

The International - *Oryza* Map Alignment Project: Development and Analysis of A Genus-wide Comparative Genomics Platform to Help Solve The 9 Billion People Question

Rod A. Wing¹, Mingsheng Chen², Lizhi Gao³, Bin Han⁴, Robert Henry⁵, Yue-ie Hsing⁶, Nori Kurata⁷, Antonio Costa de Oliveira⁸, Olivier Panaud⁹, Wen Wang¹⁰, Darshan Brar¹¹, Scott Jackson¹², Kshirod Jena¹³, Manyuan Long¹⁴, Carlos Machado¹⁵, Mike Sanderson¹⁶ and Doreen Ware¹⁷

¹Arizona Genomics Institute, University of Arizona, USA; ²Institute of Genetics & Developmental Biology, CAS, China; ³Kunming Institute of Botany, CAS, China; ⁴National Center for Gene Research, CAS, China; ⁵University of Queensland, Australia; ⁶Institute of Plant & Microbial Biology, AS, Taiwan; ⁷National Institute of Genetics (NIG), Japan; ⁸Plant Genomics & Breeding Center, Federal University of Pelotas, Brazil; ⁹Plant Genome & Development Laboratory, Perpignan University, France; ¹⁰Kunming Institute of Zoology, CAS, China; ¹¹Department of Plant Breeding & Genetics, International Rice Research Institute, Philippines; ¹²Crop & Soil Sciences Department, University of Georgia, USA; ¹³Department of Plant Breeding & Genetics, International Rice Research Institute, Philippines; ¹⁴Department of Ecology & Evolutionary Biology, University of Chicago, USA; ¹⁵Department of Biology, University of Maryland, USA; ¹⁶Department of Ecology & Evolutionary Biology, University of Arizona, USA; ¹⁷USDA-ARS, Cold Spring Harbor Laboratory, USA. ¹⁻¹⁰I-OMAP and ¹¹⁻¹⁷OMAP, OGE. Email: rwing@ag.arizona.edu

Abstract

The international Oryza Map Alignment Project (OMAP) is aimed to characterize the rice genome by establishing a genus-wide and genome-scale comparative framework from representative species. Our progress is in the analyses of these datasets and emerging comparative phylogenomics insights into Oryza evolution at chromosomal and sequence levels. Using African Oryza as a model, we demonstrated the abundance and impact of structural variations (SV) on genome diversity. The molecular basis of SV was inferred using three genus-wide vertical sequence datasets. Combined, these data demonstrated that a single reference genome sequence for the genus Oryza is insufficient to comprehensively capture the genomic and allelic diversity present within the genus. For achieving this, we outline 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

In 2011, the *Economist* magazine published an issue entitled “The 9-billion People Question” (9BPQ) which addressed a world-wide dilemma that is central to most if not all plant scientists – How can our society grow enough food to feed 2 billion additional human beings in less than 40 years (Parker 2011)? Rice (*Oryza sativa*) will play a key role in solving the 9BPQ as rice feeds half the world and it is that half that will double in size in the foreseeable future. Rice 2020 (Zhang et al 2008) is a key initiative aimed at solving the 9BPQ and

calls for the mobilization of our community to pool and coordinate all available resources with a common goal of creating a green super rice, where “green” means less input (e.g. water, fertilizer, pesticides, land), and “super” means 2-3 fold yields (Zhang 2007).

Comparative genomics

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*.

