The rice genome revolution: from an ancient grain to Green Super Rice

Rod A. Wing[1,2], Michael D. Purugganan[3,4] and Qifa Zhang[5]**

Abstract | Rice is a staple crop for half the world's population, which is expected to grow by 3 billion over the next 30 years. It is also a key model for studying the genomics of agroecosystems. This dual role places rice at the centre of an enormous challenge facing agriculture: how to leverage genomics to produce enough food to feed an expanding global population. Scientists worldwide are investigating the genetic variation among domesticated rice species and their wild relatives with the aim of identifying loci that can be exploited to breed a new generation of sustainable crops known as Green Super Rice.

Lodging resistance
The ability of plants to withstand high-velocity winds, such as those from annual typhoons in the tropics. Typically, lodging resistance occurs by breeding for stiffer stalks, short stature or both.

Heterosis
Also known as hybrid vigour. A phenomenon whereby the hybrid produced by crossing two genetically distinct breeding lines (normally inbred) agronomically outperforms each of its parents (for example, in terms of higher yield and faster growth).

[1]*University of Arizona, Tucson, AZ, USA.*

[2]*International Rice Research Institute, Los Baños, Philippines.*

[3]*New York University, New York, NY, USA.*

[4]*New York University Abu Dhabi, Abu Dhabi, United Arab Emirates.*

[5]*Huazhong Agricultural University, Wuhan, China.*

**e-mail: rwing@email.arizona.edu; qifazh@mail.hzau.edu.cn*

https://doi.org/10.1038/s41576-018-0024-z

As one of the world's major staple crops, rice is an essential component of the diets and livelihoods of over 3.5 billion people. At its current rate of growth, the global population is set to reach ~10 billion people by 2050 (REF.[1]). Much of this increase will occur in poor, densely populated regions of the world that are already highly dependent on rice (such as Africa and southern Asia)[1], which underscores the crucial role rice will play in this looming humanitarian crisis.

Traditional crop improvement programmes rely on identifying and crossing plants with agronomically desirable phenotypes. These approaches have resulted in the adoption of semi-dwarf varieties for better lodging resistance and the exploitation of heterosis, which has seen rice yields increase substantially over the past half century. However, these rises in productivity have been costly in terms of resources and their adverse environmental effects in many rice-producing areas; such effects stem from the excessive use of pesticides, fertilizers and water, among other inputs. The concept of Green Super Rice (GSR) was first proposed 10 years ago[2] as a means to meet future demands while also reducing the costs and the ecological footprint associated with increased productivity. Key attributes of GSR varieties include reduced requirements for fertilizers, pesticides and water; increased yield; more palatable and nutritious grains; the ability to grow on marginal lands; and reduced greenhouse gas emissions (FIG. 1a).

To achieve this goal, it will be necessary to better understand, manage and utilize existing genetic variation present in domesticated rice and its wild relatives. In this regard, functional and comparative genomics studies will be key to identifying and understanding the genetic components responsible for agronomically beneficial traits (FIG. 1b). Rice is well positioned to lead the way in these new genomic breeding approaches for a number of reasons. At ~400 Mb, the diploid rice genome is the smallest among the domesticated cereals, making it particularly amenable to genomic studies. Furthermore, rice was the first crop plant to be completely sequenced[3], and the availability of a high-quality genome enables evolutionary, functional and population genomic studies. Additionally, the genus *Oryza* comprises wild and domesticated species with a long evolutionary trajectory that contains a virtually untapped reservoir of genes and traits that can be used for crop improvement (FIG. 2a,b). Indeed, rice has been cultivated in the Old World for thousands of years and in the New World for hundreds of years, during which time it has adapted to a range of geographical locations and a plethora of environmental conditions (FIG. 2c); the genes selected for during domestication and adaptation can potentially be systematically harnessed for crop improvement. Indeed, although difficult to cross, hybrids between all non-AA genome types have been successfully produced, resulting in the introgression of many valuable traits into cultivated rice[4].

Here, we review how comparative and functional genomic studies of domesticated and wild rice germplasm collections can be used to inform breeding programmes, with an emphasis on how they are contributing to the development of GSR varieties. We discuss insights gained from studying the domestication of Asian rice (*Oryza sativa*) and African rice (*Oryza glaberrima*), which occurred on two different continents ~6,000 years apart. We describe how entire germplasm collections are now being sequenced and phenotyped to identify genomic regions that are important for crop productivity and adaptation and how this process has been facilitated by advances in field phenotyping technologies and the availability of an increasing number of nearly gap-free reference genomes covering the breadth of cultivated and wild *Oryza* diversity. Finally, we address the need for a functional genomics and breeding

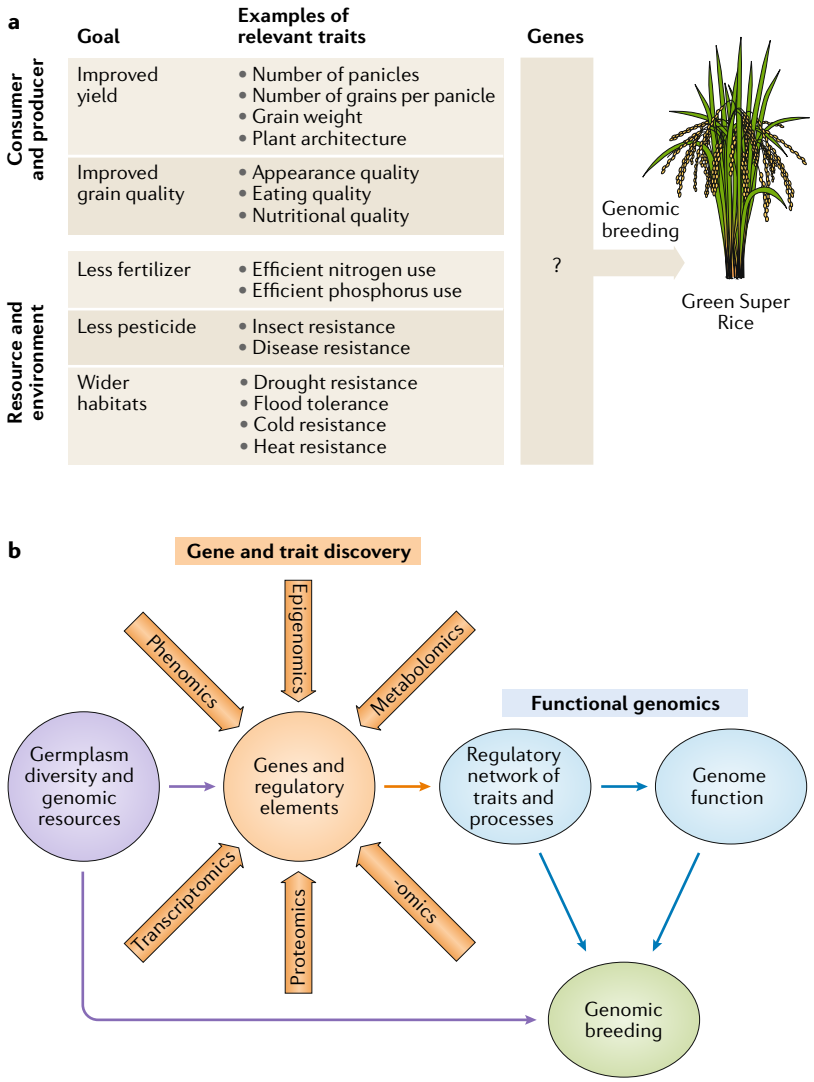


Fig. 1 | **Genomics-based strategies for developing Green Super Rice. a** | Understanding which genes regulate desirable 'green' traits can help generate improved crops, such as new varieties of Green Super Rice, through genomic breeding approaches. **b** | An outline of genomics research in the global rice community and how it relates to genomic breeding. The arrows indicate the workflow and information flow. The ultimate goal is to identify and understand the function of the full complement of genes in the rice genome. Progress made at any stage of the workflow can be applied to genomic breeding.

pipeline to accelerate the pace of crop improvement via genomic breeding technologies once candidate genes have been identified.

Genomics of domestication

During the Neolithic period, which began ~11,500 years ago, early farmers transformed wild plants into domesticated species with increased yield and yield stability[5]. For example, before the Green Revolution, yields of domesticated Asian rice (*O. sativa*) averaged 4.1 tonnes per hectare, which is 3.7 times more than the 1.12 tonnes per hectare average yield of its wild ancestor, *Oryza rufipogon*[5]. Furthermore, crop populations underwent adaptive diversification as they expanded beyond their original species ranges and encountered both new environments and new human cultural contexts. The study of these past events can identify beneficial agronomic traits that have been selected for over the course of crop domestication and diversification. Comparative, population and functional genomics approaches can be used to determine what genes are responsible for the selected traits, and this genetic information can potentially be used to shape future breeding efforts.

Introgression of genes during domestication and diversification of Asian rice. *O. sativa* displays genetic differentiation into subspecies or varietal groups, which have been validated by genomic studies[6]. These varietal groups include tropical japonica, temperate japonica and indica, aromatic and the lesser-known aus group[7] (FIG. 2b). However, it has been unclear whether these varietal groups arose from one or multiple de novo domestication events[7–10]. Recent population genetic modelling and phylogenomic analyses based on high-quality whole genome sequences from the different varietal groups and from wild *Oryza* species suggest a complex picture of rice evolution, in which extant rice populations originated from different ancestral populations of *O. rufipogon* and/or *Oryza nivara* that diverged ~300,000–400,000 years ago[10] (FIG. 3a). In this model, de novo domestication seems to have occurred only once — in japonica — and the subsequent introgression of japonica alleles into either wild rice or a proto-indica and/or proto-aus population gave rise to other Asian rice populations that today make up the different rice varietal groups[9,10].

The origin of indica (and possibly aus)[10] illustrates the role of introgression in moving selected domestication alleles among distinct populations and represents an early Neolithic example of rice improvement by hybridization between divergent populations. Analysis of the genomic regions introgressed from japonica to indica has shown that they contain a number of genes that are responsible for agronomically desirable traits, some of which are still used in breeding programmes today. These genes include the non-shattering allele of *SH4* (REF.[11]), which enhances the retention of mature seed on the rice panicle and which arrived in the South Asian subcontinent through the movement of early japonica rice, possibly through the ancient Silk Road into northwest India[12]; the colour gene *RC*, which leads to the culturally desirable white grain colour[13]; and the gene for erect growth *PROG1*, which prevents yield loss from lodging[14].

Independent domestication events in Asia and Africa selected for similar genes. Independent domestications within the genus *Oryza* provide a comparative context for the study of the evolution of key agronomic traits. While *O. sativa* evolved in Asia, a separate domestication occurred in West Africa that gave rise to *Oryza glaberrima* from its wild ancestor *Oryza barthii*[15] (FIG. 3b). Population genomic and phylogenetic analyses of whole-genome sequences suggest a primary inland domestication event for this species ~3,500 years ago, followed by diversification between inland and coastal populations, and between northern and southern groups[15,16].

