The International Oryza Map Alignment Project: development of a genus-wide comparative genomics platform to help solve the 9 billion-people question

Julie Jacquemin1,3, Dharminder Bhatia1,3, Kuldeep Singh2 and Rod A Wing1

The wild relatives of rice contain a virtually untapped reservoir of traits that can be used help drive the 21st century green revolution aimed at solving world food security issues by 2050. To better understand and exploit the 23 species of the Oryza genus the rice research community is developing foundational resources composed of: 1) reference genomes and transcriptomes for all 23 species; 2) advanced mapping populations for functional and breeding studies; and 3) in situ conservation sites for ecological, evolutionary and population genomics. To this end, 16 genome sequencing projects are currently underway, and all completed assemblies have been annotated; and several advanced mapping populations have been developed, and more will be generated, mapped, and phenotyped, to uncover useful alleles. As wild Oryza populations are threatened by human activity and climate change, we also discuss the urgent need for sustainable in situ conservation of the genus.

Addresses
1 Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, United States
2 School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India

Corresponding author: Wing, Rod A (rwing@ag.arizona.edu)

3 Co-first authors.

Current Opinion in Plant Biology 2012, 16:xx–yy
This review comes from a themed issue on Genome studies and molecular genetics
Edited by Qifa Zhang and Rod Wing

1369-5266/$ – see front matter, © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.pbi.2013.02.014

Introduction
In 2010, Science magazine published a special issue entitled ‘Feeding the Future’ that featured a series of articles concerning food security and the challenge of feeding 9 billion people [1], followed by an issue in the Economist magazine entitled ‘The 9-billion People Question’ (9BPQ) [2]. Both addressed a worldwide dilemma that is central to most if not all plant scientists, that is, how can our society grow enough food to feed 2 billion additional human beings in less than 40 years? Rice (Oryza sativa) will play a key role in helping to solve the 9BPQ, as it presently provides 20% or more daily calories to half the world’s population, and will be the developing world’s most important food crop in 2050 [3]. Rice 2020 [4] is a key initiative aimed at solving the 9BPQ and calls for community mobilization to pool and coordinate all available resources with the common goal of creating new green super rice varieties, where ‘green’ means less input (e.g. water, fertilizer, pesticides, land), and ‘super’ means two- to threefold yields [5].

One of the most important resources that can be utilized to improve cultivated rice is the virtually untapped reservoir of genetic variation hidden within the wild relatives of Oryza. The genus Oryza spans approximately 15 MY of evolutionary history (Figure 1) [6] and is composed of 21 wild and 2 domesticated (O. sativa, Oryza glaberrima) species, 10 distinct genome types (AA, BB, CC, BBCC, CCDD, EE, FF, GG, KKLL, HHJJ), and a 3.6 genome size variation. Wild Oryza species have a broad habitat distribution, including Asia, Australia, Africa, South and Central America, and many novel biotic/abiotic resistances have been identified.

To lay the foundation for a complete genomic interrogation of the wild relatives of rice the Oryza Map Alignment (OMAP) and Oryza Genome Evolution (OGEP) Projects were funded to generate a large array of publicly available genomic resources, most notably a set of manually edited BAC-based physical maps (i.e. 18 deep-cover BAC libraries — fingerprinted, end-sequenced, and FPC assembled) representing 17 of the 23 recognized Oryza species, covering all eight AA genome species and one each of the other nine genome types (BB, CC, BBCC, CCDD, EE, FF, GG, KKLL, HHJJ) [6–8]; and a set of chromosome three short arm sequences from all eight AA genome species, as well as the BB, CC, BBCC, FF, GG and Leersia perrieri, an Oryza outgroup species. All of these data and resources can be accessed through the http://www.Gramene.org and http://www.genome.arizona websites, respectively.

Analysis of these data sets revealed the following key points: first, LTR Retro-transposable element amplifications dramatically increased the size of both the Oryza
2 Genome studies and molecular genetics

Phylogenetic tree of the *Oryza* genus and outgroup *Leersia perrieri* (modified from [6]). Targeted OGE/I-OMAP species are indicated on each branch. Arrows indicate origins of allotetraploids. Dark circles indicate maternal parents. Open circles indicate unidentified diploid species. Divergence time is based on data from the literature [13,67,68,69].