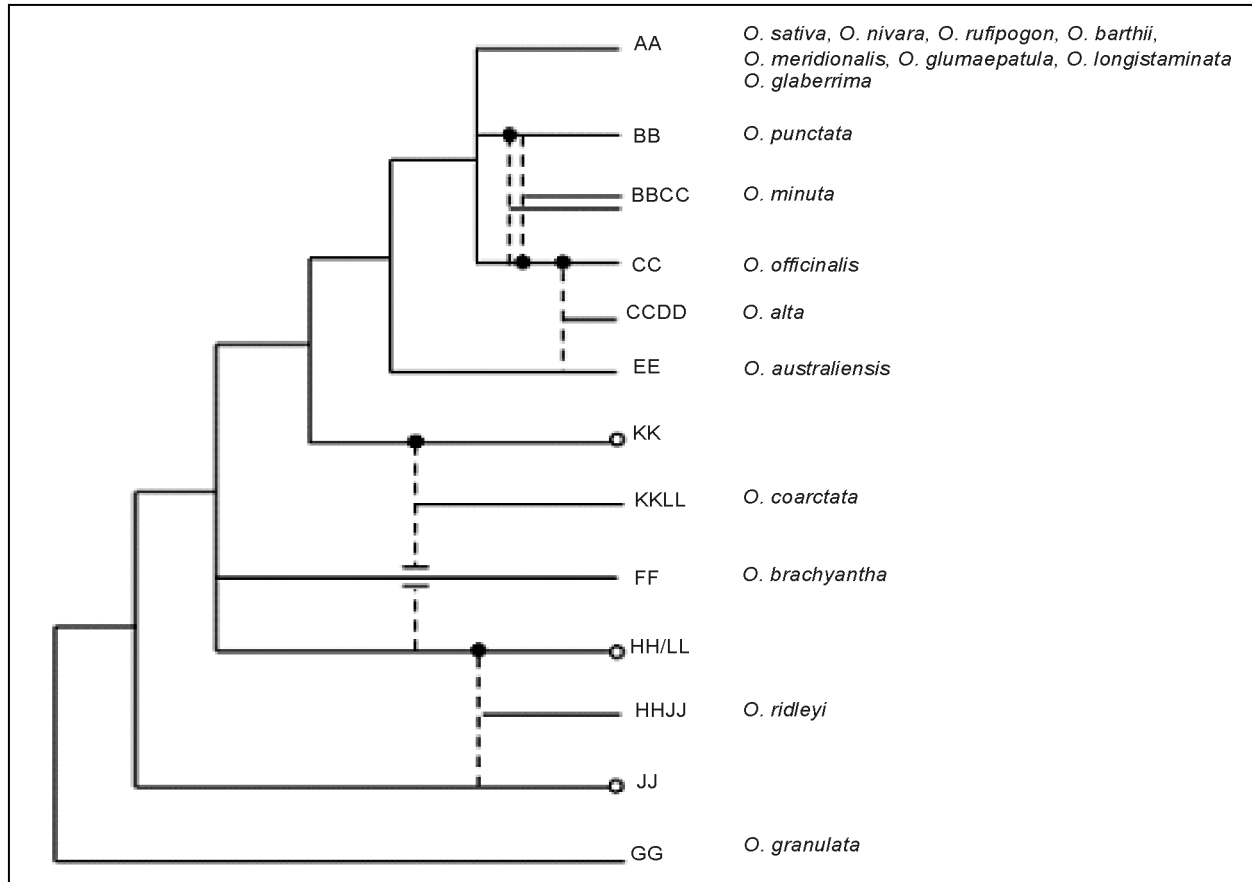


Figure 1. Phylogenetic tree of *Oryza*. Targeted OMAP species are indicated on each phylogenetic branch. Evolutionary relationships of the rice genomes were inferred from Ammiraju et al (2010), Ge et al (1999) and Lu et al (2009). Dashed lines indicate origin of allotetraploids; • indicate maternal parents. ○ indicate unidentified diploid species

The genus *Oryza* spans approximately 15 million years of evolutionary history (Fig. 1) (Ammiraju et al 2010) and is composed of 21 wild and two domesticated (*O. sativa* and *O. 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 we organized the *Oryza* Map Alignment (OMAP) and *Oryza* Genome Evolution (OGEP) Projects which have generated a large of array of publicly available genomic resources most notably a set of manually edited BAC-based physical maps (i.e. 18 deep-coverage BAC libraries - fingerprinted, end-sequenced, and FPC assembled) representing 18 of the 23 recognized *Oryza* species, covering all 8 AA genome species and one each of the other 9 genome types (BB, CC, BBCC, CCDD, EE, FF,

GG, KKLL, JJKK) (Ammiraju et al 2010, 2006, Kim et al 2008); and a set of chromosome 3 short arm sequences from all 8 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 www.Gramene.org and www.genome.arizona web sites, respectively.

Analysis of these data sets revealed the following key points:

- 1) LTR (long terminal repeat) Retro-transposable element amplifications dramatically increased the size of both the *O. australiensis* (EE) and *O. granulata* (GG) by as much as 400 and 200 Mb, respectively (Ammiraju et al 2007; Piegu et al 2006);
- 2) The AA genomes of *O. nivara*, *O. 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 (International Rice Genome Sequencing Project) RefSeq (Hurwitz et al 2010); and

3) Analysis of the *Adh1* region (~100-200 kb) across the entire *Oryza* phylogeny (diploid and polyploidy) showed significant perturbations of synteny including dynamic evolution of gene families, transposable element mediated gene movement, mutations, genome size changes, and large scale physical rearrangements (Ammiraju 2010, 2008).

