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 Han¹

¹National 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: bh@negr.ac.cn
²University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100049, China
³College of Life Sciences, Shanghai Normal University, Shanghai 200234, China; email: xhhuang@shnu.edu.cn
⁴State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China
⁵Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, Arizona 85721, USA; email: rwing@email.arizona.edu

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 ............................................................ 640
2. EVOLUTION OF ORYZA SPECIES ............................................. 641
   2.1. Natural Diversity of the Genus Oryza ............................... 641
   2.2. Differentiation of O. sativa L. ssp. indica and ssp. japonica .......... 641
3. RICE DOMESTICATION .......................................................... 642
   3.1. Features of Domesticaion Traits ...................................... 642
   3.2. The Origin of Asian Cultivated Rice .................................. 644
   3.3. Independent Domestication of African Rice .......................... 645
   3.4. The Evolution of Weedy Rice .......................................... 651
4. NATURAL VARIATION AND PLANT ADAPTATION .......................... 651
   4.1. Rice Pan-Genome Analysis .............................................. 651
   4.2. Identifications of Genes for Plant Fitness by GWAS ................. 652
   4.3. Insights into Rice Heterosis ............................................ 653
5. GENETIC IMPROVEMENT AND FUTURE RICE BREEDING .................. 655
   5.1. Characterization of Rice QTLs Underlying Complex Agronomical Traits ...... 655
   5.2. Hybrid Rice and Intersubspecific Hybrid Incompatibility .............. 656
   5.3. Genome Editing for Crop Design ..................................... 656
   5.4. Natural Variants and Genome-Wide Design ............................ 657
6. CONCLUSIONS ............................................................... 657

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. longisistamineata, 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. longisistamineata are African species, with O. barthii being mainly distributed in West Africa, while O. longisistamineata 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 bisen (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.
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 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-III, 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), Bb4 (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

<table>
<thead>
<tr>
<th>Quantitative trait locus</th>
<th>Trait</th>
<th>Gene ID</th>
<th>Encoding protein</th>
<th>Biological function</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domestication genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seed shattering in chromosome 1 (qSH1)</td>
<td>Shattering</td>
<td>Os01g0848400</td>
<td>BEL1-type homeobox</td>
<td>Absence of abscission layer formation</td>
<td>76</td>
</tr>
<tr>
<td>Shattering 4 (sb4)</td>
<td>Shattering</td>
<td>Os04g0670900</td>
<td>Transcription factor</td>
<td>Absence of abscission layer formation</td>
<td>83</td>
</tr>
<tr>
<td>Awn-1 (An-1)</td>
<td>Awn length</td>
<td>Os04g0350700</td>
<td>Helix-loop-helix protein</td>
<td>Long awn and larger grains by regulating cell division</td>
<td>97</td>
</tr>
<tr>
<td>Awn-2 (An-2)/LONG AND BARBED AWN 1 (LABA1)</td>
<td>Awn length</td>
<td>Os04g0518800</td>
<td>Cytokinin-activating enzyme</td>
<td>Reducing the cytokinin concentration</td>
<td>39, 49</td>
</tr>
<tr>
<td>GRAIN NUMBER, GRAIN LENGTH AND AWN DEVELOPMENT 1 (GAD1)/REGULATOR OF AWN ELONGATION 2 (RAE2)</td>
<td>Awn length</td>
<td>Os08g0485500</td>
<td>Secretory signal peptide</td>
<td>Peptide signal</td>
<td>4, 65, 156</td>
</tr>
<tr>
<td>Black hull 4 (Bh-6)</td>
<td>Hull color</td>
<td>Os04g0460000</td>
<td>Amino acid transporter</td>
<td>Controlling seed hull color</td>
<td>169</td>
</tr>
<tr>
<td>Red pericarp (Ri)</td>
<td>Pericarp color</td>
<td>Os07g0211500</td>
<td>Basic helix-loop-helix protein</td>
<td>Positive regulator of proanthocyanidin synthesis</td>
<td>38</td>
</tr>
<tr>
<td>Liguleless gene 1 (OsLG1)</td>
<td>Panicle shape and ligule development</td>
<td>Os04g0656500</td>
<td>SQUAMOSA promoter-binding protein (SBP)</td>
<td>Morphological change of panicle shape</td>
<td>61</td>
</tr>
<tr>
<td>QTL for grain width and weight on chromosome 5 (GWS5)/GRAIN SIZE ON CHROMOSOME 5 (GSE5)/QTL for seed width on chromosome 5 (qSW5)</td>
<td>Grain width</td>
<td>Os05g0187500</td>
<td>Plasma membrane–associated protein with IQ domains</td>
<td>Interacts with OsGSK2 in the brassinosteroid signaling pathway</td>
<td>92, 117, 142</td>
</tr>
<tr>
<td>PROSTRATE GROWTH 1 (Prog1)</td>
<td>Tiller angle</td>
<td>Os07g0153600</td>
<td>Zinc-finger nuclear transcription factor</td>
<td>Regulates tiller angle and number of tillers</td>
<td>66, 126</td>
</tr>
<tr>
<td><strong>Improvement and diversification genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ideal Plant Architecture 1 (IPA1)/SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 14 (OsSPL14)/WEALTHY FARMER’S PANICLE (WFP)</td>
<td>Plant architecture</td>
<td>Os08g0509600</td>
<td>Transcription factor containing SBP-box</td>
<td>Controlling ideal plant architecture</td>
<td>64, 103, 162</td>
</tr>
</tbody>
</table>

