

The Future of Rice Genomics: Sequencing the Collective *Oryza* Genome

Jose Luis Goicoechea · Jetty Siva S. Ammiraju · Pradeep Reddy Marri ·
Mingsheng Chen · Scott Jackson · Yeisoo Yu · Steve Rounsley · Rod A. Wing

Received: 21 August 2010 / Accepted: 2 September 2010 / Published online: 24 September 2010
© Springer Science+Business Media, LLC 2010

Abstract The main objectives of the “*Oryza* Map Alignment Project” (OMAP) are to characterize the rice genome from a comparative standpoint by establishing a genus-wide and genome-scale comparative framework from representative species. Here, we report our progress in the analyses of these datasets and emerging “comparative phylogenomics” insights into *Oryza* evolution at two different resolutions—chromosomal and sequence levels. We demonstrate the abundance and impact of structural variations (SV) on genome diversity using African *Oryza* as a model. The molecular basis of SV was inferred using three genus-wide vertical sequence datasets. Combined, these data demonstrate that a single reference genome sequence for the genus *Oryza* is insufficient to comprehensively capture the genomic and allelic diversity present within the genus. Towards this end, we present a strategy to

generate high-quality and cost-effective *de novo* reference sequences of collective *Oryza*. The application and broader scientific impact of the OMAP resources under an international cooperative effort (I-OMAP) are discussed.

Keywords Rice · *Oryza* · OMAP · Comparative genomics

Introduction

Rice (*Oryza sativa* L.) feeds more people than any other crop. The agronomic importance of rice, its shared evolutionary history with major cereal crops, and its small genome size led to the generation of two draft sequences (Goff et al. 2002; Yu et al. 2002) and a high-quality “gold standard” genome or “reference” sequence (RefSeq) from the public International Rice Genome Sequencing Project (IRGSP 2005). The highly accurate RefSeq serves as a unifying research platform for a complete functional characterization of the rice genome. With the rice-dependent population expected to double in ~25 years, breeders are faced with the enormous task of doubling rice yields with less land and water, and with poorer soils. It is therefore critical that the scientific community unites to provide the tools and biological/evolutionary insights required to meet future needs. Rice scientists around the world are moving to address these critical issues and are formulating plans to systematically characterize the function of all rice genes by 2020 (Zhang et al. 2008). This characterization will take many approaches, including comparative genomics between the cereals and within *Oryza*, and utilization of the wild *Oryza* gene pool. One significant goal is what Zhang describes as a “Super Green Rice” variety, which would have a double yield and a reduced need for nitrogen, phosphorus, biocides, and water (Zhang 2007).

Jose Luis Goicoechea and Jetty Siva S. Ammiraju contributed equally to the work.

J. L. Goicoechea · J. S. S. Ammiraju · Y. Yu · R. A. Wing (✉)
Arizona Genomics Institute, School of Plant Sciences,
BIO5 Institute, University of Arizona,
Tucson, AZ 85721, USA
e-mail: rwing@ag.arizona.edu

P. R. Marri · S. Rounsley
School of Plant Sciences, BIO5 Institute, University of Arizona,
Tucson, AZ 85721, USA

M. Chen
State Key Laboratory of Plant Genomics, Institute of Genetics
and Developmental Biology, Chinese Academy of Sciences,
Beijing, China

S. Jackson
Department of Agronomy, Purdue University,
West Lafayette, IN 47907-2054, USA

The genus *Oryza*

The genus *Oryza* is composed of two cultivated (*O. sativa* and *Oryza glaberrima*) and 22 wild species (Bao and Jackson 2007). Cultivated rice is classified as an AA genome diploid and has six wild AA genome relatives. The remaining 16 wild species (including both diploid and tetraploid species) are classified into nine other genome types. Figures 1 and 2 show a phylogenetic tree of the genus *Oryza* (inferred from Ge et al. 1999; Lu et al. 2009 and Ammiraju et al. 2010) and a photograph of the diploid *Oryza* species at the same developmental stage, respectively. The wild *Oryza* species offer a largely untapped resource of genes that have the potential to solve many of the world's rice production issues, including yield, drought and salt tolerance, and disease and insect resistance (Brar and Khush 1997, 2003).

The *Oryza* map alignment project

In 2003 and 2006, our consortium was funded by the USA National Science Foundation (NSF) to create a set of genus-level comparative genomic resources spanning the *Oryza* phylogeny that could be used as an experimental platform to address important hypothesis-driven questions in both basic and applied research. The result of this *Oryza* Map Alignment Project (OMAP) was a set of bacterial artificial chromosome (BAC)-based physical maps (i.e., BAC libraries SNaPshot fingerprinted/BAC end sequenced (BES)/(Finger Printed Contigs (FPC) assembled) from 13 *Oryza* species (four AA genome types, and a single representative species of the nine other genome types [i.e., BB, CC, BBCC, CCDD, EE, FF, HHJJ, KKLL, and GG]) aligned to the rice

reference genome (Rice Chromosome 3 Sequencing Consortium 2005; Ammiraju et al. 2006; Kim et al. 2007, 2008; Wing et al. 2007; Goicoechea 2009). In addition, we generated five chromosome 3 short arm (Chr3S) RefSeqs with Sanger sequencing (i.e., genome types—AA, BB, CC, and BBCC), and three more using second generation sequencing methods (two AA and one FF). These genomic resources provide immediate access to virtually any region of the collective *Oryza* genome for interrogation, and are unparalleled in depth and taxonomic breadth in the higher eukaryote genomics community. They are only rivaled by the *Drosophila* 12 genomes system (*Drosophila* Sequencing Consortium 2007).