Genomic breeding
Approaches that use the data, knowledge, resources, genes and technologies generated by genomic research to enhance breeding programmes.

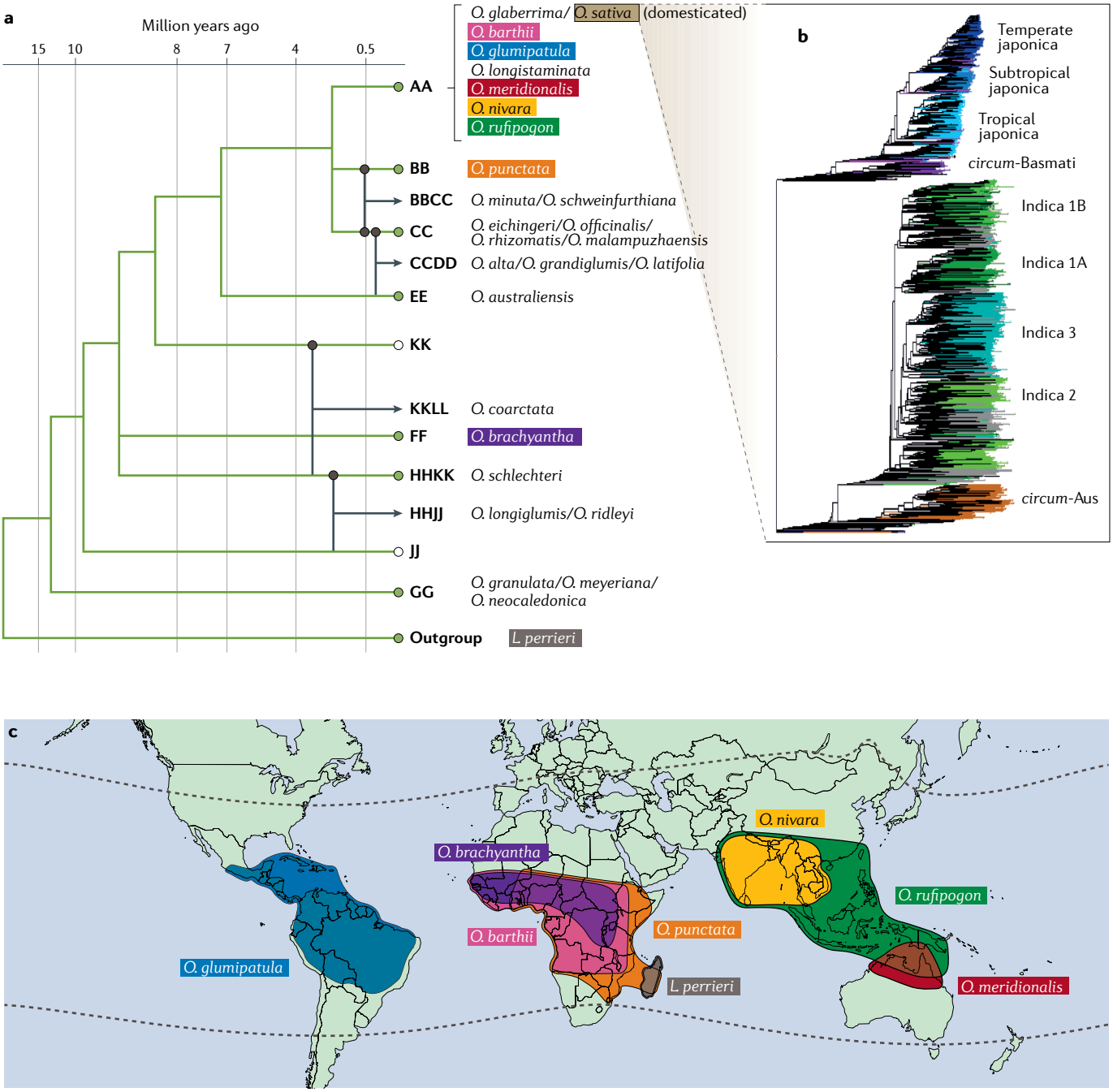


Fig. 2 | **Phylogeny and distribution of the *Oryza* genus. a** | Phylogenetic tree of the *Oryza* genus (modified from REFS[45,111]). The *Oryza* genus contains 27 species, two of which are domesticated. The genus harbours 11 extant genome types (6 diploid and 5 polyploid) that have a 3.6-fold variation in genome size (ranging from ~350 to 1283 Mb). Arrows indicate origin of the polyploids, closed circles represent maternal parents and open circles represent unidentified diploid genomes. **b** | Genomic analysis confirms that *Oryza sativa* can be subdivided into nine subpopulations[30]; shown is an analysis of the 3000 Rice Genomes Project (3K RGP) accession data set using the software program ADMIXTURE[112] **c** | World map showing that the genus *Oryza* is widely distributed. Indicated on the map are the growth limits of cultivated rice (dashed lines) and endemic locations of seven recently sequenced species, including the outgroup *Leersia perrieri*. Part **a** is adapted with permission from REF.[45], Elsevier and REF.[111], PNAS. Part **c** is adapted from REF.[42], Macmillan Publishers Limited, CC-BY-4.0.

Genome types
The *Oryza* genus is composed of ~27 extant species that harbour 11 distinct genome types (GTs), 6 of which are diploid ($n = 12$; GTs: AA, BB, CC, EE, FF and GG) and 5 of which are polyploid ($n = 24$; GTs: BBCC, CCDD, HHJJ, HHKK and KKLL). These GTs were defined based on cytogenetics (that is, chromosome number, size and shape), fluorescence in situ hybridization (FISH) and genetic hybridization.

Sequencing of the African rice genome has provided key insights into the genes underlying parallel evolution of domestication traits[15]. Comparative genomic analysis, for example, confirms that *HD1* (also known as *SE1*), a gene that controls flowering time, has been lost in African rice compared with its wild progenitor *O. barthii* and may explain the photoperiod insensitivity and synchronized flowering time of this crop species[15]. Scans for signatures of selection in the African rice genome implicate many loci that also seem to be

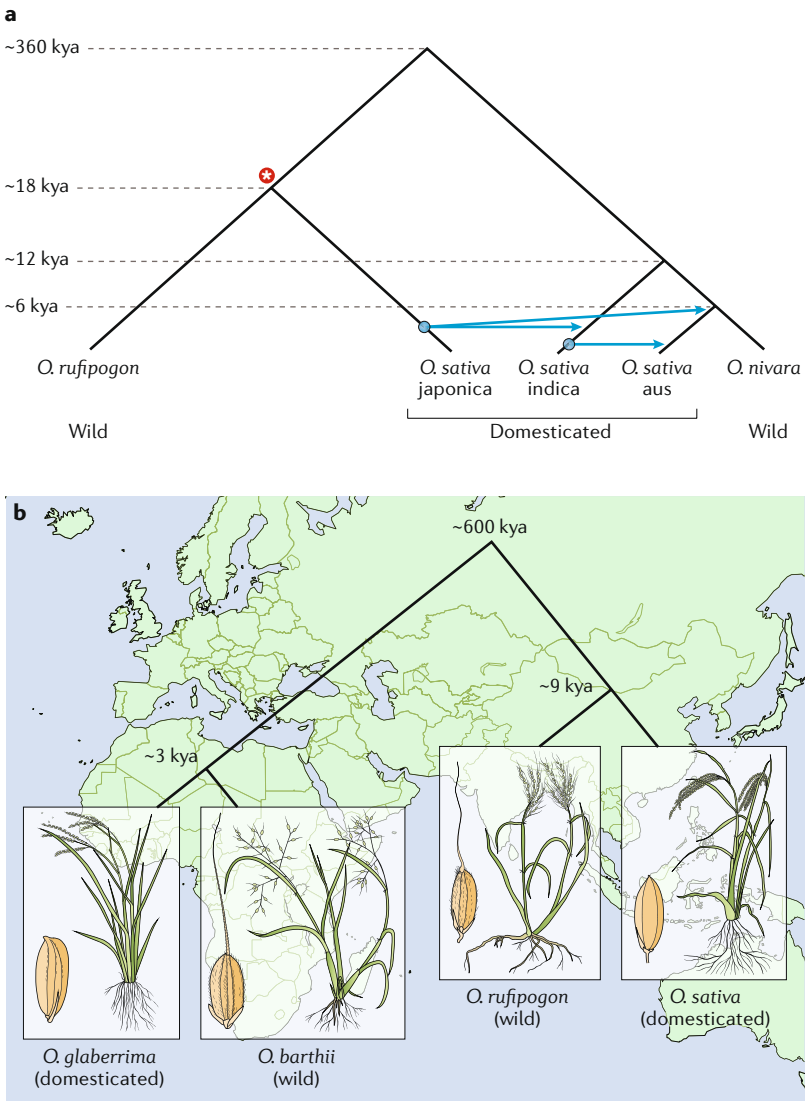


Fig. 3 | **Domestication of *Oryza sativa* and *Oryza glaberrima*. a** | Relationships between various *Oryza sativa* subspecies and their ancestors, *Oryza rufipogon* and *Oryza nivara*. The red asterisk indicates the time of divergence of japonica from extant *O. rufipogon* lineages (~18 kya). This early date may reflect divergence between current *O. rufipogon* populations and extinct wild rice populations that were the progenitors of *O. sativa* variegata japonica. Evidence suggests that hybridization resulted in introgression of genes from japonica to indica and aus (indicated by the blue arrows)[10]. **b** | Asian and African rice were domesticated independently ~6,000 years apart. kya, thousand years ago. Part **a** is adapted from REF.[10], Choi, J. Y. The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* 2017, **34**(4), 969–379, by permission of Oxford University Press. Part **b** is adapted from REF.[113], Macmillan Publishers Limited.

important for key agronomic traits in Asian rice, including the semi-dwarf gene *SD1*, the stress response gene *NAC6* and the shattering *SH4* locus[15]. Interestingly, at least two genes involved in the non-shattering trait in Asian rice (*SH1* and *SH4*) are associated with mutations that are potentially responsible for the non-shattering phenotype in *O. glaberrima*[15]. The *SH1* gene is deleted in African rice, while the *O. glaberrima SH4* orthologue shows reduced expression and is found in proximity to a selective sweep, which is consistent with selection for non-shattering in this domesticated crop species. These studies suggest that, for certain key traits, only a limited number of genes are selected for during and after domestication. Thus, genes identified as regulating important agronomic traits in one species (here, African rice) could be used to develop those traits in GSR.

Adaptive evolution: genes involved in the spread of Asian and African rice. After the onset of domestication, a crop species begins to spread from its centre of origin and move into regions with distinct environmental features; therefore, it needs to adaptively evolve to these new environments. Moreover, some loci in a given ancestral population might be selected for in new environments because they generate key traits that remain desirable in distinct cultures and locations. Some of these phenotypes (such as abiotic stress response and grain quality) could therefore be extremely important for breeding GSR.

