900	Phosphatidylethanolamine-binding protein-like	Enhanced meristematic activity	53, 123
Tiller Angle Control 1 (*TAC1*)	Tiller angle	Os09g0529300	Novel small gene family	Regulation of tiller angle	158
Grain width 2 (*GW2*)	Grain width	Os02g0244100	E3 ubiquitin ligase	Ubiquitin-proteasome pathway	120
Regulator of grain size 5 (*GS5*)	Grain length and width	Os05g0158500	Serine carboxypeptidase	Positive regulator of grain size	88
Grain weight on chromosome 6a (*GW6a*)	Grain length	Os06g0650300	Histone H4 acetyltransferase	Regulation of grain weight and yield via controlling cell number and grain filling rates	120
THOUSAND-GRAIN WEIGHT 6 (*TGW6*)	Grain weight	Os06g0623700	Indole-3-acetic acid (IAA)-glucose hydrolase	Controlling IAA supply and limiting cell size	62
Grain Length on Chromosome 7 (*GL7*)*/Grain Width on Chromosome 7* (*GW7*)	Grain length and width	Os07g0603300	TONNEAU1-recruiting motif protein	Increase in grain length and improvement of grain appearance quality	136, 139
GRAIN LENGTH AND WEIGHT ON CHROMOSOME 7 (*GLW7*)	Grain length	Os07g0505200	Transcription factor containing SBP-box (*OsSPL13*)	Regulates rice grain length via controlling cell size in the grain hull	118
GRAIN WIDTH ON CHROMOSOME 8 (*GW8*)	Grain width	Os08g0531600	Transcription factor containing SBP-box (*OsSPL16*)	Promotes cell division and grain filling	137
Heading date 3a (*Hd3a*)	Heading date	Os06g0157700	Florigen	Short-day condition promoting heading and controlling lateral branching	74
Grain number, plant height, and heading date 7 (*Ghd7*)	Heading date	Os07g0261200	CCT-domain protein	Regulating number of grains per panicle, plant height, and long-day and delaying heading	150

(*Continued*)

Table 1 (*Continued*)

Quantitative trait locus	Trait	Gene ID	Encoding protein	Biological function	Reference(s)
QTL for days to heading on chromosome 8 (*Ghd8*)	Heading date	Os08g0174500	Putative HAP3 subunit of the CCAAT box-binding transcription factor	Delaying heading data under long-day conditions and also controlling plant height as well as number of grains per panicle	152
Heading date 1 (*Hd1*)	Heading date	Os06g0275000	Florigen	Promotion of heading under short-day conditions and in inhibition under long-day conditions	157
Semidwarf 1 (*Sd1*)	Plant height	Os01g0883800	The oxidase enzyme, GA20ox	Taking part in the biosynthesis of gibberellin	80, 112
Thermosensitive male sterility 5 (*Tms5*)	Thermosensitive male sterility	Os02g0214300	RNase Z protein	Processing male sterility–related Ub_{L40} mRNAs into fragments	168
Pigm	Blast resistance	National Center for Bio-technology Information (NCBI) accession KU904633	Nucleotide-binding leucine-rich repeat receptors	Conferring durable resistance to the fungus *Magnaporthe oryzae* without yield penalty	16
Brown planthopper resistance 6 (*Bph6*)	Brown planthopper resistance	NCBI accession KX818197	Exocyst-localized protein	Increasing exocytosis, participating in cell wall maintenance and reinforcement, as well as activating cytokinin, salicylic acid, and jasmonic acid signaling pathway	40
Submergence 1A (*Sub1A*)	Submergence tolerance	NCBI accession DQ011598	Putative ethylene response factors	Crosstalk in multiple signaling pathways during submergence	148
SNORKEL1	Deep water adaptation	DNA Data Bank of Japan (DDBJ) accession AB510478	Ethylene response factors	Triggering internode elongation via gibberellin	44
SNORKEL2	Deep water adaptation	DDBJ accession AB510479	Ethylene response factors	Triggering internode elongation via gibberellin	44

(*Continued*)

Table 1 (*Continued*)

Quantitative trait locus	Trait	Gene ID	Encoding protein	Biological function	Reference(s)
Thermo-tolerance 1 (*TT1*)	Thermotolerance	Os03g0387100	α2 subunit of the 26S proteasome	Modulation of protein homeostasis by the proteasome for thermotolerance	87
Nitrate-transporter 1.1B (*NRT1.1B*)	Nitrate uptake	Os10g0554200	Nitrate transporter	Enhanced nitrate uptake	46
DEEPER ROOTING 1 (*Dro1*)	Deeper rooting	Os09g0439800	Unknown membrane protein	Controlling root growth angle to improve drought avoidance	128, 129
Grain chalkiness 5 (*Chalk5*)	Grain quality	Os05g0156900	Vacuolar H^+-translocating pyrophosphatase	Increasing the chalkiness by disturbing pH homeostasis	89
Waxy	Grain quality	Os06g0133000	Starch granule-bound starch synthase	Controlling the synthesis of amylose	141

11 significant QTLs associated with six salt tolerance traits have been identified by genome-wide association study (GWAS), suggesting adaptive geographical divergence for salt tolerance (100).

African rice has an evolutionary lineage parallel to but different from that of Asian rice (105). Some common genes have different diversity patterns within these two rice species. Those involving parallel evolutionary phenotypes such as seed shattering, hull color, and pericarp color have been investigated (130). Loss of function of *Rc* and deletion of *Hd1* have parallel roles in both *O. sativa* and *O. glaberrima* (37, 105). *Shattering 1* (*OsSh1*) in *O. sativa* encodes a YABBY transcription factor underlying the shattering phenotype. Sequence analysis shows the orthologous region in O. *glaberrima* has a 45-kb deletion that results in the complete ablation of the *O. glaberrima OsSh1* ortholog and three additional genes (135). In addition, *GL4* is the ortholog gene of *SH4*, controlling seed shattering in Asian rice, which also regulates the grain length in African rice. An SNP mutation in the *GL4* gene results in a premature stop codon and leads to small seeds and loss of seed shattering during African rice domestication (98, 146). *PROSTRATE GROWTH* 7 (*PROG*7) is the main determinant in controlling the transition of plant architecture from prostrate growth to erect growth during African rice domestication. Comparing *PROG*7 to the known *PROG1*, both genes belong to the zinc-finger domain transcription factor family (the protein sequence similarity is 35%) and locate in almost the same region, suggesting that the *PROG*7 of *O. glaberrima* is identical to the *PROG1* of *O. sativa* (47).

African cultivars and their wild rice species are known to possess enormous genetic diversity, which is of potential genetic value in terms of resistance to biotic and abiotic stress (133). Typical useful traits found in African *Oryza* species, like resistance to drought, iron toxicity, the green leafhopper, weed competition, and bacterial blight, have not been exploited rationally (91). *Thermo-tolerance 1* (*TT1*) is a major thermotolerance QTL identified in African rice that encodes an α2 subunit of the 26S proteasome involved in the degradation of ubiquitinated proteins (87). Enhanced thermotolerance in rice will increase food security in the face of global warming and unpredicted climate changes.