Key OMAP findings

Extensive structural variation

Although the role of structural variation (SV) in eukaryote genome evolution and speciation is now well recognized (Rieseberg 2001; Navarro and Barton 2003; Thomas et al. 2003; Livingstone and Rieseberg 2004; Coghlan et al. 2005; Feuk et al. 2006; Conrad and Hurler 2007; Noor et al. 2007; Rieseberg and Willis 2007; Widmer et al. 2009), the prevalence and degree to which SV affects plant genomes remain unclear. Therefore, analyses of a well-defined set of SV among closely related *Oryza* genomes is expected to bring new insights into evolutionary and mechanistic bases of these aspects (Wing et al. 2007).

Previous analysis of the comparative physical maps in the *Oryza* revealed a surprisingly high level of SV embedded within well preserved syntentic relationships (Kim et al. 2007, 2008; Goicoechea 2009). Here, we conducted a

Fig. 1 Phylogenetic tree of *Oryza* inferred from Ge et al. (1999), Lu et al. 2009 and Ammiraju et al. (2010). Genome types are indicated next to each phylogenetic branch. Seventeen representative *Oryza* species that are currently subjected to genomic investigation as part of the OMAP are indicated next to respective phylogenetic branches. *Dashed lines* indicate putative diploid parents involved in the formation of polyploids. *Filled in circles* indicate maternal parent, and *unfilled circles* indicate unidentified diploid genome.

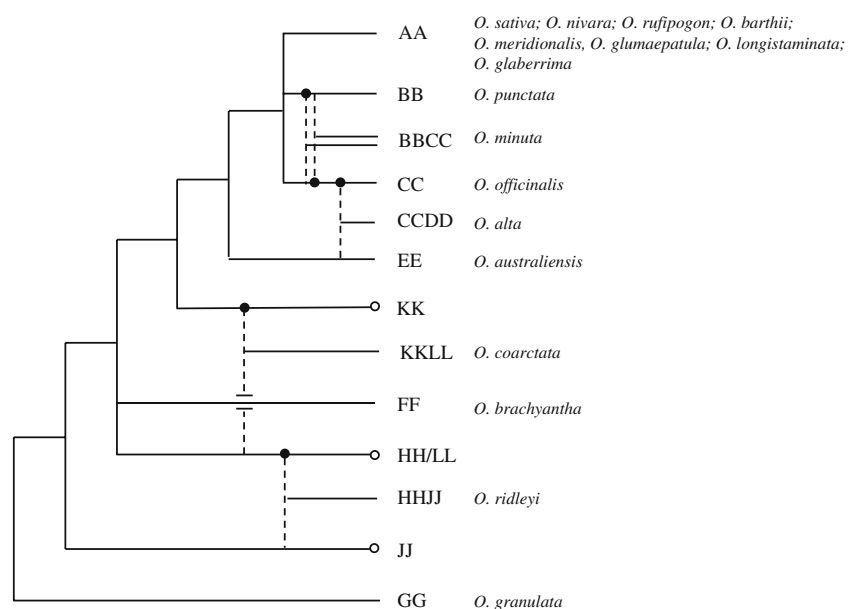
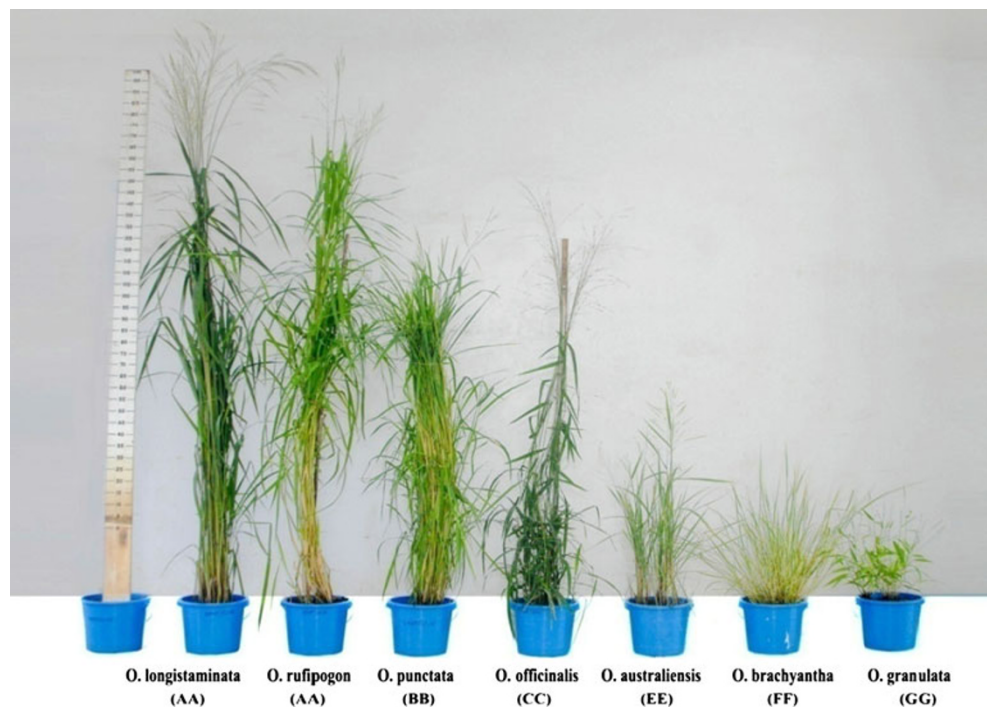


Fig. 2 Representative species of *Oryza* for each known diploid genome type.



comparative study of SV of four diploid African species with Asian cultivated rice *O. sativa* ssp. *japonica*. The compared species belong to three different genome types, namely AA (*O. glaberrima* and *Oryza barthii*), BB (*Oryza punctata*), and FF (*Oryza brachyantha*) and span major time points on the *Oryza* phylogeny (Figs. 1 and 3).