For example, studies of the evolutionary history of rice reveal the action of postive selection for the WAXY gene[17,18]. Alleles of this gene have major effects on the main determinants of cooking and eating quality of rice: amylose content, gelling temperature and gel length of the rice grain[19]. Variation in these traits has contributed greatly to the culinary diversity of rice and how it is consumed. Molecular evolutionary studies reveal that glutinous (sticky) rice carries an allele of *WAXY* with a mutant splice site, which likely originated in mainland Southeast Asia and spread across East Asia and was eventually incorporated into temperate japonica; the geographical spread of the allele is associated with cultural desirability for sticky rice[17,18].

Another clear association of an important agronomic trait and geographic adaptation is salinity tolerance in *O. glaberrima*[16]. Cultivars from across the geographic range of African rice possess some level of salt tolerance, with the exception of southerly coastal populations[16]. Genome-wide association studies (GWAS) have identified 11 loci associated with salinity tolerance[16], one of which overlaps *HAK5*, a high-affinity potassium transporter gene associated with rice salt tolerance[20] that is induced in *O. glaberrima* upon salt stress[16]. Moreover, two of the GWAS loci span genomic regions that have undergone selection associated with geographic adaptation in African rice; one of these regions contains 41 genes[16], including the *PPL* locus, which is a member of a gene family associated with rice salt tolerance[21].

GWAS have also been performed on the two major indica rice groups that have been generated by independent breeding activities in China (indica I) and Southeast Asia (indica II, which has largely been generated by the International Rice Research Institute, IRRI)[22]. This study identified ~200 regions that contained signatures of domestication or artificial selection and which spanned 7.8% of the rice genome. These regions harbour around 4,000 non-transposable element genes, including many with functions that are associated with important agronomic traits. Examples include *GN1A*[23], which affects the number of grains per panicle; *SD1* (REF.[24]), which affects plant height; *AMT1;1* (REF.[25]), which regulates nitrogen uptake; *XA4* (REF.[26]) and *XA26* (also known as *XA3*) (REF.[27]), which are involved in disease resistance; and *RF1* (REF.[28]), which restores fertility to hybrid rice.

Introgression
The transfer of genes and genomic segments from one species or population to another through hybridization.

Field phenotyping
The use of state-of-the-art sensor and camera systems, mounted on tractors, gantries and drones, to measure plant phenotypic traits (such as height, leaf angle, 1,000-grain weight, disease pressure and canopy temperature, among others) over the course of a growing season.

Reference genomes
Also referred to as a reference sequence (RefSeq). A genome assembly that is used to represent the full genome sequence of a given organism. Ideally, a RefSeq will be gap-free and have zero sequence errors. However, genome assemblies can potentially be missing up to 50% of the full genome sequence, primarily owing to the sequencing technology used (for example, short read sequencing) and the assembly tools available.

Green Revolution
The substantial increase in grain production that began in the late 1960s and early 1970s. It was a result of widespread adoption of high-yielding wheat and rice varieties bred to incorporate semi-dwarf genes and a more systematic use of nitrogen fertilizers and pesticides.

Hybridization
The process or outcome of performing genetic crosses between individuals from distinct species or highly divergent populations.

Selective sweep
A genomic region that appears to be under natural or artificial selection. In the context of this Review, we consider it to be a region of the genome including and surrounding a domestication trait (for example, yield or grain shattering).

Abiotic stress
A stress considered to be of non-biological origin, such as heat, salt, drought, nutrient, light and dark, among others.

Genome-wide association studies
(GWAS). A mapping approach that relies on an observed statistical correlation between individual genomic variants (such as single nucleotide polymorphisms (SNPs)) and specific phenotypes in a natural population.

Furthermore, grain yield was positively correlated with the number of selection signatures, which indicates that the signatures could be useful for predicting agronomic potential and implicates the selected loci as potential targets for crop improvement programmes.

Maximizing the utility of genebanks

A logical extension to the evolutionary studies discussed above is to perform systematic comparative genomic analyses of Asian and African rice populations to identify genes associated with local adaptation to abiotic stresses and other environmental and cultural pressures, thereby providing new targets for breeding efforts. Such analyses require both extensive genotyping (that is, resequencing) combined with high-throughput field and laboratory phenotyping under multiple environmental conditions with the goal of translating genotype to phenotype in a breeder friendly format.

Generating digital genebanks. Virtually all crops have germplasm collections, called genebanks, that are used to store, preserve and maintain germplasm as seed and/or live plants for breeding and conservation purposes. Such banks are essential to ensure that crop diversity is preserved in the long term and can be easily accessed, typically through a universal material transfer agreement. The majority of crop genebanks are national, and their seed is not always easily accessed by the international community. However, the Consultative Group and International Agricultural Research (CGIAR) partnership maintains a number of international genebanks where researchers from around the world can access and deposit valuable germplasm for research purposes. There are three international rice genebanks: the International Rice Genebank (IRRI, Philippines), which holds ~128,000 rice accessions and 4,464 wild relatives; the Africa Rice Genebank (Africa Rice Center (ARC), Ivory Coast), which holds ~20,000 accessions; and Oryzabase (Japan) which holds ~22,000 accessions, including cultivated, wild, mutant and genetic stocks and populations.

Resequencing large populations is an important tool used to unravel population structure, detect signatures of selection and map quantitative trait loci (QTL). As sequencing costs plummet and the throughput of technology platforms continues to increase, crop communities are now contemplating resequencing entire germplasm collections to create digital genebanks. In 2014, the Chinese Academy of Sciences (CAS), the Beijing Genomics Institute (BGI) and the IRRI performed a pilot digital genebank experiment (termed the 3000 Rice Genomes Project (3K RGP)) in which >3,000 diverse rice accessions from 89 countries were resequenced, with an average coverage of 14× (REF.[29]). The resulting >17 Tb of raw data were filtered down to 18.9 million single nucleotide polymorphisms (SNPs) when mapped to the Nipponbare reference sequence (RefSeq)[3] alone. In their paper, they proposed the use of 'xian' and 'geng' as the varietal group names, in place of indica and japonica respectively, in recognition of the fact that these names have been used in China for several thousand years, which we corroborate. Initial phylogenetic analysis of 200,000 random sets of SNP variants enabled the 3,000 accessions to be subdivided into 5 distinct varietal groups: 2 major groups (geng (or japonica) and xian (or indica)); 2 small groups (aus/boro and basmati/sadri); and a small intermediate (admixed) group. These five groups have now been further resolved into nine groups using population structure analysis[30] (FIG. 2b). In total, we estimate that more than 10,000 domesticated and wild rice accessions have been resequenced to date, and consortia are nucleating to secure funding to make digital genebanks a reality for most major crops over the next 5 years.

Towards platinum standard reference genomes. Efforts to generate digital genebanks will be most powerful in combination with high-quality, near gap-free reference genomes (known as Platinum Standard Reference Sequences, PS-RefSeqs) that span the spectrum of cultivated and wild rice. Mapping resequencing data to these reference genomes will enable the vast majority of allelic and haplotype diversity across the *Oryza* pan-genus genome to be detected. This approach would also help to identify redundancy, enabling the most diverse accessions to be targeted for future use. PS-RefSeqs are available for the African rice *O. glaberrima*[15] and for the two main varietal groups of cultivated Asian rice: *O. sativa* japonica[3] and *O. sativa* indica[31,32]. Draft genomes were first published for japonica[33] and indica (known as 93–11)[34] in 2002, but the first true reference genome for rice, and in fact of any crop species, was that of the japonica cultivar Nipponbare, published in 2005 (REF.[3]). This Nipponbare RefSeq[3] remained the highest quality RefSeq of any crop genome until 2016, when two reference genomes of similar quality were released for indica: Minghui 63 and Zhenshan 97 (REF.[31]). Comparative analyses of these two indica genomes uncovered extensive structural differences, especially with respect to inversions, translocations, insertions and deletions, and segmental duplications. More recently, a de novo assembly of the indica cultivar Shuhui498 has been reported with only five gaps and seven complete end-to-end chromosomes[32]. These three indica reference sequences have essentially replaced the original 93–11 indica sequence released in 2002.

Efforts are now underway to generate high-quality genome assemblies for representatives of all six wild AA genome *Oryza* species, as well as representative accessions from species that harbour ten other genome types. Indeed, over the next 2 years, the *Oryza* community can expect to have access to PS-RefSeqs for 2–4 representative accessions of each of the 9 subpopulations of cultivated rice, as defined by the 3K RGP[30], and a complete set of PS-RefSeqs for all 25 defined wild *Oryza* species, including 10 polyploid species (FIG. 2a,b). To date, assemblies have been published for 35 cultivated and wild species[15,31,32,35–44] (FIG. 2c; see Supplementary Table 1), and de novo sequencing and upgrades are ongoing for at least 30 genomes. These projects include de novo sequencing of three CC genome species (N. Kurata and M. Shenton, personal communication, and R.A.W., unpublished data) and the first *Oryza* polyploid species, *Oryza coarctata* (KKLL genome type), a

Artificial selection
Selection for desirable traits that is consciously and deliberately carried out by humans.

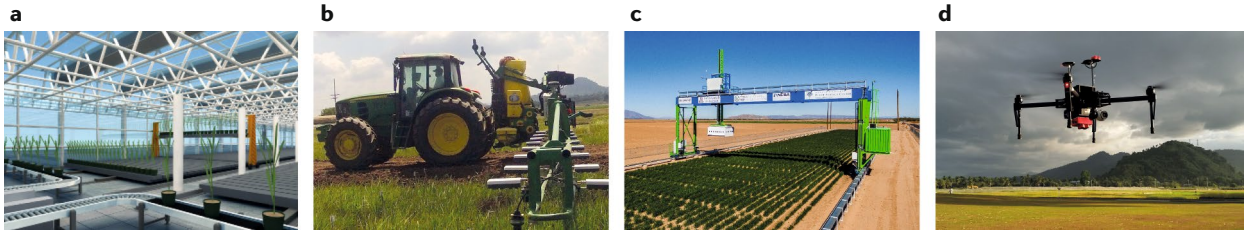
Resequencing
A technique used to sample an individual genome without the need to generate a full genome sequence. Resequencing data typically consists of short (250 bp) sequence reads at low (0.1–10-fold) genome coverage that are mapped by sequence complementarity to a reference sequence (RefSeq) to detect genetic variation (such as single nucleotide polymorphisms (SNPs) and indels) between the resequenced individual and the RefSeq.

Quantitative trait loci
(QTL). Genetic loci that contribute (positively or negatively) to non-discrete traits, such as yield, grain quality, water and heat stress.

highly salt-tolerant species (R.A.W., unpublished data). The majority of the wild *Oryza* genome sequencing efforts have been spearheaded by the International Oryza Map Alignment Project (I-OMAP), and a permanent, publicly available collection of living voucher specimens is maintained at the IRRI[42,45].