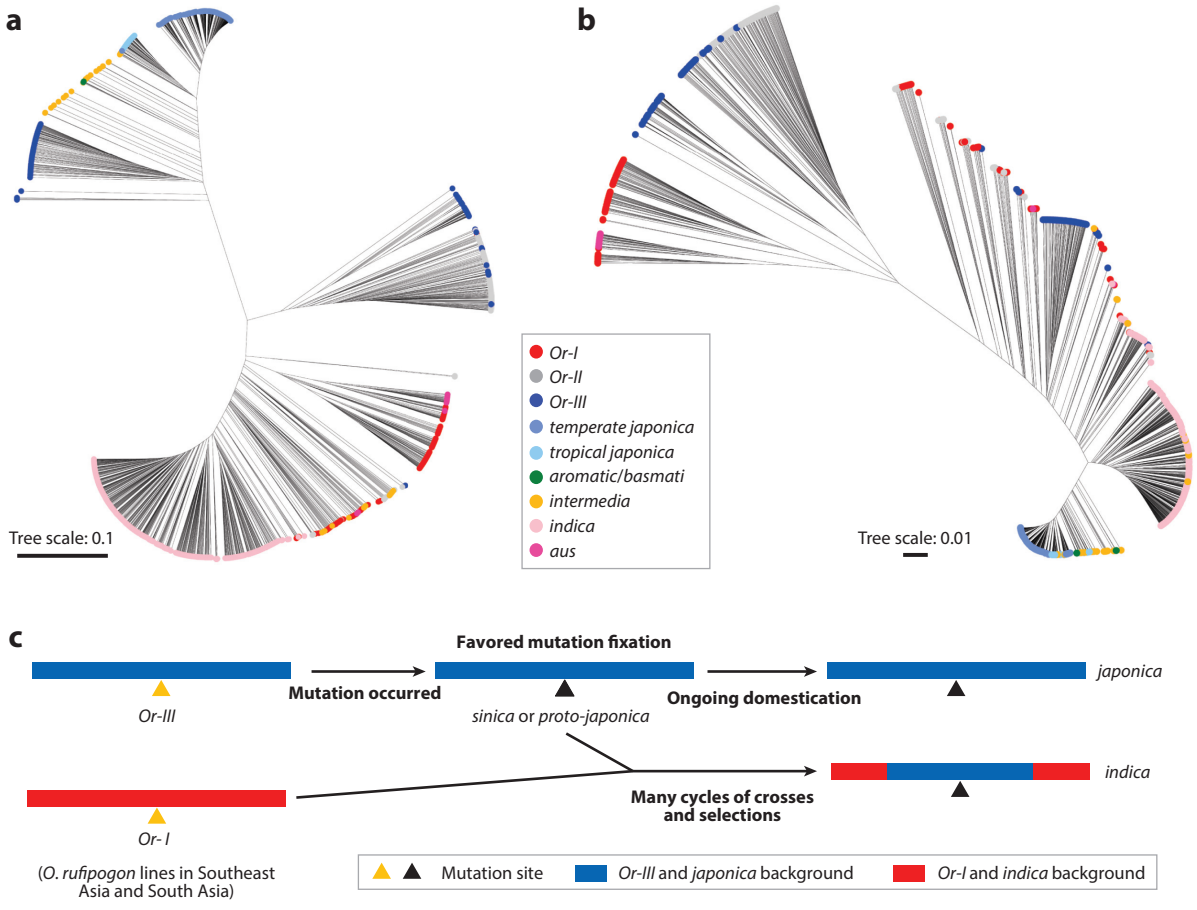

Figure 2

Phylogenetic tree of wild–domesticated rice accessions and the single-origin model of rice domestication. (*a*) Phylogenetic tree of 1,529 *Oryza rufipogon* and *Oryza sativa* accessions was constructed from single-nucleotide polymorphisms (SNPs) of the whole genome using the unweighted pair group method with arithmetic mean (UPGMA). (*b*) Phylogenetic tree of 1,529 *O. rufipogon* and *O. sativa* accessions was constructed from the SNPs of 55 domestication sweeps. The 1,529 accessions shown in panels *a* and *b* include 155 of *Or-I* (*red*), 121 of *Or-II* (*gray*), 170 of *Or-III* (*deep blue*), 409 of *temperate japonica* (*blue*), 75 of *tropical japonica* (*light blue*), 5 of *aromatic/basmati* (*dark green*), 44 of *intermedia* (*yellow*), 520 of *indica* (*pink*), and 30 of *aus* (*deep pink*). (*c*) Proposed single-origin model of rice domestication. The close ancestors (*O. rufipogon*) of Asian cultivated rice are divided into three main types, *Or-I*, *Or-II*, and *Or-III*. Useful mutations may have randomly occurred in some populations of wild rice species *Or-III* that were then selected for to generate the *sinica* or *proto-japonica* varieties. Each small triangle represents a mutation site existing in a domestication locus (*rectangle*), and the color changes indicate a favored mutation (*black*) event from wild rice (*yellow*). The *indica* varieties were subsequently developed due to acquisition of favored mutations (*black*) through the crosses between the *sinica* or *proto-japonica* varieties and the *Or-I* varieties (*O. rufipogon* lines in Southeast Asia and South Asia) after many cycles of crosses and selections. The modern *japonica* varieties were domesticated through ongoing selections. The genomic backgrounds of the wild rice species *Or-III* and *japonica* varieties (*blue*) and those of the wild rice species *Or-I* and *indica* varieties (*red*) are shown. Diverse natural variants by the introgression between *indica* and *japonica* were then widely distributed to adapt to the local environment.

3.4. The Evolution of Weedy Rice

Hybridization and introgression can play important roles in the genetic differentiation and adaptive evolution of plant species (63). Weedy rice (*Oryza sativa* f. *spontanea*), also called red rice, is a conspecific weed of cultivated rice that aggressively outcompetes crops and reduces harvests (22, 37). Its typically weedy features show highly shattering seeds, persistent seed dormancy, rapid growth, and the ability to aggressively outcompete the crop for nutrients and light (67). In focusing on SNPs identified among sequenced accessions, phylogenetic evidence lined up with other data to show that dedomestication has led to at least three types of weedy rice—those from China, US weedy rice with straw-colored hulls (SH), and US weedy rice with black hulls and long awns (BHA) (93, 131). This kind of weed–rice hybridization occurs at low frequencies during rice domestication, and crop-to-weed gene flow significantly influenced the adaptive evolution of weeds (116). The rapid gene infusion increasing weediness and invasiveness in weedy rice may become the best model to investigate the process of crop dedomestication or crop–weed coevolution.

Three questions may be involved in uncovering the process of dedomestication: (*a*) What was the process of evolution? (*b*) When did the divergence of the three weedy strains occur? (*c*) What has shaped the weedy rice genome in adaptation? The study of whole-genome sequence analyses was to examine the evolution of weedy rice within the US and Chinese strains (84). The evolutionary relationship of three weedy rice accessions shows through phylogenetic analyses that SH accessions are clustered with Southeast Asian *indica* accessions. BHA accessions are grouped with *aus* and wild rice accessions originating from the Indian subcontinent, while Chinese weed strains are similar to Chinese *indica* varieties (124). In addition, the US strains have a higher proportion of wild-specific SNPs than crop-specific SNPs, and the Chinese weeds are closely related to modern domesticated rice. Selections during weed evolution often act as genomic islands, which are enriched in genes involved in tissue development and stress response and randomly distributed across the SH and BHA genomes (84). Population analyses of 155 weedy and 76 cultivated rice accessions in four representative regions support the idea that Chinese weedy rice was dedomesticated independently from cultivated rice and has a strong genetic bottleneck (107). The standing (pre-existing) variations have a more rapid allele fixation rate than new mutations, which may contribute to rapid environmental adaptation (107).

Analyses of 12 domestication and improvement genes within three weedy rice strains have shown that dedomestication-related traits could play a crucial role in tracing and assessing the specific time of crop domestication. Three standing domestication genes, *PROG1*, *sh4*, and *OsLG1*, are all fixed within three weedy rice groups (107). However, QTLs of novel mutations associated with weediness traits may help us to investigate how these unique weeds adapt to agricultural environments quickly. Additionally, gene flow from transgenic herbicide-resistant crops to weedy rice has led to changes in biodiversity balance (95, 99, 155). Effective control of the herbicide-resistant strains will be a challenge in the future.

4. NATURAL VARIATION AND PLANT ADAPTATION

4.1. Rice Pan-Genome Analysis

Asian cultivated rice and its closest wild relative *O. rufipogon* grow in diverse ecogeographical areas across the world and harbor a high level of genetic diversity to maintain the local adaptation of the plants' response to environmental changes. There are partial or nearly no reproductive isolations within the *O. sativa–O. rufipogon* species complex (71), thus greatly facilitating the use of diverse natural alleles within the complex for genetic studies and breeding. Genomics approaches now enable us to investigate the allelic variation precisely and comprehensively. In a recent study for