*australiensis* [EE] and the *Oryza granulata* [GG] genomes by as much as 400 and 200 Mb, respectively [9,10]; second, the AA genomes of *Oryza nivara*, *Oryza rufipogon* (the putative progenitor species of *O. sativa*), and *O. glaberrima* have expanded/contracted by at least 40 Mb (>10% of their genome sizes) relative to the IRGSP RefSeq [11]; and third, analysis of the *Adh1* region (~100–200 kb) across the entire *Oryza* phylogeny (diploid and polyploid) showed significant perturbations of synteny including dynamic evolution of gene families, transposable element mediated gene movement, mutations and large scale physical rearrangements [12,13].

The overriding conclusion from these studies, and from many others (e.g. [14]), indicates that a SINGLE reference genome for the genus *Oryza* (i.e. IRGSP RefSeq) is insufficient to capture and understand the allelic diversity/natural variation hidden within the genus to help solve the 9BPQ.

To address this resource/knowledge gap the International *Oryza* Map Alignment Project (I-OMAP) was initiated in 2007 and has held six grand challenge meetings (Japan 07, Korea 08, Philippines 09, Brazil 10, Taiwan 11, Thailand 12) in conjunction with the annual International Symposia on Rice Functional Genomics (ISRFG). The three
primary focus areas of I-OMAP are to: first, generate RefSeqs & Transcriptome data sets for all eight AA genome species, and representative species of the nine other genome types; second, generate, map, and phenotype advanced ABC, CSSL, RIL, populations for the AA genome species for functional and breeding studies; and third, identify collections of naturally occurring populations of the wild *Oryza* species for diversity and evolutionary analyses, and for conservation. Here all three I-OMAP focus areas will be discussed.

### Sequencing the collective Oryza genome

As stated, a major goal of focus area 1 is to generate reference quality sequences from representatives of all 23 *Oryza* species. Table 1 lists the current status of each I-OMAP genome project to date. Sixteen of the 23 genome sequencing projects are in progress or have been completed, and include all of the diploid *Oryza* species. Draft sequences of two subspecies of *O. sativa* were published a decade ago [15,16], followed by the release of IRGSP ‘gold standard’ RefSeq of *O. sativa* ssp. *japonica* (cv. Nipponbare) in 2005 [17]. Significant progress has been achieved over the past two years with completion of the *O. glaberrima* [AA], *Oryza barthii* [AA], *Oryza longistaminata* [AA], *O. punctata* [BB], and *O. brachyantha* [FF] genomes (all unpublished but in GenBank for early access). Assembly is currently in progress for the *O. nicara* [AA], *Oryza glumaepatula* [AA] and *O. meridionalis* [AA] genomes, and sequencing is underway for a majority of the remaining diploid species.

It should be noted that the I-OMAP project has a significant advantage over other next generation genome sequencing projects (e.g. *Drosophila* 12 genomes) [18] in that physical maps are available for all AA genome species as well as representatives of all other nine genome types. Such resources facilitate the assembly of more complete genome sequences versus ones that rely solely on next generation short-read sequence data and assembly algorithms, the so called ‘gene space assemblies’.

Once the assemblies are linked to their respective physical maps and edited for mis-assemblies and errors, they will then be annotated using a common annotation platform—for example, MAKER [19]. To date, all available assemblies have been annotated, including both *O. sativa* subspecies *japonica* and *indica*. It is important that all the *Oryza* genome be annotated using a common platform for ease of downstream comparative structural and evolutionary analyses.

In the future, I-OMAP plans to sequence all the polyploid *Oryza* species but has yet to secure funding. The BBCC genome of *O. minuta* will likely be the first one attempted as RefSeqs for both the CC and the BB genomes will soon be available, which can help to guide the assembly process of the BBCC genome. In addition a robust physical map for the *O. minuta* genome is available and thus a combination of WGS and BAC pool sequencing can be used to tame this wild genome.

### Unfolding the genetic architecture of the wild relatives of rice with advanced mapping populations

The wild relatives of rice constitute an important reservoir of valuable genes but the association of these species with several weedy traits and incompatibility barriers has limited the transfer of useful genes into
cultivated species. The major consideration in alien gene transfer is to selectively transfer agronomically and commercially important genes from wild species, while at the same time avoiding linkage drag, using a combination of strategies involving conventional and molecular plant breeding and tissue culture [20]. Most of the important traits present in the wild relatives of rice are cryptic and need to be extracted in the cultivated species background. A substantial amount of ‘pre-breeding’ research can provide a useful platform to identify and transfer favorable alleles from wild and unadapted sources into elite cultivars [21].