Generating phenomic data. While digital genebanks are valuable in their own right, the ability to link genotypic information to phenotypes in a high-throughput format is of utmost importance for the rapid identification of new allele combinations suitable for the development of GSR cultivars. High-throughput greenhouse and field phenotyping platforms, in both academic and industrial settings, have advanced rapidly over the past 10 years and are continuously improving. For instance, a greenhouse-based high-throughput rice phenotyping facility has been developed in China, which includes a novel yield traits scorer that can automatically measure total spikelet number, filled grain number, spikelet fertility, yield per plant, 1,000-grain weight, and grain shape and size at a rate of one plant per minute[46] (FIG. 4a). Recently, tractor-mounted multispectral ultrasonic and reflectance sensors were used to generate phenotype data for a high-throughput field experiment on a rice population comprising 1,516 recombinant inbred lines[47] (FIG. 4b). Compared with taking measurements manually, the high-throughput tractor system was just as accurate for traits such as plant height, flowering time, grain yield and harvest index but took a fraction of the time to gather data. Other field-based high-throughput systems include gantry platforms, such as the TERRA-REF Field Scanalyzer in Arizona (FIG. 4c) and the LeasyScan Phenotyping Platform at the International Crops Research Institute for the Semi-Arid Tropics (ICRASAT)[48] in India, and unmanned aerial vehicles[49] (FIG. 4d).

Regardless of the high-throughput platform, there is a pressing need to coordinate intensive phenotyping activities on a common set of accessions under multiple biotic stress and abiotic stress conditions. In 2011, 15 partner institutions established the Global Rice Phenotyping Network to phenotype well-characterized sub-panels (such as the ORYTAGE and Phenomics of Rice Adaptation and Yield Potential (PRAY) panels) that represent indica, tropical japonica and aus subpopulations and that can be genotyped using a variety of methods[50–54], including a common 700,000 SNP high-density array[55]. Recently, interrogation of two indica and aus PRAY panels for loci associated with early responses to salinity tolerance identified a previously undetected locus on chromosome 11 that



	Greenhouse	Tractor	Gantry	Drone
Traits	Canopy height and temperature, flowering time, 3D structure, nitrogen and chlorophyll content, yield components and biomass	Days to flowering, canopy height, percent cover and nitrogen level	Canopy height and temperature, flowering time, 3D structure, nitrogen and chlorophyll content, yield components and biomass	Days to flowering, canopy height, percent cover and nitrogen level 1–2 cm pixel per spatial resolution
Capacity (Ha/hr)*	0.003	1.0	0.1	1.0 (10–20 Ha/hr sample rate)
Pros	Fully automated, high precision, unlimited sensor options and controlled environment	Mobile multipurpose platform, ease of use, efficient data handling and field environment	Fully automated, high precision, many sensor options and field environment	Rapid sampling, low cost, flexible layouts, deployable anywhere and field environment
Cons	Expensive, low throughput and non-field environment	Semi-automated, unstable platform, fixed layouts and weather-dependent	Expensive, slow, fixed location and weather-dependent	Semi-automated, sensor limitations, regulatory concerns and weather-dependent

*Capacity includes time for basic data processing

Fig. 4 | **High-throughput phenotyping platforms. a** | High-throughput rice phenotyping greenhouse facility (HRPF) in Wuhan, China. The HRPF can phenotype ~4,000 plants per day. **b** | High-throughput field scanner (HTFS) phenotyping a paddy rice field at the International Rice Research Institute (IRRI), Los Baños, Philippines. The HTFS can phenotype 3,000 25-plant plots (of identical genotype) per hour. **c** | TERRA-REF field scanner phenotyping wheat. The Lemnatec Field Scanalyzer at the University of Arizona Maricopa Agricultural Center and US Department of Agriculture Arid Land Research Station in Maricopa, Arizona, is the largest field crop analytics robot in the world. This high-throughput phenotyping field-scanning robot has a 30-ton steel gantry that autonomously moves along two 200-metre steel rails while continuously imaging the crops growing below it with a diverse array of cameras and sensors. **d** | Autonomous flight drone used for phenotyping at the IRRI. Automated image processing produces 3D model and high resolution geo-referenced orthomosaics used for plot-level data analysis. Ha/hr, hectares per hour. Part **a** is courtesy of Wanneng Yang, Huazhong Agricultural University, China. Parts **b** and **d** are courtesy of Stephen Klassen, IRRI, Philippines. Part **c** is courtesy of Mark Yori, Phoenix Drone Services, USA

promotes efficient transpiration[55]. The IRRI has recently announced another much larger effort called the Global Rice Array (GRA), which proposes to phenotype the 3K RGP accession data set at multiple locations around the world and under numerous field, greenhouse and laboratory conditions. For example, efforts are underway to import the 3K RGP accession data set into the USA for phenotyping not only in rice-growing regions of the USA but also in other regions such as the TERRA-REF Field Scanalyzer in Maricopa, Arizona (FIG. 4c), to detect and map phenotypes under semi-arid environmental conditions.

Accessing and integrating genomic and phenomic data. Easy access to genomic and phenotypic data is of critical importance for both applied and basic research in rice. A number of resources are available for accessing and exploring single rice reference genomes. For example, the Rice Annotation Project (RAP)[56] and Michigan State University DB (MSU-DB) data portals are considered the primary go-to sites for access to the Nipponbare RefSeq and all associated annotations. These sites were created independently at the onset of the original International Rice Genome Sequencing Project (IRGSP) and culminated in an assembly unification in 2013 (REF.[35]), which has since been used for all subsequent Nipponbare RefSeq annotation updates. A database for the PS-RefSeqs for indica rice cultivars Minghui 63 and Zhenshan 97 can be found at Rice Information GateWay (RIGW)[31,57,58] and R498 at MBKBase[32]. Alternatively, these and many other rice genome assemblies can be accessed through Genbank.

Although these databases and assemblies are extremely useful, they tend to be focused on single genomes and are not set up to integrate or interrogate other cultivated rice and wild rice genomes as they come on line. To meet this need, Gramene and Ensemble Plants contain genome information for both cultivated and wild *Oryza* accessions as well as multiple analytical tools for data analysis. For example, these sites allow researchers to easily compare and download large genomic regions of interest across multiple genome assemblies. Gramene also hosts the Rice Diversity database, which contains genotype and phenotype data for grain length and panicle architecture from 1,568 diverse rice varieties[59].

As digital genebanks continue to grow, databases that can store, map and deliver SNP, structural variation and phenotypic data on demand are needed. One such database for rice is called SNP-Seek[60] (FIG. 5). SNP-Seek can display millions of SNPs across hundreds to thousands of accessions in real time and contains phenotype scores for 72 traits. In a 2016 update, SNP-Seek mapped the 3K RGP accession data set to four additional draft genome assemblies (IR 64 (indica), 93–11 (indica), DJ 123 (aus) and Kasalath (aus)), resulting in the discovery of ~11 million new SNPs and ~0.5 million new insertions and deletions[30]. These results argue strongly that continued resequencing and mapping to multiple RefSeqs are needed if we are to gain a complete understanding of the genetic variation that exists in cultivated and wild rice.

From candidate locus to gene

So far, we have reviewed the use of evolutionary and association studies to identify regions of the genome that are under selection. However, many agriculturally important genes relevant for breeding GSR have already been isolated and functionally characterized using more traditional approaches, such as map-based cloning of naturally occurring or induced mutations or transposon tagging. By the end of 2017, a total of 2,996 genes had been analysed for biological functions using various approaches (data from funRiceGenes[61]) and can be classified into functional categories pertaining to the goals of GSR[2] (FIG. 6).

Resistance to diseases or pests. Utilization of disease-resistance and insect-resistance genes provides the most economical and effective approach for reducing pesticides and is therefore one of the primary targets in GSR breeding. Approximately 8% of rice genes that have been functionally characterized to date have roles in disease resistance or pest resistance (FIG. 6b). The two major diseases that frequently lead to heavy yield losses in most rice-growing areas are bacterial blight and fungal blast, which are caused by *Xanthomonas oryzae* pv. *oryzae* and *Magnaporthe oryzae*, respectively. In the past decade, 11 genes for bacterial blight resistance and 27 genes for fungal blast resistance have been characterized and are available for breeding purposes[62–65]. Comparison of the predicted gene products revealed very different features of plant–pathogen interactions between these two diseases. All but three of the fungal blast resistance loci encode nucleotide binding site and leucine-rich repeat (NBS-LRR) proteins, which are assumed to provide effector-triggered immunity and function singly or in pairs to confer race-specific resistance. By contrast, only one of the bacterial blight resistance genes encodes an NBS-LRR protein. Three of the remaining ten genes activate expression of other rice resistance genes rather than directly conferring resistance; four are recessive resistance genes that prevent invasion of the pathogen via various processes; two are assumed to confer pattern-triggered immunity; and the final gene encodes a wall-associated kinase that improves lodging resistance in addition to providing resistance to bacterial blight[65]. Clearly, the molecular mechanisms underlying bacterial blight resistance are much more diverse than for fungal blast resistance, suggesting that different strategies should be adopted to achieve durable resistance to these two diseases. Pyramiding of genes belonging to different functional categories may provide a viable solution to breeding resistance to bacterial blight, while developing a multiline series of near isogenic lines, each carrying a different resistance gene, may be the solution for durable resistance to rice blast.

Thirty genes that provide resistance to the brown planthopper, one of the most damaging rice pests, have been identified, and 12 of them have been cloned[66,67]. The molecular features of these genes are surprisingly like those that confer fungal blast resistance: 8 of the 12 genes encode NBS-LRR proteins, which presumably provide effector-triggered resistance as is observed for plant–pathogen interaction. Thus, adopting a similar multiline approach might also be an effective strategy for achieving durable resistance to this insect. Several of

Xian
Also known as indica, a major group of Asian cultivated rice that is widely grown in tropical and subtropical regions of Asia and is partially reproductively isolated from the geng rice.

Geng
Also known as japonica, a major group of Asian cultivated rice that is widely grown in temperate regions of Asia and other areas and is partially reproductively isolated from xian rice.

Living voucher specimens
Single-plant accessions selected to be representative of a particular species. In the case of most wild *Oryza* species, vouchers can be clonally propagated indefinitely.

Biotic stress
A stress considered to be of biological origin, such as plant pathogens (bacteria, fungi and viruses) and animal pests (insects and nematodes), among others.

Nucleotide binding site and leucine-rich repeat (NBS-LRR) proteins
Members of a large class of proteins encoded by many disease-resistance or insect-resistance genes (*R* genes) of plants. An NBS-LRR protein contains a nucleotide binding site (NBS) domain and a leucine-rich repeat (LRR) domain, which are believed to confer the specificity of resistance.

Effector-triggered immunity
A defence response that is initiated when a pathogen effector molecule is recognized by a cytoplasmically localized host nucleotide binding site and leucine-rich repeat (NBS-LRR) protein.

Pattern-triggered immunity
A defence response that is initiated when a pathogen-associated molecular pattern is recognized by the corresponding host pattern recognition receptor.