(Continued)
### Table 1 (Continued)

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

(Continued)
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Quantitative trait locus</th>
<th>Trait</th>
<th>Gene ID</th>
<th>Encoding protein</th>
<th>Biological function</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QTL for days to heading on chromosome 8 (Ghd8)</td>
<td>Heading date</td>
<td>Os08g0174500</td>
<td>Putative HAP3 subunit of the CCAAT box-binding transcription factor</td>
<td>Delaying heading data under long-day conditions and also controlling plant height as well as number of grains per panicle</td>
<td>152</td>
</tr>
<tr>
<td>Heading date 1 (Hd1)</td>
<td>Heading date</td>
<td>Os06g0275000</td>
<td>Florigen</td>
<td>Promotion of heading under short-day conditions and in inhibition under long-day conditions</td>
<td>157</td>
</tr>
<tr>
<td>Semidwarf 1 (Sd1)</td>
<td>Plant height</td>
<td>Os01g0883800</td>
<td>The oxidase enzyme, GA20ox</td>
<td>Taking part in the biosynthesis of gibberellin</td>
<td>80, 112</td>
</tr>
<tr>
<td>Thermosensitive male sterility 5 (Tms5)</td>
<td>Thermosensitive male sterility</td>
<td>Os02g0214300</td>
<td>RNase Z protein</td>
<td>Processing male sterility–related Ht,60 mRNAs into fragments</td>
<td>168</td>
</tr>
<tr>
<td>Pigm</td>
<td>Blast resistance</td>
<td>National Center for Bio-technology Information (NCBI) accession KU904633</td>
<td>Nucleotide-binding leucine-rich repeat receptors</td>
<td>Conferring durable resistance to the fungus Magnaporthe oryzae without yield penalty</td>
<td>16</td>
</tr>
<tr>
<td>Brown plant hopper resistance 6 (Bph6)</td>
<td>Brown plant hopper resistance</td>
<td>NCBI accession KX818197</td>
<td>Exocyst-localized protein</td>
<td>Increasing exocytosis, participating in cell wall maintenance and reinforcement, as well as activating cytokinin, salicylic acid, and jasmonic acid signaling pathway</td>
<td>40</td>
</tr>
<tr>
<td>Submergence 1A (Sub1A)</td>
<td>Submergence tolerance</td>
<td>NCBI accession DQ011598</td>
<td>Putative ethylene response factors</td>
<td>Crosstalk in multiple signaling pathways during submergence</td>
<td>148</td>
</tr>
<tr>
<td>SNORKEL1</td>
<td>Deep water adaptation</td>
<td>DNA Data Bank of Japan (DDBJ) accession AB510478</td>
<td>Ethylene response factors</td>
<td>Triggering internode elongation via gibberellin</td>
<td>44</td>
</tr>
<tr>
<td>SNORKEL2</td>
<td>Deep water adaptation</td>
<td>DDBJ accession AB510479</td>
<td>Ethylene response factors</td>
<td>Triggering internode elongation via gibberellin</td>
<td>44</td>
</tr>
</tbody>
</table>
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Quantitative trait locus</th>
<th>Trait</th>
<th>Gene ID</th>
<th>Encoding protein</th>
<th>Biological function</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermo-tolerance 1 (TTI)</strong></td>
<td>Thermotolerance</td>
<td>Os03g0387100</td>
<td>α2 subunit of the 26S proteasome</td>
<td>Modulation of protein homeostasis by the proteasome for thermotolerance</td>
<td>87</td>
</tr>
<tr>
<td><strong>Nitrate-transporter 1.1B (NRT1.1B)</strong></td>
<td>Nitrate uptake</td>
<td>Os10g05354200</td>
<td>Nitrate transporter</td>
<td>Enhanced nitrate uptake</td>
<td>46</td>
</tr>
<tr>
<td><strong>DEEPER ROOTING 1 (Dro1)</strong></td>
<td>Deeper rooting</td>
<td>Os09g0439800</td>
<td>Unknown membrane protein</td>
<td>Controlling root growth angle to improve drought avoidance</td>
<td>128, 129</td>
</tr>
<tr>
<td><strong>Grain chalkiness 5 (Chalk5)</strong></td>
<td>Grain quality</td>
<td>Os05g0156900</td>
<td>Vacuolar H⁺-translocating pyrophosphatase</td>
<td>Increasing the chalkiness by disturbing pH homeostasis</td>
<td>89</td>
</tr>
<tr>
<td><strong>Waxy</strong></td>
<td>Grain quality</td>
<td>Os06g01331000</td>
<td>Starch granule-bound starch synthase</td>
<td>Controlling the synthesis of amylose</td>
<td>141</td>
</tr>
</tbody>
</table>