the establishment of a rice pan-genome, a total of 66 phylogenetically representative accessions (including both *O. sativa* and *O. rufipogon*) were carefully selected from approximately 1,500 diverse accessions to maximize the genetic diversity in rice (166). Through deep sequencing (115-fold coverage) and whole-genome de novo assembly, 66 complete rice genome sequences were constructed and the coding genes in the genomes well annotated. Using the pan-genome data set, nearly the whole set of genes among rice was ascertained, and the complex genomic variation (including millions of structural variants) was captured. It was found that each rice gene contained 10 missense SNP and 6 polymorphic sites of relatively large effect on average, creating multiple diverse alleles (approximately 16 diverse alleles per gene locus). The naturally occurring variations were often present in only one or a few accessions. Alleles located within different regions of genes (e.g., either the intron region or coding region for protein domains) with different destructive powers (e.g., single missense mutations of a nondomain region or frameshift of the whole gene) may result in different phenotypic effects. In *waxy*, a well-known gene for grain quality, three alleles (wild type, a mutation in the intron–exon junction site, and one with a 23-bp indel) caused different levels of amylase content in rice grains. In another study of 3,000 rice genomes, it was found that approximately half (56%) of rice genes contained high-effect SNPs for gene coding, and most of the variants belonged to rare alleles present in at least four accessions in a subpopulation of cultivated rice and not found in other subpopulations (138). Moreover, according to the pan-genome analysis, 10,872 coding genes absent in the Nipponbare reference genome were identified from the 66 rice genomes, many of which may have important biological functions according to transcript evidence and known protein domains (166). The newly identified genes included several functionally important genes detected previously by map-based cloning, for example, *Sub1A*, *SNORKEL1* and *SNORKEL2* (controlling submergence tolerance), and *Pstol* (controlling phosphorus-deficiency tolerance) (28, 44, 148). In 3,000 rice genomes analyzed, 12,465 novel, full-length genes were predicted through a map-to-pan strategy (i.e., de novo assembly of the sequence reads not mapped onto the Nipponbare reference). The novel genes showed enrichment in the gene families for immune defense responses and ethylene metabolism (138). Furthermore, the construction of the high-quality reference genome of the wild species in the *Oryza* genus resulted in detection of more species-specific genes or gene families, especially for the plant disease resistance genes with rapid evolution rates. These genes, absent in the Nipponbare genome but identified from the genome studies of the diverse rice samples, may provide a valuable resource in the future, as exemplified by the wide use of the disease resistance gene *Xa21* from the wild rice *O. longistaminata* (15, 104). Taken together, the natural accessions provide a rich resource for us to search for alleles with favored genetic effects in breeding.

4.2. Identifications of Genes for Plant Fitness by GWAS

With the global distribution of rice plants, the diverse alleles for plant fitness play an important role in rice genomes (145). Forward genetics approaches are one of the best ways to explore the genes underlying plant fitness for diverse rice growth environments and clarify the biological or phenotypical effects of the natural alleles. The use of diverse rice accessions in GWAS has been developed and widely implemented for genetic mapping of many agronomically important rice traits (10, 11, 42, 54, 55, 58, 149). For instance, *GRAIN LENGTH AND WEIGHT ON CHROMOSOME 7* (*GLW7*) is a major QTL that encodes the plant-specific transcription factor *OsSPL13* and positively regulates cell size, resulting in enhanced rice grain length and yield; it was identified through implementing an approach integrating GWAS with functional analysis in a diverse rice population (118). Further analysis indicates that the allelic variations of *GLW7* are strongly associated with divergence in grain size between *tropical japonica* and *temperate japonica* rice varieties (118).

However, in conventional GWAS, it is very difficult to identify associations from rare alleles or those confounded by complex genetic structures (especially in rice with strong population differentiations), although some statistical methods (e.g., linear mixed model) provide some improvements (42). In rice, the fixation index (F_{ST}), which is a measure of population differentiation, is approximately 0.55 between *indica* and *japonica* rice and 0.17–0.36 between *O. sativa* and *O. rufipogon*, owing largely to long-term geographical isolation coupled with the self-fertilization characteristic of natural populations of rice (52). For genes underlying plant fitness, allelic distributions were strongly biased in the populations (e.g., between *indica* and *japonica* rice), suggesting selective pressure or genetic drift. Hence, many causative genes underlying plant fitness would probably be missed through GWAS using diverse rice accessions with strong population differentiations. GWAS using multiple collaborative populations for joint analyses may be needed in rice, and similar approaches have been designed and performed successfully in maize (nested association mapping populations) and *Arabidopsis thaliana* (multiparent advanced generation inter-cross populations) (25, 33, 77, 82, 122). Since there are no or only partial reproductive isolations among diverse rice accessions, the phylogenetically representative accessions in the *O. sativa–O. rufipogon* species complex could be used as the parental lines for the construction of recombinant populations, which should be useful in both molecular breeding and large-scale genetic mapping for plant fitness.

Another challenge in rice genetics, especially for studies on plant fitness, lies in the effective investigation of genetic interactions between QTLs (broadly defined as epistasis, G×G) and gene–environment interactions (G×E). For example, to date, tens of QTLs underlying grain yield have been identified and functionally validated in rice, but the genetic interactions among these QTLs are only partially clear. The genetic architecture of plant fitness may be more complicated than that for grain yield in rice, where both G×G and G×E are important for the presence of plant fitness to diverse environments (85). In future, the improvements in quantitative genetics methods coupled with functional genomics approaches (such as genome editing technology) may be needed to clarify the interactions.

4.3. Insights into Rice Heterosis

Genetic diversity is also fundamentally important in hybrid rice breeding to develop heterotic rice hybrids because the occurrence of heterosis is dependent on the genetic differences between parental lines (56). In plants, heterosis, also known as hybrid vigor or outbreeding enhancement, refers to the increased yield or biomass in a hybrid offspring over its inbred parental lines (13). In rice, the hybrid varieties (the heterozygous F_1 generation) typically display a grain yield advantage of 10–30% over their parents (96). Exploiting the heterosis phenomenon in breeding is one of the most efficient ways to increase grain yield to meet the demands of global food security, and the development of hybrid rice breeding represents one of the most exciting advances in agricultural genetics (140). The genomic basis of heterosis has been extensively studied in a number of crops and explained by several classical models (147, 167). Previous research in multiple crops has provided support for two non–mutually exclusive hypotheses of dominance (pseudo-overdominance) and single-locus overdominance for genomic loci contributing to heterosis (30, 45, 48, 50, 55, 78, 86, 108, 147, 161, 167). However, the genetic cause of heterosis in rice, specifically, the major causative genes, has long been a puzzle despite the heterotic phenomenon having been known for several decades with various genetic models suggested to explain it.

Recently, the genetic basis of heterosis for rice grain yield has been explored through an integrated genomics approach to construct a genome map for 1,495 elite hybrid rice varieties and their inbred parental lines (54, 56). To identify loci contributing to yield traits and analyze the