Using genome-wide computational scans, based on mapping “mate-paired” BES from each *Oryza* species to the Refseq, all putative major sites of genome alteration along the phylogenetic path of *Oryza* were recovered. Briefly, mate-pair BES were placed into two categories based on the mapping data; concordant pairs—when the “mapping distance” between the mate pairs reflected the expected insert size and orientation; discordant pairs—when one or both of these are unexpected due to a SV(s). A variant is termed “insertion” when the mapped distance is smaller in the Refseq compared expected insert size from an *Oryza* species, and a “contraction” for an opposite pattern. Putative inversions were detected when mate pairs had unexpected orientation in mapping (Fig. 4). The robustness of the detected SV, and their long-range extension for specific variants (e.g., inversions and trans-chromosomal events) was assessed using the order and orientation of synteny blocks from the heavily curated physical maps of the African species (Fig. 3).

Consequently, a compendium of putative intra and inter-chromosomal SV was annotated for each *Oryza* species that allowed further comparative exploration of their distribution, extent (frequency and quantity), and association with various genomic architectural features. Almost all SVs had genome-wide distribution with no

apparent bias for differential chromosomal distribution. A large portion of detected SVs were polymorphic between species and vast majorities of this unique portion were found to be due to expansions and contractions (Fig. 5). When these events are graphically plotted into 1-Mb bins, a continuous flux in sequence was observed along the entire length of the chromosomes of the *Oryza* species relative to the Refseq (Fig. 4). A significant finding was that in the two smallest genomes of *Oryza*, *O. glaberrima* and *O. brachyantha*, contractions outnumbered expansions. In the other two species, *O. barthii* and *O. punctata*, the detected expansions and contractions were nearly similar in number (Table 1), raising an intriguing possibility that the African species are under some sort of selection regime for genome size, i.e., either recalcitrant for genome expansion, or even some are on the path to genome downsizing. Another surprising finding was the high frequency of segmental reversals and trans-chromosomal events (translocations and duplicative segmental transpositions) (Fig. 6). Using known species divergence times (Ammiraju et al. 2008), and the presence or absence of a rearrangement at an orthologous location, the approximate time of origin and species/genome type specificity was determined. An association between SV and unstable genomic regions (repetitive/duplicative) was also uncovered. For example, all inter-species inversion polymorphisms were associated with either repetitive heterochromatic peri-centromeric (Table 2) regions, or enriched in segmental duplicated regions or regions rich in duplicated genes (see next section), suggesting their role in creating chromosomal instability via homologous and

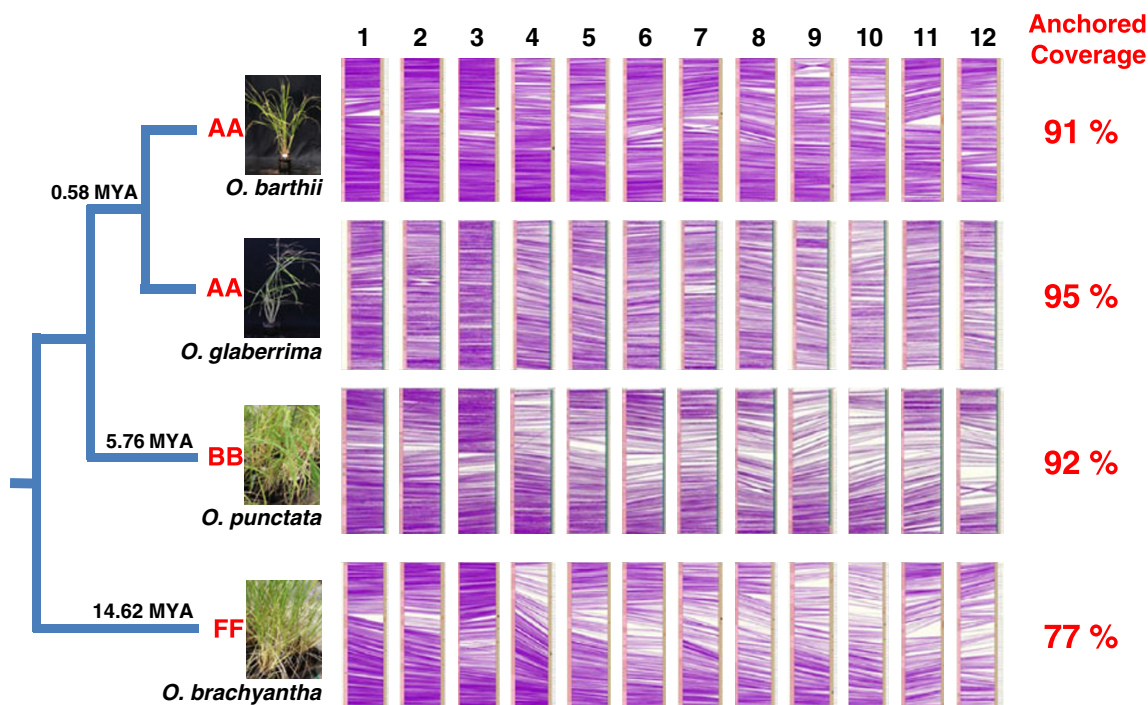


Fig. 3 Comparative physical maps of four African species of *Oryza*. On each chromosome, *left* and *right* blocks depict the physical contigs and RefSeq pseudomolecules respectively. *Purple connectors* show BES hit on the incumbent RefSeq pseudomolecule. *O. punctata* data from Kim et al. (2007).

non-homologous recombination. However, these fragile sites have been used differentially in a species-specific manner, indicating that these rearrangements have occurred recurrently at multiple occasions during *Oryza* evolution.