Conventional mapping populations including F₂, backcross, recombinant inbred lines (RILs), and doubled haploids (DH) have been used for mapping both major genes and quantitative trait loci (QTL). Although such populations are relatively simple to generate, they have

---

**Box 1 Advanced mapping populations**

**Backcross Inbred Lines (BILs).** BILs are derived by crossing a wild species with a cultivated one, followed by backcrossing with the cultivated parent two or more times and selecting for a target trait in each generation. For the purposes of genetic analysis, the advantages of using BILs include: first, high genetic and morphological similarity between lines that enables more precise estimates of quantitative traits; second, the opportunity to study QTL-by-environment interactions more accurately; and third, relatively rapid and straightforward utilization of BILs for commercial plant breeding [37].

**Near Isogenic Lines (NILs).** An alternative type of immortal experimental population commonly used in breeding is the development of sets of introgression lines or NILs. Unlike BILs, NILs differ at a limited number of loci. Small-effect QTL can be detected more precisely with NILs than with RIL populations, although the local resolution may be lower. In general, population size is more important than the number of replicates to increase the mapping power of RILs, whereas for NILs, many replicates are absolutely required [59].

**Advanced Intercrossed Lines (AILs).** AILs are produced by randomly, but sequentially, intercrossing a population that initially originated from a cross between two inbred lines or some variant, thereof. AILs provide an increased probability of recombination between any two loci. Consequently, the genetic length of the entire genome is stretched, and QTL are identified with higher precision as compared with conventional RIL populations [39]. AILs derived from crosses between known inbred lines may be a useful resource for fine genetic mapping as well; however, it takes additional time to generate a new set of inbreds after intermating.

**Chromosome Segment Substitution Lines (CSSLs).** A CSSL is an introgression line that contains a small portion of a given chromosome introgressed from another line, and a set of CSSLs encompass an entire chromosome. CSSLs are developed both from interspecific and interspecific crosses with a series of backcrosses and the identification of individual lines with molecular markers. Genome-wide CSSLs can be used as powerful QTL mapping populations for the elucidation of the molecular basis of interesting traits especially derived from wild species [23]. A major limitation of using CSSLs is that unidentified introgression of small targeted chromosomal segments, which are not tagged by markers, sometimes generate experimental noise that makes it more difficult to detect QTL effects in particular chromosomal regions. Secondly, it may be difficult to detect phenotypic differences generated by a combination of two or more donor alleles in different chromosomal regions. In this case, the mapping resolution of respective QTL can be improved by fine mapping to develop NILs using the CSSLs/BILs as base materials [22]. Graphical genotyping software programs, such as GGT [40] or CSSL Finder ([http://mapdisto.free.fr/CSSLFinder](http://mapdisto.free.fr/CSSLFinder)), are very useful for the development and usage of CSSLs.

**High-Throughput Genotyped CSSLs.** In this case, every CSSL is genotyped by whole-genome re-sequencing, and thus combines the advantages of an ultrahigh-quality physical map with high mapping accuracy. This approach has great potential value for gene discovery and genetic mapping but has not been reported for interspecific crosses of rice. High-throughput genotyped CSSLs may provide powerful tools for future whole-genome large-scale gene discovery in rice and offers foundations to enable the development of superior rice varieties [41].

**Multiparent Advanced Generation Inter-Cross (MAGIC) populations.** Under this scheme, as many diverse founder lines as an investigator wishes to deploy are used to create a mapping population. If n founder lines are chosen, they need to be intercrossed for n² generations until all founders are combined in equal proportion. Once the intercrossing is completed, RILs may be derived from these after selfing [42,43]. The major limitation of the use of MAGIC populations is that with increased founder size, intercrossing cycles also proportionately increase. Another limitation is that MAGIC populations are likely to show extensive segregation for phenological traits, like maturity and plant height. Segregation for such traits may influence the overall performance of complex traits like yield or drought tolerance, and hence may yield false QTL.