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). PROSTATE GROWTH 7 (PROG7) is the main determinant in controlling the transition of plant architecture from prostrate growth to erect growth during African rice domestication. Comparing PROG7 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 PROG7 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 rationanly (91). **Thermo-tolerance 1 (TTI)** 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 weedingness 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 weedingness 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).

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, $G \times G$) and gene–environment interactions ($G \times E$). 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 $G \times G$ and $G \times E$ 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
advantageous effects of the heterotic loci (Figure 3), specifically in terms of yield heterosis, the second filial (F₂) populations from 17 representative hybrid rice crosses were also investigated (56). It is believed that F₂ populations generated from the elite hybrids (F₁) can provide useful information for heterosis analysis. Three genotypes (homozygous male parent, heterozygous, and homozygous female parent) will be present in the F₂ 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 F2 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 S5n, 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

Chen et al.
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 architecture1 (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 XDBB0400 to B.H.). We thank Zhoulin Gu and Qiang Zhao (National Center for Gene Research) for critical discussions and input.

LITERATURE CITED


6. Borel B. 2017. CRISPR, microbes and more are joining the war against crop killers. *Nature* 543:302–4


21. Characteristics of the 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.


51. Huang X, Han B. 2015. Rice domestication occurred through single origin and multiple introgressions. Nat. Plants 2:15207


127. The 3,000 Rice Genomes Proj. 2014. The 3,000 Rice Genomes Project. *Gigascience* 3:7


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


## Contents

From Bacteriophage to Plant Genetics
*Barbara Hohn* .............................................................. 1

Assembly of the Complexes of the Oxidative Phosphorylation System
in Land Plant Mitochondria
*Etienne H. Meyer, Elina Welchen, and Chris Carrie* ......................... 23

Chloroplast Lipids and Their Biosynthesis
*Georg Hözel and Peter Dörmann* ........................................ 51

Conditional Protein Function via N-Decrorn Pathway–Mediated
Proteostasis in Stress Physiology
*Nico Dismeyer* ..................................................................... 83

The Scope, Functions, and Dynamics of Posttranslational Protein
Modifications
*A. Harvey Millar, Joshua L. Heazlewood, Carmela Giglione, Michael J. Holdsworth,
Andreas Bachmair, and Waltraud X. Schulze* ................................. 119