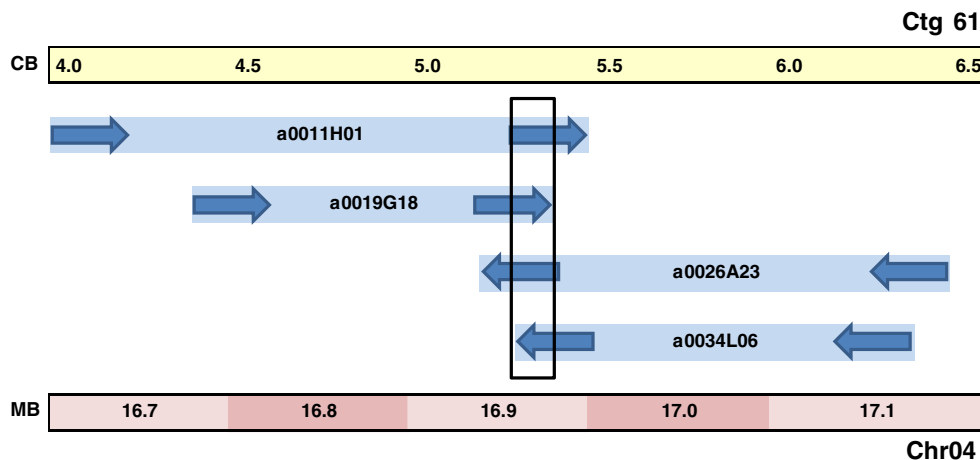
Our results indicate that SV is rampant, although clearly not random, and has played a major role in *Oryza* diversification. Because of the technical limitations in the discovery strategy, i.e., BAC size or contig size, the data presented here account only for a portion of largely uncharacterized *Oryza* SV. Major efforts are already underway to confirm and understand the biological consequences of SV in African *Oryza* (www.omap.org),

especially their possible relationship to agronomically important phenotypes, summarized by Goicoechea (2009).

Molecular nature of structural variation

As an initial micro-level foray into the molecular nature of genome flux, we conducted genus-wide, large-scale sequence analyses of three different biologically important and orthologous chromosomal regions from representative species spanning all six diploid and tetraploid genome types. These include the *Adh1–Adh2* region from Chr11 (Ammiraju et al. 2008; Bertoni 2008), the *Monoculm-1*

Fig. 4 Example of an inversion detected via mate-pair BES alignment on *O. barthii* chromosome 4. CB (consensus bands) and MB (megabase) units indicate the precise orthologous location of the rearrangement in the *O. barthii* physical map and *O. sativa* Refseq. Black blok indicates the breaking point.



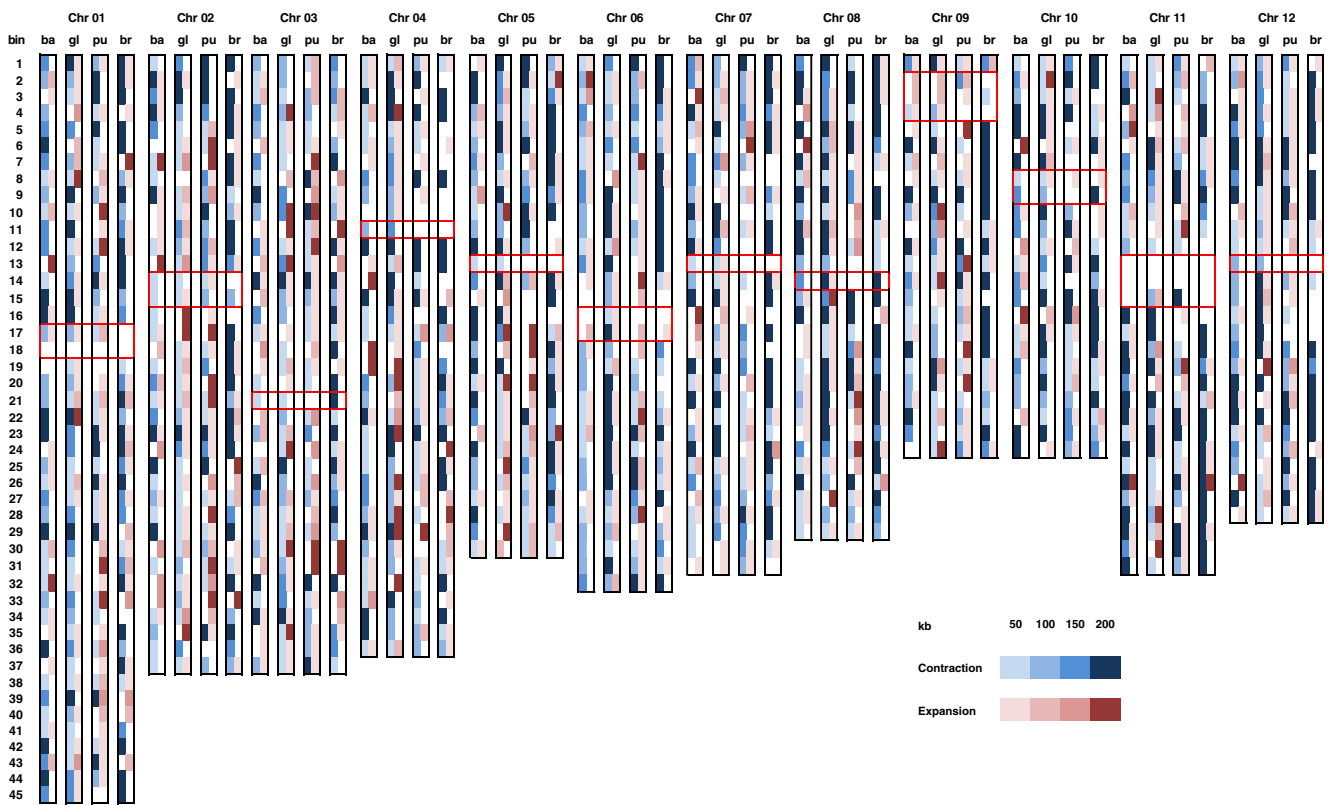


Fig. 5 Distribution of contractions and expansions by chromosome in four African species of *Oryza* mapped in 1-Mb bins. On each chromosome, contractions and expansions are mapped on each bin at left and right tracts, respectively, and their representation is not

scaled. Centromeric regions are indicated in red boxes on each chromosome. Species: ba - *O. barthii*, gl - *O. glaberrima*, pu - *O. punctata*, br - *O. brachyantha*.