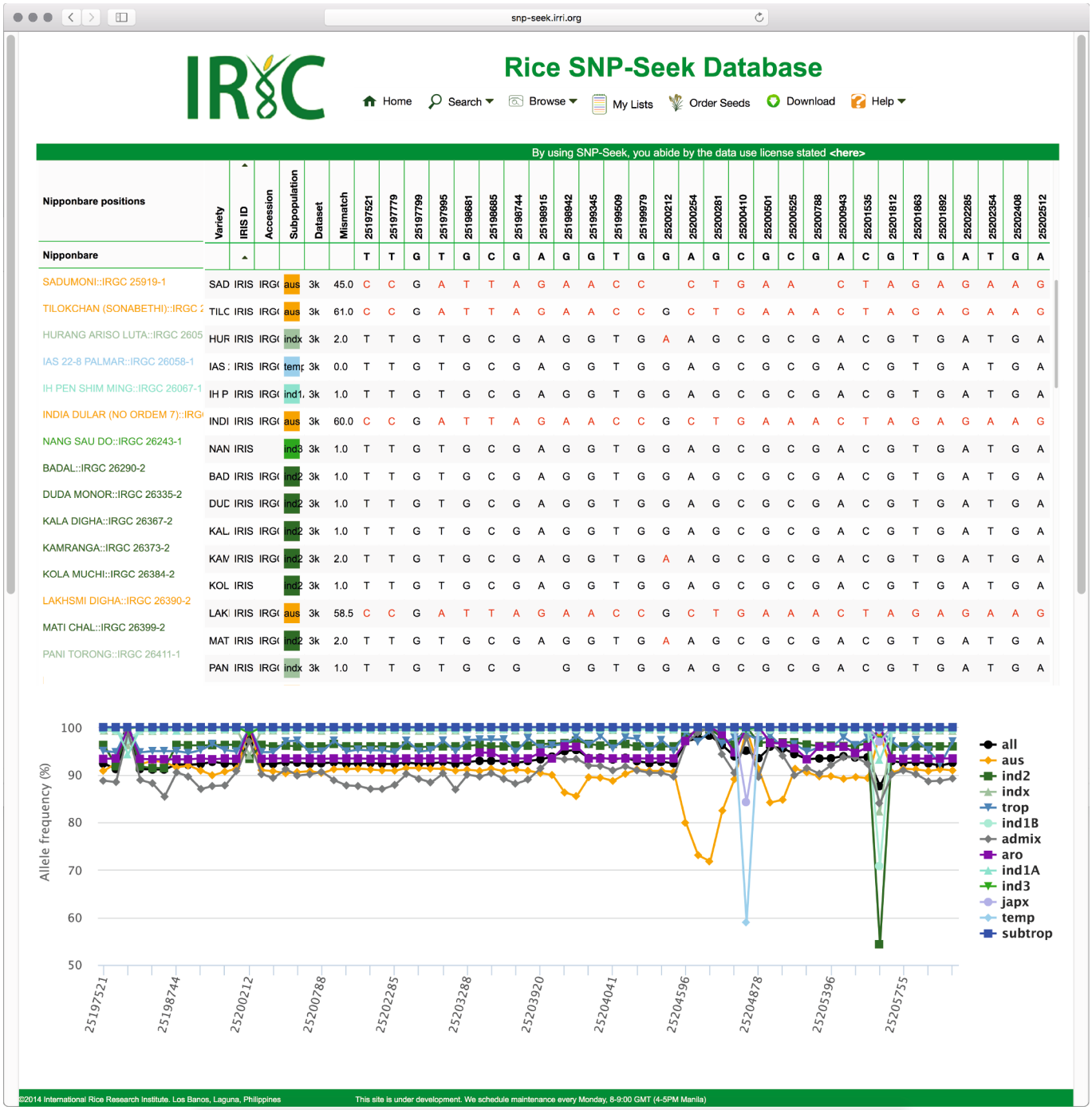


Fig. 5 | **Characterization of the 3K rice genomes through the SNP-Seek database.** Rice SNP-Seek Database output profiling single nucleotide polymorphism (SNP) calls and allele frequencies across a region of the *SH1* shattering gene. Top panel shows specific SNPs (in red) relative to the Nipponbare reference sequence (RefSeq) at 28 nucleotide positions across 15 accessions in the *SH1* gene. The lower panel shows the allele frequencies across the *SH1* gene (~8.6 kb region) on chromosome 3 for the entire 3000 Rice Genome Project (3K RGP) accession data set. Note the low allelic diversity from SNP 25197521 to 25204041 followed by a large increase in diversity, especially in temperate japonica at SNP 25204788 and indica-2, indica-1B and indica-x at SNP 25205648. These data indicate that these two SNPs are under strong artificial selection in those subpopulations.

these genes have been widely used in rice breeding programmes. For example, *BPH1*, *BPH2*, *BPH3* and *BPH4* have been incorporated into a number of rice varieties released by the IRRI, and *BPH14* and *BPH15* have been incorporated into several superior cultivars in China[67].

It should also be mentioned that numerous genes have been identified that provide quantitative, durable and race non-specific or pathogen non-specific resistances, and they might also be of utility for improving resistance to pests or pathogens[68].

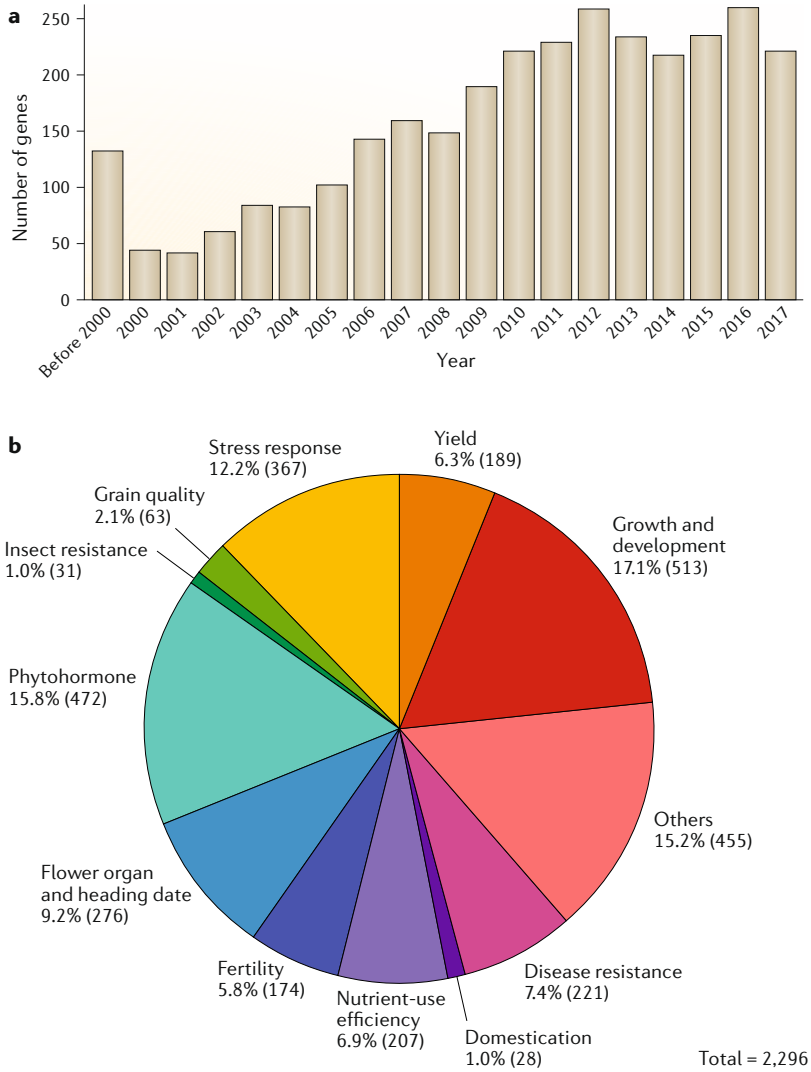


Fig. 6 | **Many rice genes have been characterized and classified by function. a** | Graph showing the number of rice genes functionally analysed in the published literature by year based on PubMed data accessed and analysed using funRiceGenes[61]. **b** | Functional classification of those genes according to traits that are desirable for developing Green Super Rice.

Resistance to abiotic stress. Providing resistance to abiotic stresses, such as drought, flood, temperature and salinity, is an intrinsic goal of GSR, and more than 360 genes have been identified that are involved in stress responses (FIG. 6b). Drought in particular has been identified as a major constraint for rice production[69], but despite tremendous efforts over the past 2 decades, very few genes have been identified that might be of practical use for breeding drought-resistant rice[70]. However, important progress has been made in identifying genes conferring resistance to a variety of abiotic stresses, including drought avoidance (*DRO1*)[71], submergence or flood tolerance (*SUB1* (REF.[72]), *SNORKEL1* (REF.[73]) and *SNORKEL2* (REF.[73])), low temperature tolerance (*qLTG3-1* (REF.[74]), *LTG1* (REF.[75]) (also known as *HBD2*) and *COLD1* (REF.[76])), heat tolerance (*TT1* (REF.[77]) and *HTAS*[78]) and salt tolerance (*SKC1*, also known as *HKT8*)[79]. Several of these genes have been incorporated into breeding programmes[70], and the best-known example is the introgression of *SUB1* from the landrace variety FR13A into a number of cultivars to greatly improve submergence tolerance[72].

Efficient nutrient use. More than 200 genes have been identified that improve the efficiency of nutrient use in crop plants (FIG. 6b), and developing plants containing such genes may provide the most effective approach to reducing fertilizer application. The molecular mechanisms underlying nitrogen and phosphorus use are better understood than for other nutrients and are therefore the main focus of GSR breeding efforts at present. A number of potentially useful genes involved in nitrogen use have been identified. For example, a naturally occurring allele of the nitrate transporter gene *NRT1.1B* from indica rice has been shown to improve nitrate uptake in japonica rice under field conditions[80]. A naturally occurring mutant of *DEP1-1*, which encodes a Gγ protein, exhibits nitrogen-insensitive vegetative growth with increased nitrogen uptake and assimilation[81]. Overexpression of *NRT2.3b* (also known as *NRT2.3*), a pH-sensitive nitrate transporter gene, enhances uptake not only of nitrogen but also of iron and phosphorus[82]. Other genes involved in efficient use of phosphorus include the protein kinase-encoding gene *PSTOL1*, which increases yields in phosphorus-deficient soil by conferring tolerance to phosphorus deficiency[83]; and the gene encoding the SULTR-like phosphorus distribution transporter, which when impaired in the nodes of rice plants reduces the allocation of phosphorous to the grain and lowers its phosphorus content without incurring a yield penalty[84].