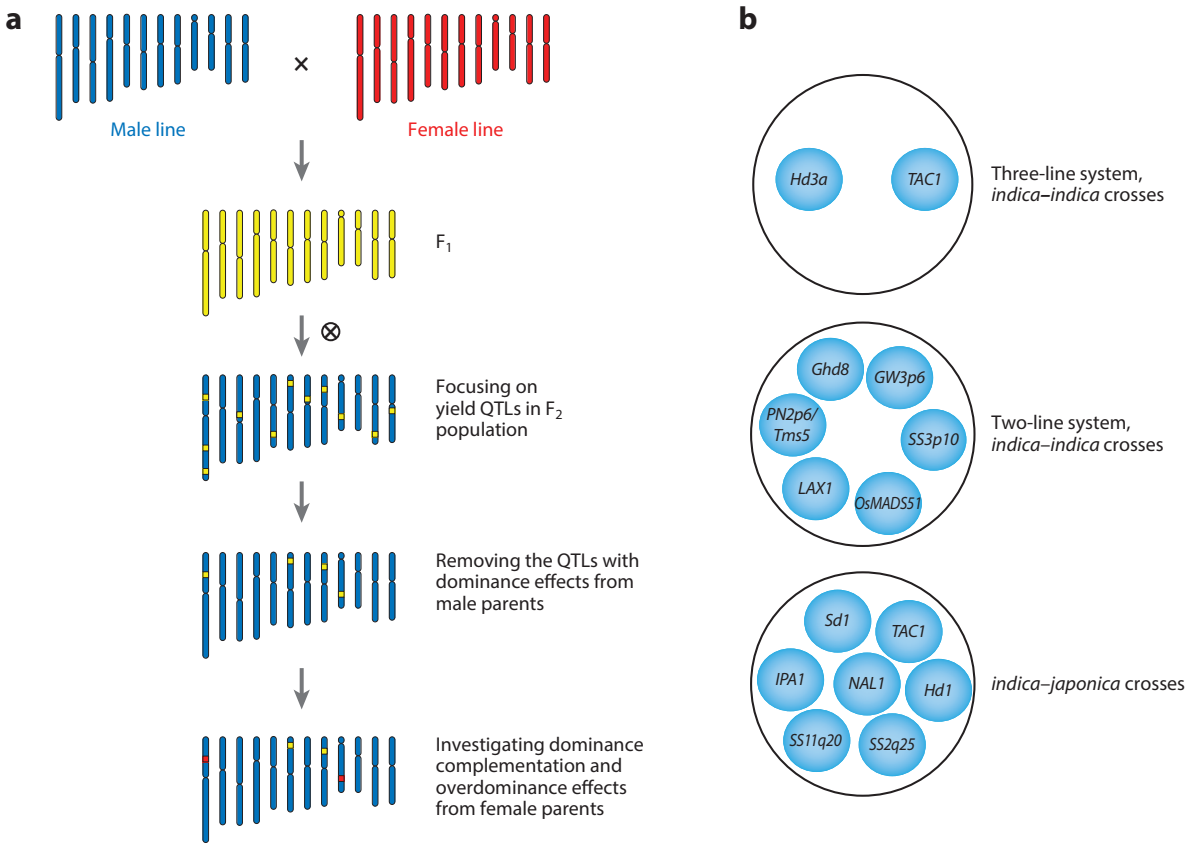

Figure 3

A genomic approach for mapping and investigating genetic effects of heterotic loci. (*a*) Crossing an elite parent (inbreeding male line; homozygous genotypes indicated in *blue*) with the alternative parent (inbreeding female line; homozygous genotypes indicated in *red*) generates the F_1 offspring (heterozygous genotypes indicated in *yellow*). To identify quantitative trait loci (QTLs) related to yield heterosis, the F_2 lines are generated, sequenced, genotyped, and phenotyped. The yield-related QTLs are mapped through a composite interval-mapping method, and the allele effect is estimated for each QTL. Since we assumed that other genomic regions except the QTLs have no or weak effects on grain yield, the genotypes of F_1 could be simplified as a chromosome segment substitution line; the inbred male line carries heterozygous genotypes only at the yield-related QTLs. Considering that heterozygous genotypes from the dominant QTLs contributed from the male parent showed phenotypes that were similar to those of homozygous male genotypes, the model is further simplified. The final model shows that dominant QTLs contributed from the female parent, as well as the overdominant QTLs, are the key heterotic loci in rice. (*b*) The lists of genomic loci or genes that explain a large proportion of the yield advantage of hybrids over their male parents in three-hybrid systems are shown: (*top*) a three-line system of *indica–indica* crosses: *Heading date 3a* (*Hd3a*) and *Tiller Angle Control 1* (*TAC1*) (74, 158); (*middle*) a two-line system of *indica–indica* crosses: *Grain number, plant height, and heading date 8* (*Ghd8*), *LAX PANICLE 1* (*LAX1*), *PN2p6/Thermo-sensitive genic male-sterile 5* (*Tms5*), *OsMADS51*, *SS3p10*, and *GW3p6* (72, 75, 152, 168); and (*bottom*) *indica–japonica* crosses existing mainly in *indica* or *japonica* parental lines: *semidwarf 1* (*Sd1*), *TAC1*, *NARROW LEAF 1* (*NAL1*), *Hd1*, *Ideal Plant Architecture 1* (*IPA1*), *SS11q20*, and *SS2q25* (26, 64, 103, 112, 157, 158, 162). *SS3p10*, *GW3p6*, *SS11q20*, and *SS2q25* are all uncharacterized QTLs.

advantageous effects of the heterotic loci (**Figure 3**), specifically in terms of yield heterosis, the second filial (F_2) populations from 17 representative hybrid rice crosses were also investigated (56). It is believed that F_2 populations generated from the elite hybrids (F_1) can provide useful information for heterosis analysis. Three genotypes (homozygous male parent, heterozygous, and homozygous female parent) will be present in the F_2 populations in a 1:2:1 ratio at the

local genomic region, which allows estimation of heterotic effects. Thus, 1,495 diverse hybrid rice varieties were collected, 17 representative hybrids selected from the large collections, and a total of 10,074 F_2 lines generated, sequenced, genotyped, and phenotyped from the 17 hybrids (56). The large amount of genomics and phenomics data from these well-designed populations helps to identify the heterosis-related genes in rice. The results suggest that modern hybrid rice varieties can be classified into three groups, representing different hybrid breeding systems. More importantly, these studies reveal that a small number of genomic loci from female parents explain a large proportion of the yield advantage hybrids have over their male parents (55, 56). For most of the heterosis-related loci identified, partial dominance of the heterozygous locus plays an important role in yield-related traits and better-parent heterosis in overall performance when all of the grain-yield traits are considered together (56).

For example, *Heading date 3a* (*Hd3a*), the ortholog in rice of the *Arabidopsis* gene *FLOWERING LOCUS T* (*FT*), which has been identified as one of the key genes for heterosis in the hybrids of *indica–indica* crosses (74), controlling both grain yield and flowering time in rice simultaneously. In the heterozygous state, *Hd3a* generally acted through incomplete dominance. When taking both grain yield and flowering time into account, the hybrids with the heterozygous *Hd3a* genotype showed an optimal combination of grain yield and flowering time (high yield and early flowering) that was better than that of both parents (high yield and late flowering for female parents or low yield and early flowering for male parents).

Applying these genetic findings, rice breeders will be able to improve combining ability by optimizing the cross designs using the molecular information provided by heterosis-related genes and their allelic distribution in germplasm collections. Moreover, the knowledge of heterosis-related QTLs or genes can also help to improve grain quality in hybrid rice. For example, many conventional parental lines in rice hybrids carried the low grain quality alleles of *waxy* (controlling amylose content and chalk grain rate) due to the genetic drag from *Hd3a* (close to the *waxy* gene on rice chromosome 6) (74, 141). Molecular breeding could be used to select the recombinants with both the high grain quality allele *waxy* and high grain yield allele *Hd3a*.

In terms of the molecular mechanisms, there remain some substantial holes in explaining the full heterotic effects. Rice breeders still do not know in detail what the exact reasons for the overdominance phenomenon are, or why the dominant effects are more frequent than the recessive effects in rice (and in maize and many other hybrid crops). For overdominance, the intermediate gene activity (gene dosage effects) is likely one reason, but the detailed molecular mechanisms at work may be much more complicated. We will need to recognize the heterosis phenomenon from the evolutionary view in rice. In summary, an integrated genomic framework that exploits population-scale genomic landscapes from a representative number of hybrid rice varieties will provide new insights into the principles of hybrid vigor and have implications for rice breeding.

5. GENETIC IMPROVEMENT AND FUTURE RICE BREEDING

5.1. Characterization of Rice QTLs Underlying Complex Agronomic Traits

With in-depth insight into diverse rice genomes, molecular cloning of agronomic QTLs that confer grain yield–related traits has been assisted mainly by GWAS, QTL mapping, and mutant analysis in the past decades (10, 11, 54, 58) (**Table 1**). The dominant yield-related trait QTLs (improvement or diversification genes) for the aerial parts (e.g., flowering time, plant architecture, and grain size) of rice have been largely exploited, whereas those for the underground parts (e.g., root absorption and iron uptake) also deserve consideration (73). QTL cloning and functional analysis will be helpful to understand gene–gene interactions in specific pathways.

semidwarf 1 (*sd1* or *OsGA20ox2*) has been reported as encoding an oxidase enzyme involved in the biosynthesis of gibberellin, reducing plant height by defects in the hormone's signaling pathway (112). Widespread adoption of semidwarf rice cultivars has brought the first Green Revolution to enhance harvest index and environmental adaptation in crop breeding history. Thus, improving crop resistance to environmental stress factors, in particular, blast disease, planthoppers, and unpredictable climate, also plays an important determinant in rice production (80). Most recently, epigenetic regulation between two paired antagonistic receptors, *PigmR* and *PigmS*, suggests a practical basis for the balance of blast disease against grain yield (16). *Bph6* confers resistance to planthoppers and has great potential for the development of elite crop varieties to control agricultural insect pests (40). Wild rice has a higher genetic diversity in terms of exploiting more yield- and stress tolerance–related QTLs, which may provide a natural allele pool to utilize.