(*MOC1*) region from Chr6 (Lu et al. 2009), and the *Heading date 1* (*Hd1*) region from Chr6 (Sanyal et al. 2010). Together, these investigations represent the most comprehensive multispecies comparative phylogenomics

analyses among plants to date that covered a broad ecological history over a short evolutionary gradient. These analyses uncovered several aspects of *Oryza* genome evolution that occurred since its origin approxi-

Table 1 Summary of major contractions and expansions observed on the aligned physical maps of four African species of *Oryza* to the RefSeq (*O. sativa*)

| Chromosome | <i>O. barthii</i> | | <i>O. glaberrima</i> | | <i>O. punctata</i> | | <i>O. brachyantha</i> | |
|------------|-------------------|------------|----------------------|------------|--------------------|------------|-----------------------|------------|
| | Contractions | Expansions | Contractions | Expansions | Contractions | Expansions | Contractions | Expansions |
| Chr01 | 3 | 3 | 5 | 3 | 3 | 3 | 3 | 0 |
| Chr02 | 4 | 2 | 4 | 3 | 5 | 1 | 9 | 0 |
| Chr03 | 3 | 3 | 4 | 4 | 1 | 5 | 2 | 2 |
| Chr04 | 5 | 8 | 2 | 1 | 2 | 3 | 9 | 0 |
| Chr05 | 2 | 0 | 8 | 0 | 1 | 1 | 8 | 0 |
| Chr06 | 2 | 2 | 2 | 1 | 3 | 2 | 4 | 0 |
| Chr07 | 1 | 1 | 8 | 2 | 1 | 6 | 4 | 0 |
| Chr08 | 2 | 1 | 5 | 1 | 2 | 2 | 11 | 0 |
| Chr09 | 3 | 3 | 2 | 1 | 3 | 1 | 4 | 0 |
| Chr10 | 4 | 3 | 5 | 2 | 5 | 2 | 8 | 0 |
| Chr11 | 5 | 4 | 4 | 2 | 5 | 5 | 4 | 1 |
| Chr12 | 4 | 2 | 4 | 1 | 3 | 5 | 5 | 0 |
| Total | 38 | 32 | 53 | 21 | 34 | 36 | 71 | 3 |

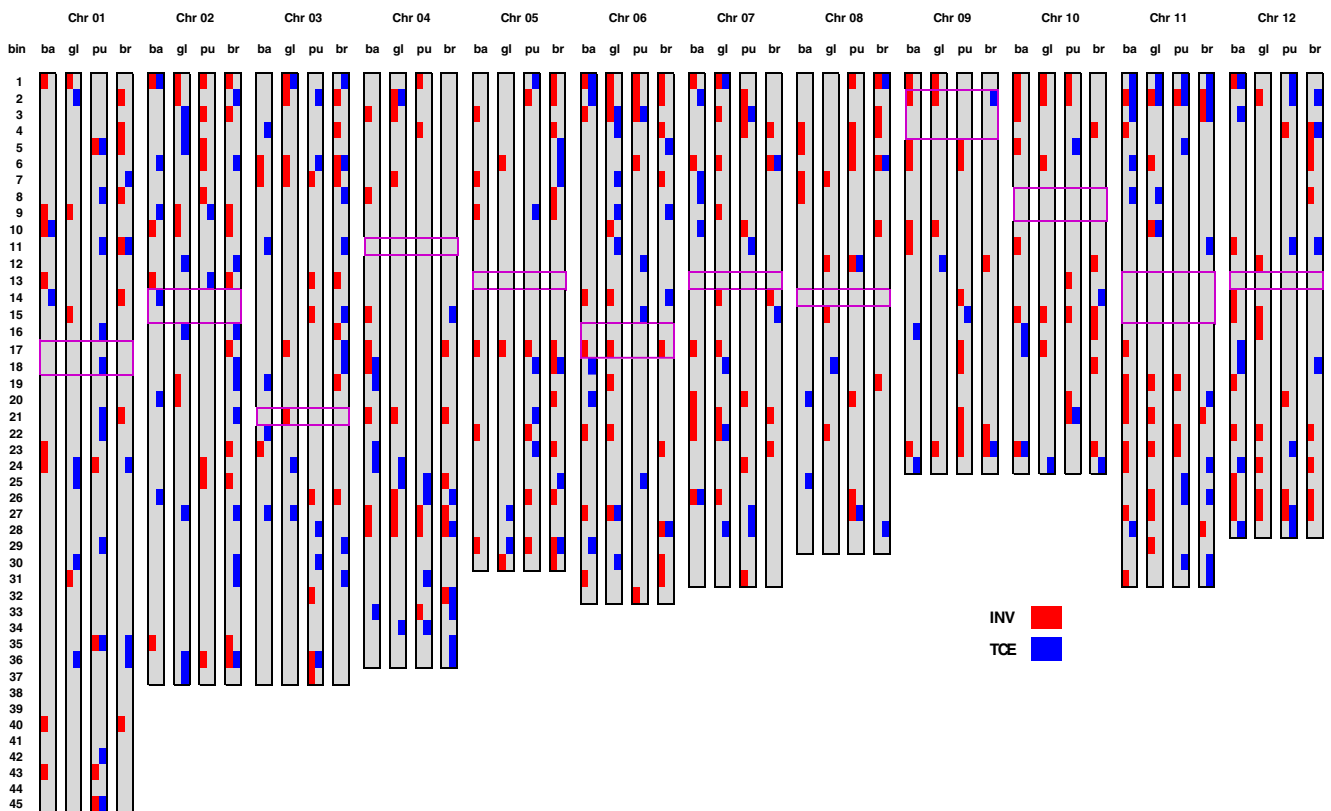


Fig. 6 Distribution of putative inversions and trans-chromosomal events by each chromosome in four African species of *Oryza* mapped in 1-Mb bins. *INV* inversions, *TCE* trans-chromosomal events.

mately 15 million years ago (MYA), providing the first insights into the nature of DNA rearrangements, their pace, chronology, mechanistic causes, and impact on genome diversity.