Improved yield. The yield of an individual rice plant is determined by three component traits: the number of panicles, the number of grains per panicle, and the grain weight[85]. At the population level, the architecture of the plant is also an important component trait of yield as it determines the number of panicles per unit area[86]. Genes that regulate yield tend to be highly pleiotropic[85], that is, they affect many different traits. Thus, obtaining optimal yield often involves balancing the different phenotypic effects. For example, the gene *GHD7* regulates the number of grains per panicle but also has a large effect on plant size and flowering time[87]. Similar effects have also been found for *DTH8* (REFS [88,89]) (also known as *GHD8* and *HD5*) *GHD7.1* (REF.[90]) (also known as *HD2*) and *HD1* (REF.[91]), and manipulating the flowering time using different combinations of these four genes has been shown to be predictive of the performance of the plants[92]. It has also been shown that downregulation of the gene *SGDP7* (also known as *FZP*) by natural variation decreases the grain size but increases the number of grains per panicle[93]. Moreover, optimal expression levels of the *IPA1* gene (also known as *WFP*) are needed to achieve an ideal plant architecture that has fewer but more productive tillers resulting in higher yield[94]. The pleiotropy associated with genes affecting yield means that they can seem to be underrepresented in functional classifications (FIG. 6b); however, genes categorized as affecting growth and development[95] and flower organ and heading date also affect yield directly or indirectly. It is important, therefore, to characterize entire regulatory networks, rather than component traits or individual genes, when considering yield. Hybrid rice

has contributed greatly to the increased grain production globally, and exploiting heterosis between indica and japonica subspecies and/or varietal groups has long been considered to be the next step towards further boosting yield levels. Unfortunately, advances have been limited by inter-subspecific hybrid sterility. However, several loci for hybrid sterility have now been identified and characterized, which promises to stimulate progress in this area[96–99].

Enhanced grain quality. Grain quality consists of a group of traits that are important to consumers and, to some extent, rice producers (that is, farmers and millers). Although preferences vary in different parts of the world, the quality of the rice grain is generally characterized by the following features: appearance (shape and translucence); milling quality (chalkiness and intactness of milled rice); cooking and eating quality (the combination of amylose content, gel consistency and gelling temperature); and nutritional quality (macronutrient and micronutrient content). Extensive genetic studies have identified many loci that regulate grain quality traits[100] (FIG. 6b), including *WAXY*, which has major effects on culinary quality[19]; *ALK*, which regulates gelling temperature[101]; and *CHALK5*, which affects translucence of the grains[102]. In addition, genes identified for their effects on grain size (length and width), such as *GS3* (REF.[103]) (also known as *LK3*) and *GW5* (REFS [104,105]), also affect many aspects of grain quality because the appearance and milling quality of grains are closely related to grain size and shape. With the increased awareness of health-promoting effects of plant products and demands for more nutritious rice, functional genomic research should also address the need for rice grains enriched in macronutrients and micronutrients, such as iron and zinc, as well as special classes of metabolites such as β-carotenoid and anthocyanin. Major progress has been made in increasing the levels of some of the nutrients using transgenic approaches[106,107]; however, public acceptance of genetically modified foods remains low. Thus, the availability of nutrient-enhanced rice varieties may depend on the identification and incorporation of naturally occurring alleles using genomic breeding approaches.

Genomic breeding of Green Super Rice

Each year for the past half century, conventional breeding approaches have generated hundreds of rice cultivars with improved yield and quality to keep pace with increased demands for food from a growing world population. In addition to increased yield and quality, the development of GSR varieties also focuses on providing resistance to multiple biotic and abiotic stresses and on the efficient use of nutrients. A large number of genes from various sources is required to confer all these traits, and combining them all in a single cultivar would be very difficult to accomplish with conventional technologies. However, our rapidly increasing understanding of the evolution, function and regulation of key rice genes and traits, and the development of sophisticated technical platforms for phenotyping and molecular breeding, means that it is increasingly becoming scientifically feasible and technically practical to generate GSR by genomic breeding (FIG. 1).

Conceptually, genomic breeding comprises two major components: genomic design and whole-genome selection. Using a specific elite cultivar as the starting point, the design phase involves matching a list of required trait improvements (such as targeted growing areas, high yield, high grain quality, efficient nutrient use, efficient water use and resistance to major diseases and insects, among others) to a list of genes that can generate the desired phenotypes. Germplasm resources for these target genes should be identified, and strategies for assembling the genes and introducing them into the elite cultivar should be determined. Clearly, different rice-growing regions face different challenges in terms of resources and environment, and each region will therefore require its own combination of green traits and the development of its own GSR variety. GSR design will also be affected by the rapid socio-economic development that is occurring in many rice-producing countries, such as China, where drastic changes in rice production systems are taking place. Thus, GSR breeding programmes should be updated regularly to include traits that would increase the efficiency of rice production in the new agricultural systems.

The selection system has two components: a platform for genotyping the plants with respect to target genes and genomic background and a gene-specific selection system. The genotyping platform takes the form of a breeding chip. Several breeding chips[108–110] have been developed based on genetic variation data obtained from large-scale resequencing projects and on gene function data obtained from the literature reporting characterization of individual genes. The gene-specific selection system should comprise a functional marker to select for each target gene and tightly linked DNA markers that flank each target gene (FIG. 7) to detect recombination between the donor and recipient genomes and to facilitate selection of lines in which the target gene has been incorporated precisely. The scientific literature contains a large amount of functional information about rice genes, which can be used to develop gene-specific selection systems.

This selection system provides an efficient means for targeted improvement of any traits in rice for which a causative locus has been identified. For example, genomic breeding has been used to generate a fungal blast-resistant multiline series (Q.Z., unpublished data and F. Zhou, personal communication): four blast-resistance genes (*PI1*, *PI2* (also known as *PIZ*), *PI9* and *PIGM*) were individually incorporated into the genome of Kongyu131, the most widely cultivated rice cultivar in the Heilongjiang province of northeast China. The entire process involved one cross, four backcrosses and one generation of selfing (FIG. 7). During the breeding process, the 6K breeding chip[108] and the gene-specific selection system developed for each gene facilitated selection for the target genes, recombination events and the genomic background. Four near isogenic lines (NILs) were obtained, each of which contained a very short DNA fragment (<200 kb) that contained one of the relevant blast-resistance genes from the corresponding donor line. A large-scale (~500 Ha) field test showed that these lines were phenotypically highly uniform with the required levels of resistance, both in pure stands of individual lines and as a mixture of all four lines. This promising strategy for generating multilines with durable resistance

Breeding chip
A microarray that enables high-throughput genotyping and selection of offspring in breeding programmes.

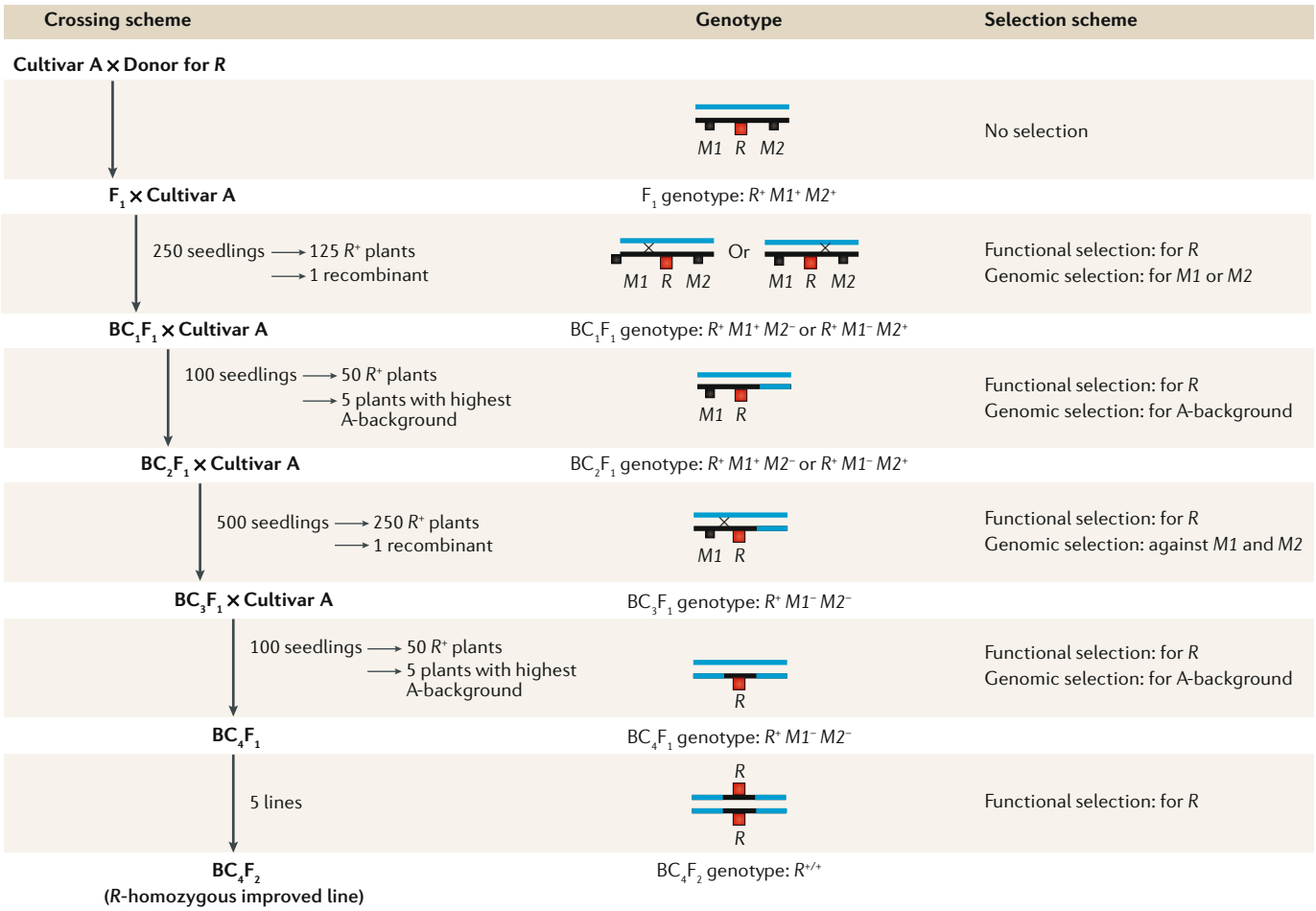


Fig. 7 | **A genomic breeding scheme for precisely introducing a gene into the background of an elite cultivar.** A target gene (*R*) is introduced into the background of Cultivar A by successive backcrossing. Selection is performed using a gene-specific selection system consisting of a functional marker for selecting the target gene (forward selection) and DNA markers on each side of the target gene (*M1* and *M2*) to identify recombination between the target gene and the genetic background (backward selection). The DNA markers should each be <100 kb from the target gene. The numbers of plants indicated for each generation are expectations based on an average physical to genetic distance ratio of 250 kb/cM across the rice genome.

under large-scale production conditions illustrates two distinct advantages of genomic breeding compared with conventional breeding. The first is speed: improved homozygous lines were obtained in 2.5 years (with three generations per year), whereas 5–6 years are required using conventional breeding methods. The second, and most distinct, advantage is the precise incorporation of the target gene into the genome of the recipient line. This feature allows a series of NILs to be generated, each differing from the others by only a very small genomic fragment, which provides genetically diverse but phenotypically uniform resistance in a rice field. A large number of genes that confer blast resistance have been cloned in rice that could be used to generate a large number of NILs via this multiline approach. Spatial and/or temporal rotation of different mixtures comprising four to five NILs from the larger group might provide a strategy for durable resistance in rice production. In addition, genes for different traits can be easily combined into a homogenous genetic background with a single cross. Thus, GSR varieties could be developed by incorporating multiple genes, each for a different green trait, into a single line.