5.2. Hybrid Rice and Intersubspecific Hybrid Incompatibility

The new potential for green high-yield rice breeding is dependent on ideal plant architecture, superior grain quality, and lodging resistance (163). In the past decades, hybrid rice has contributed greatly to Chinese agriculture, which represents more than half of total rice cultivation and has displayed greater production than that offered with elite inbred varieties. The success of hybrid rice is due in large part to the utilization of thermo- or photoperiod-sensitive male sterility lines and cytoplasmic male sterility (CMS) lines in three- and two-line systems (17, 151, 165).

Hybrid incompatibility (e.g., hybrid sterility, inviability, breakdown, and weakness) and low genetic diversity are two main genetic barriers in modern rice breeding (94). Previous studies have suggested that genetic diversity was lower in CMS/A and restorer (R) lines for three-line hybrid rice (48). For CMS, a large number of sterility and restorer of fertility (*Rf*) genes have been identified and many main CMS and R lines developed from them. Compared to *indica–indica* or *japonica–japonica* hybrids, *indica–japonica* hybrids have the highest yield potentials due to greater genetic divergence (79). On the other hand, wide-compatibility (WC), conferred by S_5^n, is recognized as an important trait and can overcome the fertility barrier in the *indica–japonica* hybrids (153). However, *qHMS7*, a selfish genetic element gene, confers hybrid male sterility, which suggests a toxin/antidote system to maintain genome stability between wild and cultivated rice (159).

5.3. Genome Editing for Crop Design

Sequence-specific nucleases have been developed as an effective tool to perform genome editing (CRISPR/Cas9 or CRISPR/Cpf1) (2, 8, 68). Due to such rapid innovations in genome editing, gene modification has greatly contributed to improving crop yield and disease resistance (3, 6, 23, 90, 102, 114, 154). One successful example is an application for controlling tomato productivity, which demonstrates a useful way of using CRISPR/Cas9 genome editing to generate diverse *cis*-regulatory alleles (5). By screening the phenotype of a large number of progeny through crossing the editing cassette to normal lines, it was possible to induce diverse alleles to investigate gene expression (109). This finding may provide a foundation for future main crop precision breeding assisted by genome editing.

In rice, 474 genomic regions contributing to heterosis have been identified through hybrid populations, and male parents (restorer lines) contributed more beneficial alleles than female parents in terms of the yield performance of F1 lines (56). The heterosis-related QTLs *Hd3a*

and *TILLER ANGLE CONTROL1* (*TAC1*) were the main contributors that distributed mostly in restorer lines, and the performance of *Hd3a/hd3a* and *TAC1/tac1* was better than that of either of their homozygous genotypes (74, 158). The identification of *qWS8/ipa1–2D* has revealed a practical approach to manipulate expression of the key locus *ideal plant architecture*1 (*IPA1*) by fine-tuning its traits (162). Meanwhile, the main restorer lines (male parent) in both the three- and two-line systems performed far better than female parent lines, contributing many beneficial grain-yield QTL alleles. It is feasible to modify the key heterosis locus on restorer lines or superior varieties assisted by gene modification.

5.4. Natural Variants and Genome-Wide Design

Efforts to increase crop production have faced limitations in traditional breeding practices. Thus, advances in breeding assisted by pan-genome analysis and GWAS will accelerate progress in this area. Pan-genome analysis provides a whole diverse allele data set of key genes from different natural variants across the whole genome (166). Modern crop breeding is focusing on combinations of multiple beneficial alleles that determine the good agronomic traits in the rice genome (106). How to effectively utilize these genes will be the great challenge in facilitating rice breeding.

A genomic map for 1,495 elite hybrid rice varieties and their inbred lines has been constructed through an integrated genomic approach. What's more, numerous beneficial alleles that contribute to heterosis have been revealed (55). Improving grain quality and nutrients is still an obstacle to pyramiding superior-quality varieties. *Chalk5* encodes a vacuolar H^+-translocating pyrophosphatase and is the first cloned and functionally characterized gene that controls rice grain chalkiness (89). By using multiple collaborative populations of 10,074 F_2 lines in hybrid rice, the genetic basis of grain shape and chalkiness traits has been well dissected, and results indicate that *GW5*, *ALK*, and *Waxy* control chalky grain rate and also play a major role in endosperm development (29, 35, 117, 141, 142). In addition, a perspective on the advantages of breeding high-yield, superior-quality hybrid super rice by rational design has been suggested (160). To further improve grain yield and overcome the pressures of population growth and severe climate, the genetic basis of agronomically important complex traits should be well characterized through an integrated genomic approach.

6. CONCLUSIONS

In this article, we reviewed comprehensive studies on sequencing and comparative analysis of *Oryza* species genomes and factors underlying the relationships and genome structural differences seen in rice evolution and domestication. Clearly, cultivated rice has a physiologically and phenotypically striking difference compared to its progenitor wild rice, which is dominated by standing domestication traits. A population genomic study on *O. sativa–O. rufipogon* genome complex is a powerful approach to reveal a single origin and multiple introgressions of the rice domestication processes. Recent research on the domestication of African rice illustrates a distinct but parallel evolutionary event different from the development of Asian rice. Pan-genomic research on *O. sativa* and *O. rufipogon* species will facilitate the use of diverse natural alleles for future rice breeding. New breeding practices assisted by precision gene modification of key heterotic QTLs on superior parent lines will optimize traits and develop the ideal phenotype. Identifying those genes conferring agronomically important traits assisted by pan-genome analysis and GWAS will optimize rice breeding practices.

FUTURE ISSUES

1. Genome sequencing analysis of more wild species in the *Oryza* genus, assisted by advanced sequencing methods, and further investigation of the evolutionary relationship across the whole genome are needed.
2. Exploiting and discovering more environment-related QTLs assisted by pan-genome analysis and GWAS will facilitate development of new elite rice varieties to adapt to various environments.
3. Further characterizing new selective sweeps and the continued use of functional analysis to study domestication genes will be helpful in uncovering the mystery of rice domestication.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grants 31788103 and 31630055 to B.H.), Shanghai Engineering Research Center of Plant Germplasm Resources (grant 17DZ2252700 to X.H.), and the Chinese Academy of Sciences (grant XDPB0400 to B.H.). We thank Zhoulin Gu and Qiang Zhao (National Center for Gene Research) for critical discussions and input.

LITERATURE CITED

1. Ammiraju JS, Lu F, Sanyal A, Yu Y, Song X, et al. 2008. Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* 20:3191–209
2. Arora L, Narula A. 2017. Gene editing and crop improvement using CRISPR-Cas9 system. *Front. Plant Sci.* 8:1932
3. Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V. 2013. Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods* 9:39
4. Bessho-Uehara K, Wang DR, Furuta T, Minami A, Nagai K, et al. 2016. Loss of function at *RAE2*, a previously unidentified EPFL, is required for awnlessness in cultivated Asian rice. *PNAS* 113:8969–74
5. Birchler JA. 2017. Editing the phenotype: a revolution for quantitative genetics. *Cell* 171:269–70
6. Borel B. 2017. CRISPR, microbes and more are joining the war against crop killers. *Nature* 543:302–4
7. Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, et al. 2007. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLOS Genet.* 3:1745–56
8. Cermak T, Curtin SJ, Gil-Humanes J, Cegan R, Kono TJY, et al. 2017. A multipurpose toolkit to enable advanced genome engineering in plants. *Plant Cell* 29:1196–217
9. Chen J, Huang Q, Gao D, Wang J, Lang Y, et al. 2013. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat. Commun.* 4:1595
10. Chen W, Gao Y, Xie W, Gong L, Lu K, et al. 2014. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46:714–21
11. Chen W, Wang W, Peng M, Gong L, Gao Y, et al. 2016. Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nat. Commun.* 7:12767