Look Closely, the Beautiful May Be Small: Precursor-Derived
Peptides in Plants
*Vilde Olsson, Lisa Joos, Shanshoo Zbu, Kris Gevaert, Melinka A. Butenko,
and Ive De Smet* ............................................................... 153

Next-Gen Approaches to Flavor-Related Metabolism
*Guangtao Zhu, Junbo Gou, Harry Klee, and Sanwen Huang* .............. 187

Heterotrimeric G-Protein Signaling in Plants: Conserved and Novel
Mechanisms
*Sona Pandey* ..................................................................... 213

Division Plane Establishment and Cytokinesis
*Pantelis Livanos and Sabine Müller* ........................................... 239

Control of Meristem Size
*Munemori Kitagawa and David Jackson* ....................................... 269

The Dynamics of Cambial Stem Cell Activity
*Urs Fischer, Melis Kucukoglu, Ykä Helariutta, and Rishikesb P. Bhalerao* .. 293
Thermomorphogenesis
Jorge J. Casal and Sureshkumar Balasubramanian ............................................. 321

Leaf Senescence: Systems and Dynamics Aspects
Hye Ryun Woo, Hyo Jung Kim, Pyung Ok Lim, and Hong Gil Nam ......................... 347

Molecular Mechanisms of Plant Regeneration
Momoko Ikeuchi, David S. Fucero, Yuki Sakamoto, Akira Iwase,
Duncan Coleman, Bart Rymen, and Keiko Sugimoto ......................................... 377

Functional Status of Xylem Through Time
Craig R. Brodersen, Adam B. Roddy, Jay W. Wason, and Andrew J. McElrone ...... 407

Molecular Networks of Seed Size Control in Plants
Na Li, Ran Xu, and Yunbai Li .................................................................................. 435

Molecular and Environmental Regulation of Root Development
Hans Motte, Steffen Vanneste, and Tom Beeckman ............................................. 465

MicroRNAs and Their Regulatory Roles in Plant–Environment Interactions
Xianwei Song, Yan Li, Xiaofeng Cao, and Yijun Qi .............................................. 489

Molecular Interactions Between Plants and Insect Herbivores
Matthias Erb and Philippe Reymond ..................................................................... 527

A Molecular View of Plant Local Adaptation: Incorporating Stress-Response Networks
Acer VanWallendael, Ali Soltani, Nathan C. Emery, Murilo M. Peixoto,
Jason Olsen, and David B. Lowry ........................................................................ 559

Evolution of Glucosinolate Diversity via Whole-Genome Duplications, Gene Rearrangements, and Substrate Promiscuity
Brenden Barco and Nicole K. Clay ......................................................................... 585

Comparative and Functional Algal Genomics
Crysten E. Blaby-Haas and Sabeeha S. Merchant .................................................. 605

The Genomics of Oryza Species Provides Insights into Rice Domestication and Heterosis
Erwang Chen, Xuehui Huang, Zbixi Tian, Rod A. Wing, and Bin Han .................. 639

CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture
Kanling Chen, Yanpeng Wang, Rui Zhang, Huaweizhang, and Caixia Gao .......... 667

Risk Assessment and Regulation of Plants Modified by Modern Biotechniques: Current Status and Future Challenges
Joachim Schiemann, Antje Dietz-Pfeilstetter, Frank Hartung, Christian Kobl,
Jorg Romeis, and Thorben Sprink ...................................................................... 699
Crop Biodiversity: An Unfinished Magnum Opus of Nature
Matthew B. Hufford, Jorge C. Berny Mier y Teran, and Paul Gepts................. 727

Crop Improvement Through Temperature Resilience
Jingyu Zhang, Xin-Min Li, Hong-Xuan Lin, and Kang Chong ....................... 753

Water Use Efficiency as a Constraint and Target for Improving the
Resilience and Productivity of C3 and C4 Crops
Andrew D.B. Leakey, John N. Ferguson, Charles P. Pignon, Alex Wu,
Zhenong Jin, Graeme L. Hammer, and David B. Lobell ......................... 781

A Fruitful Journey: Pollen Tube Navigation from Germination to
Fertilization
Mark A. Johnson, Jeffrey F. Harper, and Ravishankar Palanivelu ............... 809

Errata

An online log of corrections to Annual Review of Plant Biology articles may be found at
http://www.annualreviews.org/errata/arplant