Conclusions and future perspectives

Cultivars with individual GSR traits (such as blast resistance, brown planthopper resistance, drought resistance or more efficient nitrogen use) have already been developed and planted in large areas of China and other Asian countries and in demonstration trials in several African countries. Furthermore, major projects for developing GSR have been initiated in China and by the Bill and Melinda Gates Foundation for International Cooperation.

Recent technological advances in genome biology (such as sequencing, genome editing and molecular breeding) have opened the door for plant biologists to rapidly explore and test an unlimited supply of natural variation across the *Oryza* genus at the DNA sequence level. The most urgent and challenging task facing the community now is how to associate this natural sequence variation with phenotypes under varied ecosystems. To close this genotype-to-phenotype gap, we propose that rice researchers around the world partner with international rice centres (such as IRRI, AfricaRice and International Center for Tropical Agriculture (CIAT)). Each centre would perform baseline high-throughput

phenotyping on large panels of resequenced rice accessions (such as the 3K RGP accession data set) under varied field conditions. International topic-specific teams would interrogate these panels for their specific suite of traits, such as agronomic performance, root architecture, grain quality and microbiomes, among others. All data would be deposited in a central database (such as SNP-Seek at the International Rice Informatics Consortium (IRIC)) and used to guide and accelerate the genomic breeding process to develop GSR cultivars adapted to anywhere in the world. Baseline phenotyping work could be funded by the CGIAR centre donors and topic-specific interrogations by national granting agencies. This approach, if adopted, could serve as a model for other crop systems.

Indeed, progress towards the development of GSR is already setting a paradigm for breeding goals in other crops. For example, in 2017, the Ministry of Agriculture in China implemented Green Varieties, a new varietal evaluation and certification system for the release of major crop varieties that require less fertilizer, pesticide or water. This scheme has, for the first time, opened the door for the recognition of green varieties in an official varietal certification system. Such initiatives will foster the development of green agriculture systems and contribute to the ultimate aim of ensuring sustainable food production for the 10 billion people predicted to populate the world by 2050.

Published online: 05 June 2018

1. United Nations Department of Economic and Social Affairs/Population Division. World population prospects: key findings and advance tables. *ESA* https://esa.un.org/unpd/wpp/publications/Files/WPP2017_KeyFindings.pdf (2017).
2. Zhang, Q. Strategies for developing green super rice. *Proc. Natl Acad. Sci. USA* **104**, 16402–16409 (2007).
This paper introduces the concept of GSR.
3. Matsumoto, T. et al. The map-based sequence of the rice genome. *Nature* **436**, 793–800 (2005).
This paper describes the finished sequence of the rice genome, the first crop genome to be sequenced.
4. Yang, C., Yang, Z., Hu, J., He, G. & Shu, L. Study on the brown planthopper resistance in introgressive lines from wild rice. *Acta Phytophylac. Sin.* **26**, 197–202 (1999).
5. Buckler, E. S., Thornsberry, J. M. & Kresovich, S. Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77**, 213–218 (2001).
6. Caicedo, A. L. et al. Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet.* **3**, 1745–1756 (2007).
7. Civáň, P., Craig, H., Cox, C. J. & Brown, T. A. Three geographically separate domestications of Asian rice. *Nat. Plants* **1**, 15164 (2015).
8. Molina, J. et al. Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl Acad. Sci. USA* **108**, 8351–8356 (2011).
9. Huang, X. et al. A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497–501 (2012).
10. Choi, J. Y. et al. The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* **34**, 969–979 (2017).
This study uses multiple reference genomes and demographic models to support a single domestication event for japonica rice followed by introgression to produce indica and aus varieties.
11. Li, C., Zhou, A. & Sang, T. Rice domestication by reducing shattering. *Science* **311**, 1936–1939 (2006).
12. Stevens, C. J. et al. Between China and South Asia: a middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *Holocene* **26**, 1541–1555 (2016).
13. Sweeney, M. T. et al. Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* **3**, e133 (2007).
14. Tan, L. et al. Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* **40**, 1360–1364 (2008).
15. Wang, M. et al. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat. Genet.* **46**, 982–988 (2014).
This paper reports the first available reference genome for African rice and provides evidence for convergent yet independent selection of a common set of genes during two geographically and culturally distinct domestication processes.
16. Meyer, R. S. et al. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat. Genet.* **48**, 1083–1088 (2016).
17. Olsen, K. M. & Purugganan, M. D. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**, 941–950 (2002).
18. Olsen, K. M. et al. Selection under domestication: evidence for a sweep in the rice waxy genomic region. *Genetics* **173**, 975–983 (2006).
This study provides an early demonstration of the genomic footprint of selection associated with a culturally significant trait.
19. Tan, Y. F. et al. The three important traits for cooking and eating quality of rice grains are controlled by a single locus in an elite rice hybrid, Shanyou 63. *Theor. Appl. Genet.* **99**, 642–648 (1999).
20. Yang, T. et al. The role of a potassium transporter OsHAK5 in potassium acquisition and transport from roots to shoots in rice at low potassium supply levels. *Plant Physiol.* **166**, 945–959 (2014).
21. Ruan, S. L. et al. Proteomic identification of OsCYP2, a rice cyclophilin that confers salt tolerance in rice (*Oryza sativa* L.) seedlings when overexpressed. *BMC Plant Biol.* **11**, 34 (2011).
22. Xie, W. et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc. Natl Acad. Sci. USA* **112**, E5411–5419 (2015).
This study identifies two major groups of indica rice, based on breeding signatures, that resulted from independent breeding activities in different regions of Asia.
23. Wang, J. et al. Artificial selection of Gn1a plays an important role in improving rice yields across different ecological regions. *Rice* **8**, 37 (2015).
24. Spielmeyer, W., Ellis, M. H. & Chandler, P. M. Semidwarf (sd-1), "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl Acad. Sci. USA* **99**, 9043–9048 (2002).
25. Hoque, M. S., Masle, J., Udvardi, M. K., Ryan, P. R. & Upadhyaya, N. M. Over-expression of the rice OsAMT1-1 gene increases ammonium uptake and content, but impairs growth and development of plants under high ammonium nutrition. *Funct. Plant Biol.* **33**, 153–163 (2006).
26. Khush, G. S., Mackill, D. J. & Sidhu, G. S. in *Bacterial Blight of Rice* (eds Banta, S. J., Cervantes, E. & Mew, T. W.) 207–217 (International Rice Research Institute, 1989).
27. Sun, X. et al. Xa26, a gene conferring resistance to *Xanthomonas oryzae* pv. oryzae in rice, encodes an LRR receptor kinase-like protein. *Plant J.* **37**, 517–527 (2004).
28. Kazama, T. & Toriyama, K. A pentatricopeptide repeat-containing gene that promotes the processing of aberrant atp6 RNA of cytoplasmic male-sterile rice. *FEBS Lett.* **544**, 99–102 (2003).
29. The 3000 Rice Genomes Project. The 3,000 rice genomes project. *GigaScience* **3**, 7 (2014).
30. Wang, W. et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **557**, 43–49 (2018).
This paper describes pan-genome analyses that discovered, of 29 million SNPs, 2.4 million short indels, over 90,000 structural variants and more than 10,000 novel genes, all of which contribute to population variation in Asian cultivated rice. Additionally, detected patterns of introgression at several domestication genes support multiple independent domestications in Asian rice.
31. Zhang, J. et al. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA* **113**, E5163–E5171 (2016).
32. Du, H. et al. Sequencing and de novo assembly of a near complete indica rice genome. *Nat. Commun.* **8**, 15324 (2017).
33. Goff, S. A. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). *Science* **296**, 92–100 (2002).
34. Yu, J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp indica). *Science* **296**, 79–92 (2002).
35. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
36. Yamamoto, T. et al. Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* **11**, 267 (2010).
37. Takagi, H. et al. MutMap-Gap: whole-genome resequencing of mutant F2 progeny bulk combined with de novo assembly of gap regions identifies the rice blast resistance gene Pii. *New Phytol.* **200**, 276–283 (2013).
38. Lu, L. et al. Tracking the genome-wide outcomes of a transposable element burst over decades of amplification. *Proc. Natl Acad. Sci. USA* **114**, E10550–E10559 (2017).
39. Yu, J. et al. The Genomes of *Oryza sativa*: a history of duplications. *PLoS Biol.* **3**, e38 (2005).
40. Reddy, M. M. & Ulaganathan, K. Draft genome sequence of *Oryza sativa* elite indica cultivar RP Bio-226. *Front. Plant Sci.* **6**, 896 (2015).
41. Mahesh, H. B. et al. Indica rice genome assembly, annotation and mining of blast disease resistance genes. *BMC Genomics* **17**, 242 (2016).
42. Stein, J. C. et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nat. Genet.* **50**, 285–296 (2018).
43. Zhang, Y. et al. Genome and comparative transcriptomics of African wild rice *Oryza longistaminata* provide insights into molecular mechanism of rhizomatousness and self-incompatibility. *Mol. Plant* **8**, 1683–1688 (2015).
44. Chen, J. F. et al. Whole-genome sequencing of Oryza brachyantha reveals mechanisms underlying Oryza genome evolution. *Nat. Commun.* **4**, 1595 (2013).
45. Jacquemin, J., Bhatia, D., Singh, K. & Wing, R. A. The international *Oryza* map alignment project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question. *Curr. Opin. Plant Biol.* **16**, 147–156 (2013).
46. Yang, W. et al. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* **5**, 5087 (2014).
This work reports a high-throughput phenotyping facility and demonstrates that the data can be used for GWAS of agronomic traits.
47. Tanger, P. et al. Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. *Sci. Rep.* **7**, 42839 (2017).
48. Vadez, V. et al. LeasyScan: a novel concept combining 3D imaging and lysimetry for high-throughput phenotyping of traits controlling plant water budget. *J. Exp. Bot.* **66**, 5581–5593 (2015).
49. Shi, Y. et al. Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS ONE* **11**, e0159781 (2016).