12. Cheng KS, Oka HI. 1986. Classification of rice cultivars into *Hsien* and *Keng* types. *Jpn. J. Breed.* 36(Ext. 2):14–15
13. Cheng SH, Zhuang JY, Fan YY, Du JH, Cao LY. 2007. Progress in research and development on hybrid rice: a super-domesticate in China. *Ann. Bot.* 100:959–66
14. Choi JY, Platts AE, Fuller DQ, Hsing YI, Wing RA, Purugganan MD. 2017. The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* 34:969–79
15. da Silva FG, Shen Y, Dardick C, Burdman S, Yadav RC, et al. 2004. Bacterial genes involved in type I secretion and sulfation are required to elicit the rice *Xa21*-mediated innate immune response. *Mol. Plant Microbe Interact.* 17:593–601
16. Deng Y, Zhai K, Xie Z, Yang D, Zhu X, et al. 2017. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* 355:962–65
17. Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, et al. 2012. A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *PNAS* 109:2654–59
18. Ding Y. 1949. [The origin of rice cultivation in China]. *Agron. Bull. Nat. Zhong Shang Univ.* 7:1–18 (In Chinese)
19. Doebley JF. 2004. The genetics of maize evolution. *Annu. Rev. Genet.* 38:37–59
20. Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127:1309–21
21. **Doebley JF, Stec A, Hubbard L. 1997. The evolution of apical dominance in maize. *Nature* 386:485–88**
22. Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, et al. 2010. Crops gone wild: evolution of weeds and invasives from domesticated ancestors. *Evol. Appl.* 3:494–504
23. Filler Hayut S, Melamed Bessudo C, Levy AA. 2017. Targeted recombination between homologous chromosomes for precise breeding in tomato. *Nat. Commun.* 8:15605
24. Food Agric. Organ. U.N. (FAO). 2009. *Proceedings of the Expert Meeting on How to Feed the World in 2050*. Rome: FAO
25. Fournier-Level A, Perry EO, Wang JA, Braun PT, Migneault A, et al. 2016. Predicting the evolutionary dynamics of seasonal adaptation to novel climates in *Arabidopsis thaliana*. *PNAS* 113:2812–21
26. Fujita D, Trijatmiko KR, Tagle AG, Sapasap MV, Koide Y, et al. 2013. *NAL1* allele from a rice landrace greatly increases yield in modern *indica* cultivars. *PNAS* 110:20431–36
27. Fuller DQ. 2007. Contrasting patterns in crop domestication and domestication rates: recent archaeobotanical insights from the Old World. *Ann. Bot.* 100:903–24
28. Gamuyao R, Chin JH, Pariasca-Tanaka J, Pesaresi P, Catausan S, et al. 2012. The protein kinase *Pstol1* from traditional rice confers tolerance of phosphorus deficiency. *Nature* 488:535–39
29. Gao ZY, Zeng D, Cui X, Zhou Y, Yan M, et al. 2003. Map-based cloning of the *ALK* gene, which controls the gelatinization temperature of rice. *Sci. China C Life Sci.* 46:661–68
30. Gao ZY, Zhao SC, He WM, Guo LB, Peng YL, et al. 2013. Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *PNAS* 110:14492–97
31. **Ge S, Sang T, Lu BR, Hong DY. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *PNAS* 96:14400–5**
32. Glaszmann JC. 1987. Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* 74:21–30
33. Gnan S, Priest A, Kover PX. 2014. The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. *Genetics* 198:1751–58
34. Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, et al. 2010. Food security: the challenge of feeding 9 billion people. *Science* 327:812–18
35. Gong J, Miao J, Zhao Y, Zhao Q, Feng Q, et al. 2017. Dissecting the genetic basis of grain shape and chalkiness traits in hybrid rice using multiple collaborative populations. *Mol. Plant* 10:1353–56
36. Gross BL, Olsen KM. 2010. Genetic perspectives on crop domestication. *Trends Plant Sci.* 15:529–37
37. Gross BL, Reagon M, Hsu SC, Caicedo AL, Jia Y, Olsen KM. 2010. Seeing red: the origin of grain pigmentation in US weedy rice. *Mol. Ecol.* 19:3380–93
38. Gross BL, Steffen FT, Olsen KM. 2010. The molecular basis of white pericarps in African domesticated rice: novel mutations at the *Rc* gene. *J. Evol. Biol.* 23:2747–53

21. Characterizes *tb1* as a major contributor in increasing apical dominance in maize when compared with its probable wild ancestor *teosinte*, suggesting that gene regulatory changes underlie the evolutionary divergence.

31. Shows that the evolutionary relationships among rice species are inferred from *Adh1*, *Adh2*, and *matK*, especially for origins of allotetraploid species.

39. Gu B, Zhou T, Luo J, Liu H, Wang Y, et al. 2015. *An-2* encodes a cytokinin synthesis enzyme that regulates awn length and grain production in rice. *Mol. Plant* 8:1635–50
40. Guo J, Xu C, Wu D, Zhao Y, Qiu Y, et al. 2018. *Bph6* encodes an exocyst-localized protein and confers broad resistance to planthoppers in rice. *Nat. Genet.* 50:297–306
41. Guo YL, Ge S. 2005. Molecular phylogeny of *Oryzeae* (*Poaceae*) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am. J. Bot.* 92:1548–58
42. Han B, Huang X. 2013. Sequencing-based genome-wide association study in rice. *Curr. Opin. Plant Biol.* 16:133–38
43. Hancock JF. 2005. Contributions of domesticated plant studies to our understanding of plant evolution. *Ann. Bot.* 96:953–63
44. Hattori Y, Nagai K, Furukawa S, Song XJ, Kawano R, et al. 2009. The ethylene response factors *SNORKEL1* and *SNORKEL2* allow rice to adapt to deep water. *Nature* 460:1026–30
45. Hollick JB, Chandler VL. 1998. Epigenetic allelic states of a maize transcriptional regulatory locus exhibit overdominant gene action. *Genetics* 150:891–97
46. Hu B, Wang W, Ou S, Tang J, Li H, et al. 2015. Variation in *NRT1.1B* contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* 47:834–38
47. Hu M, Lv S, Wu W, Fu Y, Liu F, et al. 2018. The domestication of plant architecture in African rice. *Plant J.* 94:661–69
48. Hu YY, Mao BG, Peng Y, Sun YD, Pan YL, et al. 2014. Deep re-sequencing of a widely used maintainer line of hybrid rice for discovery of DNA polymorphisms and evaluation of genetic diversity. *Mol. Genet. Genom.* 289:303–15
49. Hua L, Wang DR, Tan L, Fu Y, Liu F, et al. 2015. *LABA1*, a domestication gene associated with long, barbed awns in wild rice. *Plant Cell* 27:1875–88
50. Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang Q. 2002. Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics* 162:1885–95
51. Huang X, Han B. 2015. Rice domestication occurred through single origin and multiple introgressions. *Nat. Plants* 2:15207
52. Huang X, Kurata N, Wei X, Wang ZX, Wang A, et al. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497–501
53. Huang X, Qian Q, Liu Z, Sun H, He S, et al. 2009. Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat. Genet.* 41:494–97
54. Huang X, Wei X, Sang T, Zhao Q, Feng Q, et al. 2010. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42:961–67
55. Huang X, Yang S, Gong J, Zhao Y, Feng Q, et al. 2015. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6:6258
56. Huang X, Yang S, Gong J, Zhao Q, Feng Q, et al. 2016. Genomic architecture of heterosis for yield traits in rice. *Nature* 537:629–33
57. Huang X, Zhao Q, Han B. 2015. Comparative population genomics reveals strong divergence and infrequent introgression between Asian and African rice. *Mol. Plant* 8:958–60
58. Huang X, Zhao Y, Wei X, Li C, Wang A, et al. 2011. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* 44:32–39
59. Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, et al. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44:808–11
60. Int. Rice Genome Seq. Proj., Sasaki T. 2005. The map-based sequence of the rice genome. *Nature* 436:793–800
61. Ishii T, Numaguchi K, Miura K, Yoshida K, Thanh PT, et al. 2013. *OsLG1* regulates a closed panicle trait in domesticated rice. *Nat. Genet.* 45:462–65
62. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, et al. 2013. Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat. Genet.* 45:707–11
63. Jiang Z, Xia H, Basso B, Lu BR. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theor. Appl. Genet.* 124:309–22