Table 2 Summary of major putative inversions observed in the physical maps of four African species of *Oryza* compared with the RefSeq (*O. sativa*)

| Chromosome | Start | End | Size (Mb) | ba | gl | pu | br |
|------------|-------|------|-----------|----|----|----|----|
| Chr01 | 27.9 | 28.2 | 0.3 | + | + | + | - |
| | 32.8 | 33.3 | 0.5 | + | + | - | - |
| Chr02 | 15.2 | 16.3 | 1.1 | - | + | - | - |
| Chr04 | 11.3 | 11.9 | 0.6 | - | - | - | + |
| | 13.9 | 15.4 | 1.5 | + | + | + | - |
| Chr05 | 10.8 | 11.8 | 1.0 | - | - | - | + |
| | 21.8 | 22.3 | 0.5 | + | + | + | - |
| Chr07 | 13.7 | 14.2 | 0.5 | - | - | + | - |
| | 14.1 | 15.7 | 1.6 | + | - | - | - |
| Chr08 | 7.1 | 7.5 | 0.4 | + | - | - | - |
| | 19.6 | 20.7 | 1.1 | - | - | + | - |
| Chr09 | 1.1 | 2.1 | 1.0 | + | - | - | - |
| Chr10 | 3.6 | 4.0 | 0.4 | - | + | - | - |

(+) presence, (-) absence, *ba* - *O. barthii*, *gl* - *O. glaberrima*, *pu* - *O. punctata*, *br* - *O. brachyantha*

Centromeric regions are indicated in red boxes on each chromosome. Species: *ba* - *O. barthii*, *gl* - *O. glaberrima*, *pu* - *O. punctata*, *br* - *O. brachyantha*.

A key outcome is that the *Oryza* genomes, despite a well-conserved gene structure, content, order, and orientation, have undergone rapid and lineage specific changes. A large number of them are recent, at least some are frequent, and show regional biases leading to varying degrees of genomic structural instability.

Here, we briefly highlight a few major forces and mechanisms contributing to genomic instability. The studied euchromatic genomic regions differ in their gene composition and recombinational properties; 78% of the genes from the *Adh1* region belong to eight different tandemly arrayed clusters, whereas the *MOC1* and *Hd1* regions are composed essentially of low copy genes. Compared with the *MOC1* and *Hd1* regions, the *Adh1* region showed the highest rate of structural evolution, and frequent exceptions to colinearity, thereby leading to an interesting correlation between regional genome stability, genic composition, recombination, and evolution of colinearity across a broad range of recently diverged lineages. A particular source of this instability is the presence of gene families, which exhibited remarkable plasticity in the copy number by continuous lineage specific birth, death, and divergence of individual members or the entire clusters at every phylogenetic node (Fig. 7).

Heterogeneous evolution of transposable elements (TEs) has played a major role in shaping the highly variable intergenic landscape and the genome sizes of *Oryza*, through independent expansions and contractions. In particular, these and related sequence datasets also revealed intriguing evolutionary dynamics of an ancient retrotransposon family “RWG”. Intense proliferation waves within the last 2–5 MYA of this single family alone is responsible for a dramatic increase of genome size in two diploid species, *Oryza granulata* [GG] and *Oryza australiensis* [EE] (Piegu et al. 2006; Ammiraju et al. 2007). Strikingly, the two proliferation bursts were independent and occurred after speciation, suggesting a non-linear evolution of genome size in *Oryza*. In addition, we showed that one TE family jumped multiple times “horizontally” in geographically and reproductively isolated species, thus constituting an important evolutionary force for *Oryza* genome diversification (Roulin et al. 2008). Single gene loss

(excluding the members of tandem clusters), gene transpositions, and *de novo* creation of new genes are common on all phylogenetic branches of *Oryza* and acts as a source for inter-species presence or absence polymorphisms. Specific examples include (a) the deletion of a very important gene *Hdl* in *O. glaberrima* (Sanyal et al. 2010), (b) a Pack-MULE mediated movement of full length and apparently functional LRR Kinase in AA genomes (Ammiraju et al. 2008), and (c) a *de novo* origin of and DNA/RNA-based recombination initiated new genes (Lu et al. 2009).

These analyses also revealed that the unequal homologous and non-homologous recombination rates can vary by more than two-fold within a 15-MYA time frame and play a major role in the creation of abundant and species-specific structural variation—e.g., a 300- and 75-kb inversions in *O. australiensis* [EE] and *O. brachyantha* [FF], respectively, and 200-kb deletion in *O. granulata* at the *Adh1* region [GG] (Ammiraju et al. 2008). Considering the