50. Courtois, B. et al. Genome-wide association mapping of root traits in a japonica rice panel. *PLoS ONE* **8**, e78037 (2013).
51. Rebolledo, M. C. et al. Phenotypic and genetic dissection of component traits for early vigour in rice using plant growth modelling, sugar content analyses and association mapping. *J. Exp. Bot.* **66**, 5555–5566 (2015).
52. Qiu, X. et al. Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS ONE* **10**, e0145577 (2015).
53. Rebolledo, M. C. et al. Combining image analysis, genome wide association studies and different field trials to reveal stable genetic regions related to panicle architecture and the number of spikelets per panicle in rice. *Front. Plant Sci.* **7**, 1384 (2016).
54. Kikuchi, S. et al. Genome-wide association mapping for phenotypic plasticity in rice. *Plant Cell Environ.* **40**, 1565–1575 (2017).
55. Al-Tamimi, N. et al. Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat. Commun.* **7**, 13342 (2016).
56. Sakai, H. et al. Rice annotation project database (RAP-DB): an integrative and interactive database for rice genomics. *Plant Cell Physiol.* **54**, e6 (2013).
57. Zhang, J. et al. Building two indica rice reference genomes with PacBio long-read and Illumina paired-end sequencing data. *Sci. Data* **3**, 160076 (2016).
58. Song, J. M. et al. Rice information gateway: a comprehensive bioinformatics platform for Indica rice genomes. *Mol. Plant* **11**, 505–507 (2018).
59. McCouch, S. R. et al. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* **7**, 10532 (2016).
60. Mansueto, L. et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res.* **45**, D1075–D1081 (2017).
61. Yao, W., Li, G., Yu, Y. & Ouyang, Y. funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *Gigascience* **7**, 1–9 (2018).
62. Zhang, H. T. & Wang, S. P. Progress in functional genomic studies of rice disease resistance. *Chinese Bulletin of Life Sciences.* **28**, 1189–1199 (2016).
63. Deng, Y. et al. Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* **355**, 962–965 (2017).
64. Li, W. et al. A natural allele of a transcription factor in rice confers broad-spectrum blast resistance. *Cell* **170**, 114–126 (2017).
65. Hu, K. et al. Improvement of multiple agronomic traits by a disease resistance gene via cell wall reinforcement. *Nat. Plants* **3**, 17009 (2017).
66. Zhao, Y. et al. Allelic diversity in an NLR gene BPH9 enables rice to combat planthopper variation. *Proc. Natl Acad. Sci. USA* **8**, 12850–12855 (2016).
67. Hu, J., Xiao, C. & He, Y. Recent progress on the genetics and molecular breeding of brown planthopper resistance in rice. *Rice* **9**, 30 (2016).
68. Ke, Y., Deng, H. & Wang, S. Advances in understanding broad-spectrum resistance to pathogens in rice. *Plant J.* **90**, 738–748 (2017).
69. Lin, J. Y. & Shen, M. *Rice production constraints in China* (CAB International, 1996).
70. Hu, H. & Xiong, L. Genetic engineering and breeding of drought-resistant crops. *Annu. Rev. Plant Biol.* **65**, 715–741 (2014).
71. Uga, Y. et al. Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. *Nat. Genet.* **45**, 1097–1102 (2013).
72. Xu, K. et al. Sub1A is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708 (2006).
This paper describes the cloning and characterization of the *SUB1A* gene. This gene has since been introgressed into Asian mega-varieties to help withstand submergence flooding of up to 2 weeks.
73. Hattori, Y. et al. The ethylene response factors SNORKEL1 and SNORKEL2 allow rice to adapt to deep water. *Nature* **460**, 1026–1030 (2009).
74. Fujino, K. et al. Molecular identification of a major quantitative trait locus, qLTG3-1, controlling low-temperature germinability in rice. *Proc. Natl Acad. Sci. USA* **105**, 12623–12628 (2008).
75. Lu, G. et al. Rice LTG1 is involved in adaptive growth and fitness under low ambient temperature. *Plant J.* **78**, 468–480 (2014).
76. Ma, Y. et al. COLD1 confers chilling tolerance in rice. *Cell* **160**, 1209–1221 (2015).
77. Li, X. M. et al. Natural alleles of a proteasome α2 subunit gene contribute to thermotolerance and adaptation of African rice. *Nat. Genet.* **47**, 827–833 (2015).
78. Liu, J. et al. The RING finger ubiquitin E3 ligase OsHTAS enhances heat tolerance by promoting H2O2-induced stomatal closure in rice. *Plant Physiol.* **170**, 429–443 (2016).
79. Ren, Z. H. et al. A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat. Genet.* **37**, 1141–1146 (2005).
80. Hu, B. et al. Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* **47**, 834–838 (2015).
81. Sun, H. et al. Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Genet.* **46**, 652–656 (2014).
82. Fan, X. et al. Overexpression of a pH-sensitive nitrate transporter in rice increases crop yields. *Proc. Natl Acad. Sci. USA* **113**, 7118–7123 (2016).
83. Gamuyao, R. et al. The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. *Nature* **488**, 535–539 (2012).
84. Yamaji, N. et al. Reducing phosphorus accumulation in rice grains with an impaired transporter in the node. *Nature* **541**, 92–95 (2017).
85. Xing, Y. & Zhang, Q. Genetic and molecular bases of rice yield. *Annu. Rev. Plant Biol.* **61**, 421–442 (2010).
86. Wang, Y. & Li, J. Molecular basis of plant architecture. *Annu. Rev. Plant Biol.* **59**, 253–279 (2008).
87. Xue, W. et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **40**, 761–767 (2008).
88. Wei, X. et al. DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiol.* **153**, 1747–1758 (2010).
89. Yan, W.-H. et al. A major QTL. *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Mol. Plant* **4**, 319–330 (2011).
90. Yan, W. et al. Natural variation in Ghd7.1 plays an important role in grain yield and adaptation in rice. *Cell Res.* **23**, 969–971 (2013).
91. Zhang, Z.-H. et al. Pleiotropism of the photoperiod-insensitive allele of Hd1 on heading date, plant height and yield traits in rice. *PLoS ONE* **7**, e52538 (2012).
92. Zhang, J. et al. Combinations of the Ghd7, Ghd8 and Hd1 genes largely define the ecogeographical adaptation and yield potential of cultivated rice. *New Phytol.* **208**, 1056–1066 (2015).
93. Bai, X. et al. Duplication of an upstream silencer of FZP increases grain yield in rice. *Nat. Plants* **3**, 885–893 (2017).
94. Zhang, L. et al. A natural tandem array alleviates epigenetic repression of IPA1 and leads to superior yielding rice. *Nat. Commun.* **8**, 14789 (2017).
95. Zhang, D. & Yuan, Z. Molecular control of grass inflorescence development. *Annu. Rev. Plant Biol.* **65**, 553–578 (2014).
96. Chen, J. et al. A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. *Proc. Natl Acad. Sci. USA* **105**, 11436–11441 (2008).
97. Long, Y. et al. Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl Acad. Sci. USA* **105**, 18871–18876 (2008).
98. Mizuta, Y., Harushima, Y. & Kurata, N. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl Acad. Sci. USA* **107**, 20417–20422 (2010).
99. Yang, J. et al. A killer-protector system regulates both hybrid sterility and segregation distortion in rice. *Science* **337**, 1336–1340 (2012).
This study characterizes the *S5* locus for reproductive isolation between indica and japonica subspecies, which consists of three adjacent genes forming a killer–protector system. This gene is widely used in intersubspecific hybrid rice breeding.
100. Yu, Y., Wing, R. A. & Li, J. in *Genetics and Genomics of Rice* (eds Zhang, Q. & Wing, R. A.) 237–254 (Springer-Verlag, 2013).
101. Gao, Z. et al. Map-based cloning of the ALK gene, which controls the gelatinization temperature of rice. *Sci. China C. Life Sci.* **46**, 661–668 (2003).
102. Li, Y. et al. Chalk5 encodes a vacuolar H(+)-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* **46**, 398–404 (2014).
103. Fan, C. et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor. Appl. Genet.* **112**, 1164–1171 (2006).
104. Weng, J. et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18**, 1199–1209 (2008).
105. Shomura, A. et al. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **40**, 1023–1028 (2008).
106. Ye, X. et al. Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* **287**, 303–305 (2000).
107. Zhu, Q. et al. Development of "Purple Endosperm Rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Mol. Plant* **10**, 918–929 (2017).
108. Yu, H., Xie, W., Li, J., Zhou, F. & Zhang, Q. A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* **12**, 28–37 (2014).
109. Chen, H. et al. A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant* **7**, 541–553 (2014).
110. Singh, N. et al. Single-copy gene based 50 K SNP chip for genetic studies and molecular breeding in rice. *Sci. Rep.* **5**, 11600 (2015).
111. Ge, S., Sang, T., Lu, B. R. & Hong, D. Y. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc. Natl Acad. Sci. USA* **96**, 14400–14405 (1999).
112. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
113. Purugganan, M. D. An evolutionary genomic tale of two rice species. *Nat. Genet.* **46**, 931–932 (2014).

Acknowledgements
R.A.W. was supported by the Bud Antle Endowed Chair of Excellence in Agriculture and Life Sciences, the AXA Research Fund and NIFA-HATCH ARZT-1360510-H25-230. M.D.P. was supported by grants from the US National Science Foundation Plant Genome Research Program, the Zegar Family Foundation and the New York University Abu Dhabi Research Institute. Q.Z. was supported by grants from the National 863 Program 2104AA10A604, the National Key Research and Development Program 2016YFD0100903, the Earmarked Fund for the China Agriculture Research System of China (CARS-01-05) and the Bill and Melinda Gates Foundation. The authors also thank K. McNally and S. Klassen for critically reading the manuscript prior to publication.

Author contributions
All authors contributed to all aspects of writing this Review.

Competing interests
The authors declare no competing interests.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information
Nature Reviews Genetics thanks Guo-Liang Wang, Jean Christophe Glaszmann, and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Supplementary information
Supplementary information is available for this paper at https://doi.org/10.1038/s41576-018-0024-z.

RELATED LINKS
Africa Rice: http://www.africarice.org/warda/genebank.asp
Consultative Group and International Agricultural Research (CGIAR): https://www.cgiar.org/
Ensemble Plants: http://plants.ensembl.org/index.html
funRiceGenes: http://funricegenes.ncpgr.cn/
GenBank: https://www.ncbi.nlm.nih.gov/assembly/?term=Oryza
Global Rice Phenotyping Network: http://ricephenonetwork.irri.org/
Gramene: http://oge.gramene.org/genome_browser/index.html
International Center for Tropical Agriculture (CIAT): http://ciat.cgiar.org/
International Rice Genebank: http://irri.org/our-work/research/genetic-diversity/international-rice-genebank
International Rice Informatics Consortia (IRIC): http://iric.irri.org/
International Rice Research Institute (IRRI): www.irri.org
Rice Annotation Project (RAP): http://rapdb.dna.affrc.go.jp/
Michigan State University DB (MSU-DB): http://rice.plantbiology.msu.edu/
Oryzabase: https://shigen.nig.ac.jp/rice/oryzabase/
Rice Information GateWay (RIGW): http://rice.hzau.edu.cn/cgi-bin/rice/download_ext
R498 at MBKBASE: www.mbkbase.org/R498
Rice Diversity database: http://www.ricediversity.org/data/index.cfm
SNP-Seek: http://snp-seek.irri.org/
TERRA-REF Field Scanalyzer in Arizona: www.terraref.org