52. Uncovers the origin of Asian domesticated rice based on a gene variation map.

54. Constructs the first rice haplotype map (HapMap) used for genome-wide association studies in rice.

56. Provides insights into the genomic architecture of heterosis by using different hybrid breeding systems.

64. Jiao Y, Wang Y, Xue D, Wang J, Yan M, et al. 2010. Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42:541–44
65. Jin J, Hua L, Zhu Z, Tan L, Zhao X, et al. 2016. *GAD1* Encodes a secreted peptide that regulates grain number, grain length, and awn development in rice domestication. *Plant Cell* 28:2453–63
66. Jin J, Huang W, Gao JP, Yang J, Shi M, et al. 2008. Genetic control of rice plant architecture under domestication. *Nat. Genet.* 40:1365–69
67. Kane NC, Rieseberg LH. 2008. Genetics and evolution of weedy *Helianthus annuus* populations: adaptation of an agricultural weed. *Mol. Ecol.* 17:384–94
68. Kang BC, Yun JY, Kim ST, Shin Y, Ryu J, et al. 2018. Precision genome engineering through adenine base editing in plants. *Nat. Plants* 4(7):427–31. Erratum. 2018. *Nat. Plants* 4(9):730
69. Kato S. 1930. On the affinity of the cultivated varieties of rice plants, *Oryza sativa* L. *J. Dep. Agric. Kyushu Imp. Univ.* 2:241–76
70. Khush GS. 1997. Origin, dispersal, cultivation and variation of rice. *Plant Mol. Biol.* 35:25–34
71. Kim H, Jung J, Singh N, Greenberg A, Doyle JJ, et al. 2016. Population dynamics among six major groups of the *Oryza rufipogon* species complex, wild relative of cultivated Asian rice. *Rice* 9:56
72. Kim SL, Lee S, Kim HJ, Nam HG, An G. 2007. *OsMADS51* is a short-day flowering promoter that functions upstream of *Ehd1*, *OsMADS14*, and *Hd3a*. *Plant Physiol.* 145:1484–94
73. Klee HJ. 2017. Genetic control of floral architecture: insights into improving crop yield. *Cell* 169:983–84
74. Kojima S, Takahashi Y, Kobayashi Y, Monna L, Sasaki T, et al. 2002. *Hd3a*, a rice ortholog of the *Arabidopsis FT* gene, promotes transition to flowering downstream of *Hd1* under short-day conditions. *Plant Cell Physiol.* 43:1096–105
75. Komatsu M, Maekawa M, Shimamoto K, Kyozuka J. 2001. The *LAX1* and *FRIZZY PANICLE 2* genes determine the inflorescence architecture of rice by controlling rachis-branch and spikelet development. *Dev. Biol.* 231:364–73
76. Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, et al. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312:1392–96
77. Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, et al. 2009. a multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLOS Genet.* 5:e1000551
78. Krieger U, Lippman ZB, Zamir D. 2010. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nat. Genet.* 42:459–63
79. Kubo T, Yoshimura A. 2005. Epistasis underlying female sterility detected in hybrid breakdown in a Japonica–Indica cross of rice (*Oryza sativa* L.). *Theor. Appl. Genet.* 110:346–55
80. Kuroha T, Nagai K, Gamuyao R, Wang DR, Furuta T, et al. 2018. Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Science* 361:181–86
81. Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, et al. 2014. Current perspectives and the future of domestication studies. *PNAS* 111:6139–46
82. Li C, Sun B, Li Y, Liu C, Wu X, et al. 2016. Numerous genetic loci identified for drought tolerance in the maize nested association mapping populations. *BMC Genom.* 17:894
83. Li C, Zhou A, Sang T. 2006. Rice domestication by reducing shattering. *Science* 311:1936–39
84. Li LF, Li YL, Jia Y, Caicedo AL, Olsen KM. 2017. Signatures of adaptation in the weedy rice genome. *Nat. Genet.* 49:811–14
85. Li X, Guo T, Mu Q, Li X, Yu J. 2018. Genomic and environmental determinants and their interplay underlying phenotypic plasticity. *PNAS* 115:6679–84
86. Li X, Li X, Fridman E, Tesso TT, Yu J. 2015. Dissecting repulsion linkage in the dwarfing gene *Dw3* region for sorghum plant height provides insights into heterosis. *PNAS* 112:11823–28
87. Li XM, Chao DY, Wu Y, Huang X, Chen K, et al. 2015. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. *Nat. Genet.* 47:827–33
88. Li Y, Fan C, Xing Y, Jiang Y, Luo L, et al. 2011. Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43:1266–69
89. Li Y, Fan C, Xing Y, Yun P, Luo L, et al. 2014. *Chalk5* encodes a vacuolar H^+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* 46:398–404

84. Proposes a dedomestication model of weed rice in the coevolution of cultivated and wild rice.

85. Reveals that the genetic effect continuum across the environmental gradient and gene–gene interaction is contributing to phenotypic plasticity.

90. Liang G, Zhang H, Lou D, Yu D. 2016. Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Sci. Rep.* 6:21451
91. Linares OF. 2002. African rice (*Oryza glaberrima*): history and future potential. *PNAS* 99:16360–65
92. Liu J, Chen J, Zheng X, Wu F, Lin Q, et al. 2017. *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat. Plants* 3:17043
93. Londo JP, Schaal BA. 2007. Origins and population genetics of weedy red rice in the USA. *Mol. Ecol.* 16:4523–35
94. Long Y, Zhao L, Niu B, Su J, Wu H, et al. 2008. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *PNAS* 105:18871–76
95. Lu BR, Yang X, Ellstrand NC. 2016. Fitness correlates of crop transgene flow into weedy populations: a case study of weedy rice in China and other examples. *Evol. Appl.* 9:857–70
96. Luo D, Xu H, Liu Z, Guo J, Li H, et al. 2013. A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45:573–77
97. Luo J, Liu H, Zhou T, Gu B, Huang X, et al. 2013. *An-1* encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell* 25:3360–76
98. Lv S, Wu W, Wang M, Meyer RS, Ndjiondjop MN, et al. 2018. Genetic control of seed shattering during African rice domestication. *Nat. Plants* 4:331–37
99. Merotto A Jr., Goulart IC, Nunes AL, Kalsing A, Markus C, et al. 2016. Evolutionary and social consequences of introgression of nontransgenic herbicide resistance from rice to weedy rice in Brazil. *Evol. Appl.* 9:837–46

100. Indicates a protracted period of population size reduction in the predomestication phase and adaptive geographical divergence for salt tolerance of African rice.

100. Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, et al. 2016. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* 48:1083–88

101. Meyer RS, Purugganan MD. 2013. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14:840–52
102. Miki D, Zhang W, Zeng W, Feng Z, Zhu JK. 2018. CRISPR/Cas9-mediated gene targeting in *Arabidopsis* using sequential transformation. *Nat. Commun.* 9:1967
103. Miura K, Ikeda M, Matsubara A, Song XJ, Ito M, et al. 2010. *OsSPL14* promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* 42:545–49
104. Park CJ, Ronald PC. 2012. Cleavage and nuclear localization of the rice XA21 immune receptor. *Nat. Commun.* 3:920
105. Purugganan MD. 2014. An evolutionary genomic tale of two rice species. *Nat. Genet.* 46:931–32
106. Qian Q, Guo LB, Smith SM, Li JY. 2016. Breeding high-yield superior quality hybrid super rice by rational design. *Natl. Sci. Rev.* 3:283–94
107. Qiu J, Zhou Y, Mao L, Ye C, Wang W, et al. 2017. Genomic variation associated with local adaptation of weedy rice during de-domestication. *Nat. Commun.* 8:15323
108. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, et al. 2012. Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* 44:217–20

109. Provides foundation for unraveling complex relationships between gene-regulatory changes and control of quantitative traits by genome editing.

109. Rodriguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB. 2017. Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171:470–80.e8

110. Ross-Ibarra J, Morrell PL, Gaut BS. 2007. Plant domestication, a unique opportunity to identify the genetic basis of adaptation. *PNAS* 104(Suppl. 1):8641–48
111. Sang T, Ge S. 2007. Genetics and phylogenetics of rice domestication. *Curr. Opin. Genet. Dev.* 17:533–38
112. Sasaki A, Ashikari M, Ueguchi-Tanaka M, Itoh H, Nishimura A, et al. 2002. Green revolution: a mutant gibberellin-synthesis gene in rice. *Nature* 416:701–2
113. Second G. 1985. Evolutionary relationships in the *Sativa* group of *Oryza* based on isozyme data. *Genet. Sel. Evol.* 17:89–114
114. Shan Q, Wang Y, Li J, Zhang Y, Chen K, et al. 2013. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31:686–88
115. Shimamoto K, Kyozuka J. 2002. Rice as a model for comparative genomics of plants. *Annu. Rev. Plant Biol.* 53:399–419
116. Shivrain VK, Burgos NR, Gealy DR, Sales MA, Smith KL. 2009. Gene flow from weedy red rice (*Oryza sativa* L.) to cultivated rice and fitness of hybrids. *Pest Manag. Sci.* 65:1124–29

117. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, et al. 2008. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* 40:1023–28
118. Si L, Chen J, Huang X, Gong H, Luo J, et al. 2016. *OsSPL13* controls grain size in cultivated rice. *Nat. Genet.* 48:447–56
119. Simons KJ, Fellers JP, Trick HN, Zhang Z, Tai YS, et al. 2006. Molecular characterization of the major wheat domestication gene *Q*. *Genetics* 172:547–55
120. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39:623–30
121. Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, et al. 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* 50:285–96
122. Stich B. 2009. Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* 183:1525–34
123. Sun H, Qian Q, Wu K, Luo J, Wang S, et al. 2014. Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Genet.* 46:652–56
124. Sun J, Qian Q, Ma DR, Xu ZJ, Liu D, et al. 2013. Introgression and selection shaping the genome and adaptive loci of weedy rice in northern China. *New Phytol.* 197:290–99
125. Sweeney M, McCouch S. 2007. The complex history of the domestication of rice. *Ann. Bot.* 100:951–57
126. Tan L, Li X, Liu F, Sun X, Li C, et al. 2008. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* 40:1360–64
127. The 3,000 Rice Genomes Proj. 2014. The 3,000 Rice Genomes Project. *Gigascience* 3:7
128. Uga Y, Okuno K, Yano M. 2011. *Dro1*, a major QTL involved in deep rooting of rice under upland field conditions. *J. Exp. Bot.* 62:2485–94
129. Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, et al. 2013. Control of root system architecture by *DEEPER ROOTING 1* increases rice yield under drought conditions. *Nat. Genet.* 45:1097–102
130. van Andel TR, Meyer RS, Aflitos SA, Carney JA, Veltman MA, et al. 2016. Tracing ancestor rice of Suriname Maroons back to its African origin. *Nat. Plants* 2:16149
131. Vigueira CC, Li W, Olsen KM. 2013. The role of *Bh4* in parallel evolution of hull colour in domesticated and weedy rice. *J. Evol. Biol.* 26:1738–49
132. Wambugu PW, Brozynska M, Furtado A, Waters DL, Henry RJ. 2015. Relationships of wild and domesticated rice (*Oryza* AA genome species) based upon whole chloroplast genome sequences. *Sci. Rep.* 5:13957
133. Wambugu PW, Furtado A, Waters DL, Nyamongo DO, Henry RJ. 2013. Conservation and utilization of African *Oryza* genetic resources. *Rice* 6:29
134. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, et al. 2005. The origin of the naked grains of maize. *Nature* 436:714–19
135. Wang M, Yu Y, Haberer G, Marri PR, Fan C, et al. 2014. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* 46:982–88
136. Wang S, Li S, Liu Q, Wu K, Zhang J, et al. 2015. The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* 47:949–54
137. Wang S, Wu K, Yuan Q, Liu X, Liu Z, et al. 2012. Control of grain size, shape and quality by *OsSPL16* in rice. *Nat. Genet.* 44:950–54
138. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, et al. 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43–49
139. Wang Y, Xiong G, Hu J, Jiang L, Yu H, et al. 2015. Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat. Genet.* 47:944–48
140. Wang Y, Xue Y, Li J. 2005. Towards molecular breeding and improvement of rice in China. *Trends Plant Sci.* 10:610–14
141. Wang ZY, Zheng FQ, Shen GZ, Gao JP, Snustad DP, et al. 1995. The amylose content in rice endosperm is related to the post-transcriptional regulation of the *waxy* gene. *Plant J.* 7:613–22
142. Weng J, Gu S, Wan X, Gao H, Guo T, et al. 2008. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res.* 18:1199–209

135. Provides high-quality assembly and annotation of the *O. glaberrima* genome and reveals that *O. glaberrima* is domesticated in a single region.

143. Wing RA, Purugganan MD, Zhang Q. 2018. The rice genome revolution: from an ancient grain to green super rice. *Nat. Rev. Genet.* 19:505–17
144. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, et al. 2005. The effects of artificial selection on the maize genome. *Science* 308:1310–14
145. Wright SI, Gaut BS. 2005. Molecular population genetics and the search for adaptive evolution in plants. *Mol. Biol. Evol.* 22:506–19
146. Wu W, Liu X, Wang M, Meyer RS, Luo X, et al. 2017. A single-nucleotide polymorphism causes smaller grain size and loss of seed shattering during African rice domestication. *Nat. Plants* 3:17064
147. Xiao J, Li J, Yuan L, Tanksley SD. 1995. Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics* 140:745–54
148. Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, et al. 2006. *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* 442:705–8
149. Xu X, Liu X, Ge S, Jensen JD, Hu F, et al. 2011. Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30:105–11
150. Xue W, Xing Y, Weng X, Zhao Y, Tang W, et al. 2008. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40:761–67
151. Yamagata Y, Yamamoto E, Aya K, Win KT, Doi K, et al. 2010. Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *PNAS* 107:1494–99
152. Yan WH, Wang P, Chen HX, Zhou HJ, Li QP, et al. 2011. A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Mol. Plant* 4:319–30
153. Yang J, Zhao X, Cheng K, Du H, Ouyang Y, et al. 2012. A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science* 337:1336–40
154. Yang N, Wang R, Zhao Y. 2017. Revolutionize genetic studies and crop improvement with high-throughput and genome-scale CRISPR/Cas9 gene editing technology. *Mol. Plant* 10:1141–43
155. Yang X, Xia H, Wang W, Wang F, Su J, et al. 2011. Transgenes for insect resistance reduce herbivory and enhance fecundity in advanced generations of crop-weed hybrids of rice. *Evol. Appl.* 4:672–84
156. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, et al. 2016. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48:927–34
157. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, et al. 2000. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell* 12:2473–84
158. Yu B, Lin Z, Li H, Li X, Li J, et al. 2007. *TAC1*, a major quantitative trait locus controlling tiller angle in rice. *Plant J.* 52:891–98
159. Yu X, Zhao Z, Zheng X, Zhou J, Kong W, et al. 2018. A selfish genetic element confers non-Mendelian inheritance in rice. *Science* 360:1130–32
160. Zeng D, Tian Z, Rao Y, Dong G, Yang Y, et al. 2017. Rational design of high-yield and superior-quality rice. *Nat. Plants* 3:17031
161. Zhang J, Chen LL, Xing F, Kudrna DA, Yao W, et al. 2016. Extensive sequence divergence between the reference genomes of two elite *indica* rice varieties Zhenshan 97 and Minghui 63. *PNAS* 113:5163–71
162. Zhang L, Yu H, Ma B, Liu G, Wang J, et al. 2017. A natural tandem array alleviates epigenetic repression of *IPA1* and leads to superior yielding rice. *Nat. Commun.* 8:14789
163. Zhang Q. 2007. Strategies for developing Green Super Rice. *PNAS* 104:16402–9
164. Zhang QJ, Zhu T, Xia EH, Shi C, Liu YL, et al. 2014. Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *PNAS* 111:4954–62
165. Zhang Y, Zhang S, Liu H, Fu B, Li L, et al. 2015. Genome and comparative transcriptomics of African wild rice *Oryza longistaminata* provide insights into molecular mechanism of rhizomatousness and self-incompatibility. *Mol. Plant* 8:1683–86
166. Zhao Q, Feng Q, Lu H, Li Y, Wang A, et al. 2018. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* 50:278–84
167. Zhou G, Chen Y, Yao W, Zhang C, Xie W, et al. 2012. Genetic composition of yield heterosis in an elite rice hybrid. *PNAS* 109:15847–52

162. Indicates that different IPA traits are determined in a dosage-dependent manner and that fine-tuned expression of *IPA1* could optimize plant architecture and improve rice yield.

166. Constructs a pan-genome data set of the *O. sativa–O. rufipogon* species complex and identifies extent of the rice genome variations.

168. Zhou H, Zhou M, Yang Y, Li J, Zhu L, et al. 2014. RNase Z^{S1} processes Ub_{L40} mRNAs and controls thermosensitive genic male sterility in rice. *Nat. Commun.* 5:4884
169. Zhu BF, Si L, Wang Z, Zhou Y, Zhu J, et al. 2011. Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* 155:1301–11
170. Zou XH, Du YS, Tang L, Xu XW, Doyle JJ, et al. 2015. Multiple origins of BBCC allopolyploid species in the rice genus (*Oryza*). *Sci. Rep.* 5:14876
171. Zou XH, Yang Z, Doyle JJ, Ge S. 2013. Multilocus estimation of divergence times and ancestral effective population sizes of *Oryza* species and implications for the rapid diversification of the genus. *New Phytol.* 198:1155–64