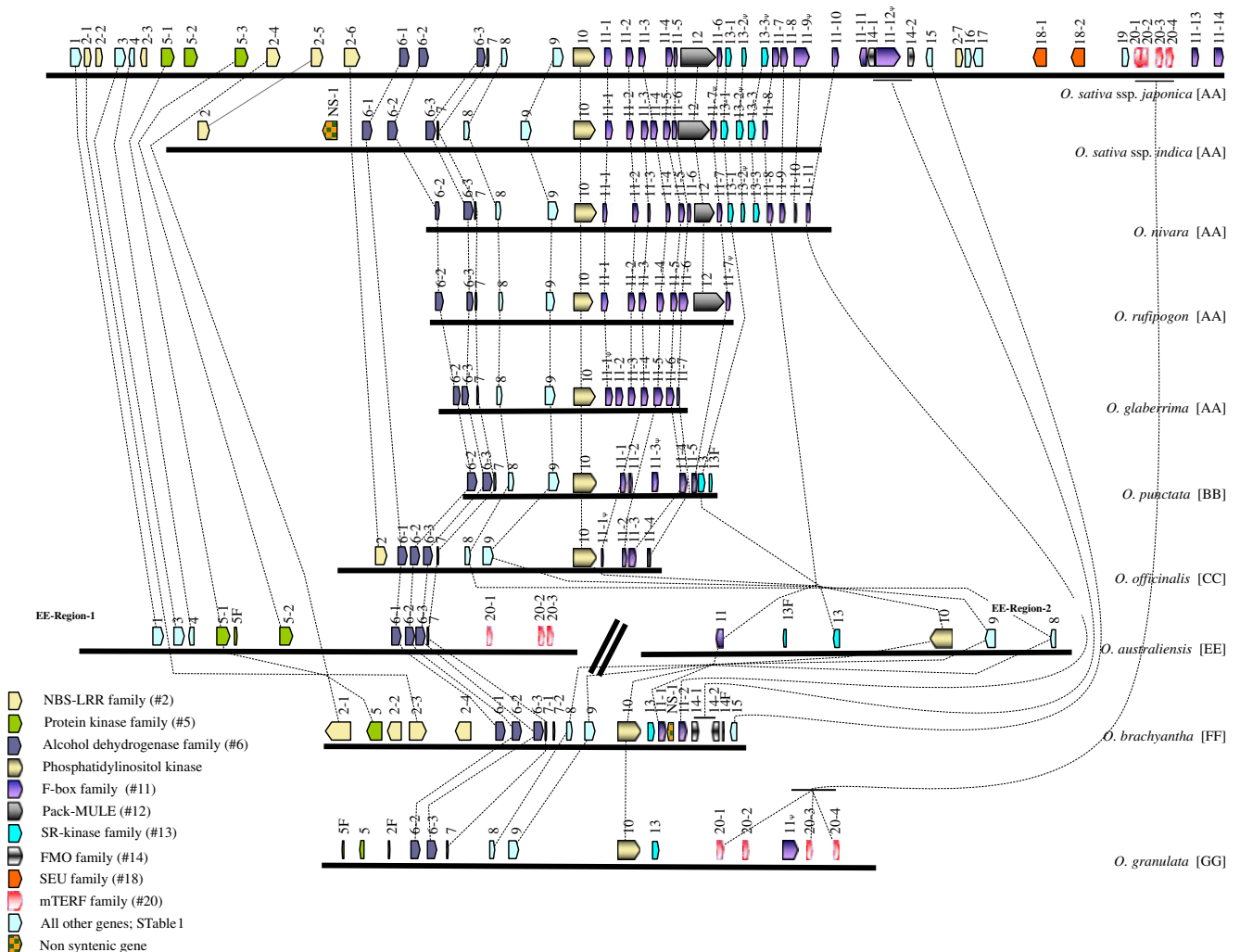


Fig. 7 Comparative phylogenomic view of the orthologous *Adh1* vertical dataset across the genus *Oryza*. Adopted from Ammiraju et al. (2008). Each gene family is color-coded to depict dynamics of gene family evolution.

rapid and dynamic changes observed in the diploid *Oryza* lineages, the natural and young (<2 MYA) *Oryza* polyploid species exhibit remarkable stability in the genome and the genome microstructure (Lu et al. 2009; Ammiraju et al. 2010).

RefSeq or ReSeq? Method development for de novo sequencing plant genomes using next-generation platforms and BAC pooling

Our comparative analysis of AA and BB genome species indicates that despite extensive colinearity their genomes can differ by 10% to 25%. With this in mind, we developed a method to cost-effectively generate high-quality *de novo* reference sequences of large plant genomes by combining “old school” physical maps with second generation sequencing technology. As a pilot experiment, we sequenced the 18 Mb Chr3S from *O. barthii* (the wild progenitor of *O. glaberrima*) using six pools of 28 BACs (each pool ~3 Mb) selected from the BAC minimum tiling path. Each pool was sequenced with 454 Titanium and GS FLX paired end chemistries, and assembled. The result was a chromosome arm sequence with a contigN50 of 14.3 kb, a scaffoldN50 of 3.16 Mb, and 90% of it is present in just six of the 44 scaffolds—the largest being over 6 Mb in length (Rounsley et al. 2009). N50 is a length-weighted average of contig or scaffold size, such that the average nucleotide in an assembly will appear in a contig (or scaffold) of N50 size or greater.

Long-range accuracy of the assembly was demonstrated by alignment against the *O. sativa* RefSeq where no rearrangements were observed. Nucleotide accuracy was assessed by comparing the overlaps between neighboring pools. Since these regions had been sequenced independently in each pool, any differences would reflect an error in one of the assembled pools. Only 95 bp of non-matching sequence was identified—an error rate of *only 2.2 bp per 10 kb*—very similar to the 1 in 10 kb “Bermuda accuracy” standard for finished sequence. The majority of errors were indels in homopolymer runs. To assess the utility of the resulting sequence for gene identification, we mapped known *O. sativa* genes onto the *O. barthii* Chr3S assembly. Of the 3,127 genes identified, 2,333 (75%) were completely covered, and 92% had greater than 90% coverage. This method was subsequently used to sequence the Chr3 arms of *Oryza nivara* [AA], *Oryza rufipogon* [AA] and *O. brachyantha* [FF], and the *O. glaberrima* genome with comparable results (unpublished).

Based on these data, our philosophy is to first generate a high-quality RefSeq using our NextGen/BAC Pool approach (or other cost-effective whole genome shotgun strategies) and then utilize re-sequencing methods (i.e., Illumina or SOLiD) to capture species-specific allelic variation.