**Association Panels.** Association panels are open mapping populations that utilize samples of individuals from germplasm collections or natural populations. Compared with conventional and linkage disequilibrium (LD) mapping, the use of nonrandom associations of loci in haplotypes is an effective tool to connect structural genomics and phenomics in plants. Association mapping exploits historical recombination events that have occurred during establishment of the sample population [44]. A number of publications have appeared on association mapping in rice. The most comprehensive study was published by Huang et al. [45] who used 950 diverse lines representing the worldwide collection of rice germplasm to map flowering time and grain yield traits. Major limitations of using association panels are population stratification, and an unequal distribution of alleles within a population which results in non-functional, spurious associations.

**Nested Association Mapping (NAM).** NAM populations are developed by pooling equal numbers of progenies from a large number of crosses involving one common parent. NAM, as currently implemented for maize, is an extremely powerful strategy for dissecting the genetic basis of quantitative traits in a species with low LD. The controlled crosses reduce the confounding effects of population structure, while the large numbers of progeny derived from such crosses allow for family mapping with substantial statistical power [46]. The major limitations are the large number of crosses required to generate NAM populations, and QTL interactions with genetic backgrounds cannot be examined as one parent is common in all component subpopulations.

A comparison of all the advanced mapping populations discussed above with respect to practical utility as mapping resources and their suitability toward mapping different genetic components is presented in [Table 2](#).
one or more significant limitations. $F_2$ and backcross mapping populations are the simplest to generate but are not efficient for mapping QTL as they cannot be replicated over multiple environments, and single plant data is not reliable. On the other hand, RIL and DH populations are immortal, and can be replicated over locations and years. However, both undergo limited recombination events, and mapping analyses can be limited because of the masking effects of major QTL and epistatic interactions of multiple QTL. Advanced

## Table 2

<table>
<thead>
<tr>
<th>Properties</th>
<th>RILs/DH</th>
<th>BILs</th>
<th>NILs</th>
<th>ALIs</th>
<th>CSSL</th>
<th>MAGIC</th>
<th>Assoc.</th>
<th>NAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Founder parents</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>$\geq$8</td>
<td>$\geq$100</td>
<td>$\geq$10</td>
</tr>
<tr>
<td>Crossing requirement</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Time to establish</td>
<td>Moderate</td>
<td>Long</td>
<td>Long</td>
<td>Long</td>
<td>Long</td>
<td>Long</td>
<td>Low</td>
<td>Long</td>
</tr>
<tr>
<td>Population size</td>
<td>$\sim$200</td>
<td>$\sim$100</td>
<td>$\sim$100</td>
<td>$&gt;200$</td>
<td>$&gt;200$</td>
<td>$&gt;1000$</td>
<td>$\sim$100</td>
<td>$&gt;1000$</td>
</tr>
<tr>
<td>Suitability for fine mapping</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Amount of genotyping required</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Amount of phenotyping required</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Statistical complexity</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Use of germplasm variation</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Suitability for mapping QTL in interspecific crosses</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Practical utility</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Suitability for measuring genetic components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i. Additive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ii. Dominance</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>iii. Additive $\times$ additive</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>iv. Additive $\times$ dominance</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Modified from [42,43].

## Table 3

<table>
<thead>
<tr>
<th>Population</th>
<th>Donor parent (accession)</th>
<th>Recurrent parent (cultivar)</th>
<th>Traits introgressed/mapped</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL</td>
<td>O. glaberrima (Tog5681)</td>
<td>Indica (IR64)</td>
<td>Drought tolerance</td>
<td>[47]</td>
</tr>
<tr>
<td>CSSL</td>
<td>O. glaberrima (IRGC103544)</td>
<td>Tropical japonica (Caiapo)</td>
<td>Rice stripe necrosis virus resistance</td>
<td>[48]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. glaberrima (Tog5675)</td>
<td>Indica (IR64)</td>
<td>BPH resistance (Bph1)</td>
<td>[49]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. glaberrima (IRGC96717)</td>
<td>Japonica (WAB56-104)</td>
<td>Drought resistance, early vigor</td>
<td>[50]</td>
</tr>
<tr>
<td>CSSL</td>
<td>O. glaberrima</td>
<td>Japonica (Koshihikari)</td>
<td>Glabrous gene</td>
<td>[51]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. glaberrima (IRGC103544)</td>
<td>Indica (Milyang 23)</td>
<td>Yield and yield components</td>
<td>[52]</td>
</tr>
<tr>
<td>CSSL</td>
<td>O. glaberrima</td>
<td>Japonica (Wuyuing-7)</td>
<td>Spreading panicle</td>
<td>[53]</td>
</tr>
<tr>
<td>IL/BIL</td>
<td>O. nivara (IRGC105444)</td>
<td>Japonica (Taichung 65)</td>
<td>Pollen fertility, gene (S27-nivs)</td>
<td>[54]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. nivara (IRGC105444)</td>
<td>Japonica (Koshihikari)</td>
<td>Hybrid breakdown locus [-hbd 1(t)]</td>
<td>[55]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. rufipogon (W630)</td>
<td>Japonica (Nippobaire)</td>
<td>Drought tolerance</td>
<td>[56]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. rufipogon (IRGC105491)</td>
<td>Tropical japonica (Jefferson)</td>
<td>Early flowering</td>
<td>[57]</td>
</tr>
<tr>
<td>BIL/CSSL</td>
<td>O. rufipogon (YJCW)</td>
<td>Indica (Xiejingzao B)</td>
<td>WBPH and BPH resistance</td>
<td>[58]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. rufipogon (YJCW)</td>
<td>Indica (93-11, restorer line)</td>
<td>Yield-related traits</td>
<td>[59]</td>
</tr>
<tr>
<td>CSSL</td>
<td>O. rufipogon</td>
<td>Indica (Teqing)</td>
<td>Small grain panicle and dwarfishity</td>
<td>[60]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. rufipogon (IRGC105491)</td>
<td>Indica (IR64)</td>
<td>Yield and yield components</td>
<td>[61]</td>
</tr>
<tr>
<td>NIL</td>
<td>O. rufipogon (IRGC105491)</td>
<td>Japonica (Hwaseongbyeo)</td>
<td>Yield components</td>
<td>[62]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. rufipogon (YJCW)</td>
<td>Indica (Teqing)</td>
<td>Yield and yield components</td>
<td>[63]</td>
</tr>
<tr>
<td>BIL/NIL</td>
<td>O. longistaminata</td>
<td>Indica (RD23)</td>
<td>Pollen/spikelet fertility, plant height</td>
<td>[64]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. glumeaepatula (RS-16)</td>
<td>Indica (B980-2)</td>
<td>Grain yield, cooking quality</td>
<td>[65]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. minutula (IRGC101141)</td>
<td>Indica (IR3917-45-3-2)</td>
<td>BPH resistance</td>
<td>[49]</td>
</tr>
<tr>
<td>BIL</td>
<td>O. brachyantha (IRGC101232)</td>
<td>Indica (IR56)</td>
<td>Bacterial blight</td>
<td>[66]</td>
</tr>
</tbody>
</table>

Modified from [23].
AMPS developed with the wild relatives of rice help not only to minimize genetic noise and reveal the true expression of individual loci, but also permit the ability to map and clone these cryptic alleles. While in case of interspecific crosses, especially with wild and unadapted germplasm, conventional mapping populations again lose their worth for these additional reasons: first, hybrid sterility often hampers their development and second, progeny derived from wide crosses often exhibit too much phenotypic noise and wide variation in heading date, particularly as a result of strong transgressive segregation. Therefore, it may be difficult to precisely evaluate phenotypic traits among segregants [22]. For transfer and mapping of QTL from wild and unadapted germplasm, BILs, NILs, AILs, CSSL are the most powerful populations that can be used. A number of such populations have already been developed for rice and are summarized in Table 3. Developing AMPs can require more than six years, even with the adoption of accelerated generations [22]. But once developed, they have the potential to uncover new alleles from the unadapted, non-productive wild rice accessions, develop genome-wide genetic stocks and clone identified genes/QTLs for functional genomic research [23].

CSSLs and ILs using different wild rice accessions have or are being developed around the world (e.g. USA, China, India, Japan, the Philippines, South America, and Europe) (for details see [23]). At Punjab Agricultural University in India, BILs have been developed in the background of two indica cvs. PR114 and Pusa44 by crossing these with 72 accessions selected from African cultivated rice O. glaberrima and five AA genome wild species (O. barthii, O. nivara, O. rufipogon, O. longistaminata and O. glumeapatula), followed by two successive backcrosses and selfing. These BILs are currently being evaluated over a five-year period for variation in yield and yield component traits (K Singh, personal communication). Alleles derived from the wild species in the cultivated recurrent parent background have already been shown to contribute to significant increases in yield component traits including spikelets/panicle and grain weight as shown in Figure 2.