International *Oryza* map alignment project

The broad utility and support of OMAP led to three 1-day International OMAP (I-OMAP) Grand Challenge meetings of basic researchers and breeders held in conjunction with the 4th, 5th, and 6th International Symposia for Rice Functional Genomics Meetings in Japan (October 2007), South Korea (November 2008), and the Philippines (November 2009). These meetings identified four top research priorities for the utilization of *Oryza* species to address fundamental questions in basic and applied research: (1) RefSeqs for all eight AA genome species and a representative species of the nine other genome types; (2) baseline transcriptome and small RNA data sets, for the 17 species from priority no. 1, as an aid to annotation, positional cloning, and new gene discovery; (3) backcross introgression lines and chromosome segment substitution lines of the AA genome species for functional and breeding studies; and (4) collections of naturally occurring populations of the wild *Oryza* species for diversity, conservation, population, and evolutionary analyses.

These I-OMAP priorities help to shape future research directions and also tie directly into the broader goals of the international community to functionally characterize all rice genes by 2020 and provide a functional toolkit for translational genomics in rice. With the rice-dependent population expected to double in 20–25 years, it is critical that this goal be achieved to help meet the food security needs of the future (Zhang et al. 2008).

Acknowledgement This work was supported by the USA National Science Foundation grant nos. 0638541, 0321678, and 0218794 to RAW and SJ.

References

- Ammiraju JS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, et al. The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*. *Genome Res.* 2006;16:140–7.
- Ammiraju JS, Zuccolo A, Yu Y, Song X, Piegue B, Chevalier F, et al. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. *Plant J.* 2007;52:342–51.
- Ammiraju JSS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, et al. Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell.* 2008;20:3191–209.
- Ammiraju JSS, Fan C, Yu Y, Song X, Cranston KA, Pontarolli AC, et al. Spatio-temporal patterns of genome evolution in allotetraploid species of the genus *Oryza*. *Plant J.* 2010;63:430–42.
- Bao Y, Jackson M. Wild rice taxonomy. 2007. <http://www.knowledgebank.irri.org/wildricetaxonomy/default.htm>.

- Bertoni G. Dynamic evolution of *Oryza* genomes. *Plant Cell*. 2008;20:3184.
- Brar DS, Khush GS. Alien introgression in rice. *Plant Mol Biol*. 1997;35:35–47.
- Brar DS, Khush GS. Utilization of wild species of genus *Oryza* in rice improvement. In: Nanda JS, Sharma SD, editors. Monograph on genus *Oryza*. Enfield: Science; 2003. p. 283–309.
- Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L. Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet*. 2005;21:673–82.
- Conrad DF, Hurler ME. The population genetics of structural variation. *Nat Genet*. 2007;39:S30–6.
- Drosophila Sequencing Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 2007;450:203–18.
- Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet*. 2006;7:85–97.
- Ge S, Sang T, Lu BR, Hong DY. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA*. 1999;96:14400–5.
- Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*. 2002;296:92–100.
- Goicoechea JL. Structural comparative genomics of four African species of *Oryza*. Ph.D. Dissertation. University of Arizona; 2009. p. 215
- International Rice Genome Sequencing Project (IRGSP). The map-based sequence of the rice genome. *Nature*. 2005;436:793–800.
- Kim H, San Miguel P, Nelson W, Collura K, Wissotski M, Walling JG, et al. Comparative physical mapping between *Oryza sativa* (AA genome type) and *O. punctata* (BB genome type). *Genetics*. 2007;176:379–90.
- Kim H, Hurwitz B, Yu Y, Collura K, Gill N, SanMiguel P, et al. Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*. *Genome Biol*. 2008;9:R45.
- Livingstone K, Rieseberg L. Chromosomal evolution and speciation: a recombination-based approach. *New Phytol*. 2004;161:107–12.
- Lu F, Ammiraju JSS, Sanyal A, Zhang SL, Song RT, Chen JF, et al. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci USA*. 2009;106:2071–6.
- Navarro A, Barton NH. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science*. 2003;300:321–4.
- Noor MAF, Garfield DA, Schaeffer SW, Machado CA. Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics*. 2007;177:1417–28.
- Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, et al. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res*. 2006;16:1262–9.
- Rice Chromosome 3 Sequencing Consortium. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. *Genome Res*. 2005;15:1284–91.
- Rieseberg LH. Chromosomal rearrangements and speciation. *Trends Ecol Evol*. 2001;16:351–8.
- Rieseberg LH, Willis JH. Plant speciation. *Science*. 2007;317:910–4.
- Roulin A, Piegu B, Wing RA, Panaud O. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE1 within the genus *Oryza*. *Plant J*. 2008;53:950–9.
- Rounsley S, Marri PR, Yu Y, He R, Sisneros N, Goicoechea JL, et al. *De novo* next generation sequencing of plant genomes. *Rice*. 2009;2:35–43.
- Sanyal A, Ammiraju JSS, Lu F, Yu Y, Rambo T, Currie J, et al. Extensive sequence co-linearity in a 155 kb region surrounding Hd1, A mayor domestication locus of rice, despite lability of the Hd1 Gene Itself. *Mol Biol Evol*. 2010. doi:10.1093/molbev/msq133.
- Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, et al. Comparative analyses of multi-species sequences from targeted genomic regions. *Nature*. 2003;424:788–93.
- Widmer A, Lexer C, Cozzolino S. Evolution of reproductive isolation in plants. *Heredity*. 2009;102:31–8.
- Wing R, Kim H, Goicoechea JL, Yu Y, Kudrna D, Zuccolo A, et al. The *Oryza* map alignment project (OMAP): A new resource for comparative genomics studies within *Oryza*. In: Brar DS, Mackill DJ, Hardy B, editors. Rice genetics V: Proc Fifth Intern Rice Genet Symp. Phillipines: IRRRI; 2007. p. 51–64.
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*. 2002;296:79–92.
- Zhang QF. Strategies for developing green super rice. *Proc Natl Acad Sci USA*. 2007;104:16402–9.
- Zhang Q, Li J, Xue Y, Han B, Deng XW. Rice 2020: a call for an international coordinated effort in rice functional genomics. *Molec Plant*. 2008;1:715–9.