With the advent of next generation sequencing, combined with genetic maps and molecular markers, it is now possible to rapidly map and identify regions of the genome associated with specific components of a phenotype and to determine which parental line contributes the favorable allele(s) at a particular locus. These are helpful tools only with precise phenotyping which is critical to the success of gene/QTL mapping experiments. Unfortunately, the importance of refining and developing new methods for precise phenotypic measurements has so far been neglected in the genomics era. Overcoming the phenotyping gap represents the next greatest challenge for scientists involved in the rice genomics research.

Conservation of the wild relatives of rice

Whereas progress in genomics and breeding techniques will certainly enable the successful transfer of complex traits present in the wild relatives of rice, including such divergent species such as Porteresia or Hygropyra, into cultivated rice, the long-term availability of these wild genetic resources is not assured [24]. Risks for populations of the wild relatives of rice, and plant biodiversity overall, include threats related to human activities and climate change [25**]. The expansion of urban and agriculture areas leads to habitat destruction and fragmentation. For example, in India (Andhra Pradesh), a population of O. officinalis ssp. malampuzhaensis is endemic to the Nallamalais of Eastern Ghats, and needs urgent collection and conservation as its narrow distribution makes it more vulnerable to the disruption of its habitat [26]. Populations of O. rufipogon have been threatened by overgrazing and its consequences on water flow in Queensland, Australia [27], and by the construction of buildings in the Central Plains of Thailand [28]. Rising sea levels may be a problem for Oryza populations in the wetlands of Northern-Australia [27]. Populations of O. meridionalis in the same area are threatened by the infestation of an exotic grass (Para Grass) that inhibits germination [29], showing that invasive species are also a potential risk. When populations are fragmented, their genetic diversity is reduced, and they may not be able to adapt to rapid environmental changes. Sustainable conservation of the wild relatives of rice is thus an urgent problem that needs to be addressed before described and undescribed populations that could provide desirable traits of interest are lost forever.

The characterization and collection of species and populations are the first step for conservation and are well documented in many regions of the world where the wild relatives of rice are found [30,31]. However, extensive identification and collection are still needed in some areas, for example, Australia [27] and Venezuela [32]. To implement conservation strategies with an optimum result, it is urgent to understand the genetic diversity and structure among populations. A strong correlation between genetic and geographic distances, as well as high enzymatic polymorphism, was found among O. glumeapatula accessions in South-America [33]. This kind of information is necessary to evaluate the need to protect specific populations, and increased international efforts should be organized to systematically assess genetic diversity and screen germplasm for useful genes [24]. A critical step is the assessment of existing risks and conservation status of species and populations. Unfortunately, the IUCN Red List, which applies the most common criteria to assess threats to wild taxa, currently covers a low percentage of crop wild relatives (CWR). Only three Oryza species are listed (O. neocaledonica, O. officinalis, O. rufipogon), where O. neocaledonica is considered endangered. Moreover there is a need to develop new criteria suitable to estimate genetic diversity loss within a species [25**].
Figure 2

(a) Introgression for spikelet number/panicle in BILs derived from *O. longistaminata* acc. IRGC104301 (left) and *O. glumaepatula* acc. IRGC104387 (right) in the background of *indica* cv. PR114 (middle). (b) Introgression for grain weight in BILs derived from *O. rufipogon* acc. IRGC104301 (left) and *O. glumaepatula* acc. IRGC104387 (right) in the background of *indica* cv. PR114 (middle).
Genome studies and molecular genetics

Once such information is available, it will be possible to implement a systematic conservation plan for the wild relatives of rice. Ex situ conservation is already well developed, with extensive seed collections in GenBanks, the two largest being the International Rice Research Institute in the Philippines (4370 wild species and hybrids accessions at IRRI, http://irri.org), and Oryzabase in Japan (1703, http://www.shigen.nig.ac.jp). The downside of ex situ conservation is genetic erosion resulting from the lack of exposure to environmental variability and selection pressures, as well as genetic drift associated with germplasm regeneration [34*, 35]. Comparison of ex situ and in situ conservation efficiency of O. rufipogon populations in Dongxiang, China, showed that ex situ failed to maintain genetic diversity, thereby reducing allelic polymorphism by 34%, and genetic heterozygosity by 16% in 13 years [34*]. On the contrary, in situ conservation of germplasm, along with its habitat, is a dynamic form of protection. It allows populations to adapt to environmental stress, leading to the creation of genetic novelties useful for future research and breeding activities. It is necessary to complement ex situ with in situ conservation to protect genetic diversity effectively, as well as ecosystem’s biodiversity. However full practical implementation of in situ for the wild relatives of rice remains limited [36]. On the nine populations of O. rufipogon in Dongxiang studied in 1978, only three remained in 1995 [34*]. A brick fence was constructed to protect two of these populations threatened by human activity. The challenges to implement and manage new protected sites on the long term remain the cost and the need for collaboration between national governments with international agriculture research programs. A short-term solution would be to improve the management of populations of the wild relative of rice found within existing protected areas.

On the positive side, the value of CWRs and the importance of their conservation have been of increased interest and are now considered a high priority for numerous national and international initiatives, like the FAO of the United Nation’s project to establish a network for the in situ conservation of CWRs [25**]. As nations are inter-dependent with regards to the global issues of food security and the loss of biodiversity, these problems should be addressed by all parties interested in agriculture and conservation.

Opinion

As stated in the introduction, the improvement of rice in terms of increased yield, reduced environmental impact, and enhanced nutrition are important goals and key factors in helping to solve the 9 billion people question by 2050. In reality, we only have about 25 years to solve this pressing issue if we are to supply breeders with designed germplasm that will need to be adapted to different and changing growing conditions across the planet.

The ultimate goal of the International Oryza Map Alignment Project is to create a basic and translational research platform that can provide immediate access to virtually any region of the collective Oryza genome both genomically (i.e. access to high quality reference genome sequences linked to population resequencing data, and clone resources), and functionally (i.e. access to phenotyped advanced interspecific mapping populations and in situ conserved natural populations). Such a platform would facilitate the rapid identification of genes, molecular markers and germplasm from an evolutionary perspective that could be used to efficiently transfer a number of potentially important traits into cultivated rice.

A strong, collaborative and active synthesis of the three I-OMAP focus areas will be required if we are to accelerate the understanding and incorporation of useful wild alleles into cultivated rice and jump over the domestication bottleneck hurdle that has slowed conventional rice breeding for years. Ideally, advanced mapping populations should be developed and genotyped for a minimum of 10 diverse accessions of each Oryza species crossed with several elite varieties derived from the major and emerging rice growing regions around the world. Priority should be given to wild accessions possessing known abiotic and biotic stress resistances/tolerances (e.g. salt tolerance). Once created the AMPs should be extensively phenotyped under both field and greenhouse conditions in facilities like those designed at Huazhong Agricultural University (e.g. rice high-throughput automatic phenotyping center). Once a specific phenotype is linked to a set of markers the genomic region of the wild Oryza species can be pinpointed, extensively characterized, and targeted for introgression into elite cultivars and molecular cloning.

Given our accelerated time frame to solve the 9BPQ, the process described above should be transparent and fully integrated with the above Rice 2020 and Green Super Rice initiatives overall.

Conclusion

OMAP, OGE & I-OMAP consortia have generated a vast array of Oryza genomic tools and data that can now be used to help solve the 9BPQ. It is anticipated that a full array of 16 reference quality Oryza genome sequences will be available by the summer of 2013 at the latest. Such a data set will facilitate rapid gene discovery and provide the evolutionary insights needed to feed the future.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


This paper proposes a step-by-step guide to identify and estimate the need to conserve crop wild relatives and implement systematic and effective conservation. They give multiple examples of studies on CWR distributions, as well as a map of the locations of high priority for CWR ex situ and in situ conservation.


This paper shows the importance of in situ conservation in maintaining the genetic diversity of populations of O. rufipogon from Dongxiang county, China, by comparing the structure of ex situ and in situ populations. O. rufipogon accessions from Dongxiang are very closely related to the domesticated rice japonica, they are valuable germplasms that need to be protected. An in situ conservation plan was implemented to protect their habitat from human activity.


36. Lu B: The challenge of in situ conservation of crop wild relatives in the biotechnology era—a case study of wild rice
10 Genome studies and molecular genetics


47. Bocca R, Lorieux M, Seck PA, Futakuchi K, Manneh B, Baimey H, Ndjiondjop MN: Agro-morphological characterization of a population of introgression lines derived from crosses between IR64 (O. sativa ssp. indica) and TQG5681 (O. glaberrima) for drought tolerance. Plant Sci, in press.