The overriding conclusion from these studies, and many others, indicates that a SINGLE reference genome for the genus *Oryza* (i.e. IRGSP RefSeq) is not sufficient to capture and understand the allelic diversity/natural variation hidden with the genus to help solve the 9BPQ. To address this resource or knowledge gap we organized the International *Oryza* Map Alignment Project (I-OMAP) and have held five grand challenge meetings (Japan 07, Korea 08, Philippines 09, Brazil 10, Taipei 11) in conjunction with the annual International Symposiums on Rice Functional Genomics (ISRFG).

The three primary goals of I-OMAP are to:

- 1) Generate RefSeqs & Transcriptome data sets for all eight AA genome species, and a representative species of the nine other genome types;
- 2) Generate, map, and phenotype advanced ABC (advanced backcross), CSSL (chromosome substitution lines), RIL (recombinant inbred line) populations for the AA genome species for functional and breeding studies; and

3) Identify collections of naturally occurring populations of the wild *Oryza* species for diversity, conservation, population and evolutionary analyses.

We focus on goal number 1 and will report the current status of the *Oryza* Ref Seq project. Table 1 lists the status of each *Oryza* genome project as of June 2012. Sixteen of the 23 genome sequencing project are in progress or have been completed, and include all the diploid *Oryza* species. Draft sequences of two subspecies of *O. sativa* were published a decade (Goff 2002; Yu et al 2002) ago followed by the IRGSP “gold standard” RefSeq of *O. sativa* ssp. *japonica* (cv. Nipponbare) in 2005 (Nature 2005). Significant progress has been achieved over the past two years with completion of the *O. glaberrima* (AA), *O. barthii* (AA), *O. longistaminata* (AA) and *O. brachyantha* (FF) genomes (all unpublished but in Genbank). Assembly is currently in progress for the *O. nivara* (AA), *O. glumaepatula* (AA) and *O. punctata* (BB) genomes, and sequencing is underway for a majority of the remaining diploid species.

It should be noted that the I-OMAP project has a huge advantage over other next generation genome sequencing projects (e.g. *Drosophila* 12 genomes) (Nature 2007) in that physical maps are available for all AA genome species as well as representatives of all other 9 genome type.

Table 1. Sequencing status of 16 *Oryza* genomes and *Leersia perrieri* (*Oryza* outgroup species)

Species (genome type)	Genome Size	Lead Seq. Group	Status
1) <i>O. sativa</i> ssp. <i>indica</i> (AA)	~400 Mb	BGI	2002 (Draft)
2) <i>O. sativa</i> ssp. <i>japonica</i> (AA)	~400 Mb	IRGSP, Sasaki	2004 (RefSeq)
3) <i>O. glaberrima</i> (AA)	~354 Mb	AGI, Wing	2010 (RefSeq)
4) <i>O. barthii</i> (AA)	~411 Mb	AGI, Wing	2012 (RefSeq)
5) <i>O. brachyantha</i> (FF)	~260 Mb	CAS, Chen	2011 (RefSeq)
6) <i>O. longistaminata</i> (AA)	~352 Mb	CAS, Wang	2011 (Draft)
7) <i>O. nivara</i> (AA)	~448 Mb	AS, Hsing	Assembly IP
8) <i>O. rufipogon</i> (AA)	~445 Mb	NIG, Kurata	Assembly IP
9) <i>O. glumaepatula</i> (AA)	~464 Mb	Oliveira	Assembly IP
10) <i>O. punctata</i> (BB)	~423 Mb	AGI, Wing	Assembly IP
11) <i>O. meridionalis</i> (AA)	~435 Mb	UQ, Henry	Sequencing IP
12) <i>O. australiensis</i> (EE)	~960 Mb	UP, Panaud	Sequencing IP
13) <i>O. officinalis</i> (CC)	~653 Mb	NIG, Kurata	Sequencing IP
14) <i>O. eichingeri</i> (CC)	~650 Mb	NIG, Kurata	Sequencing IP
15) <i>O. rhizomatis</i> (CC)	~650 Mb	NIG, Kurata	Sequencing IP
16) <i>O. granulata</i> (GG)	~862 Mb	CAS, Gao	Sequencing IP
17) <i>L. perrieri</i> (outgroup)	~323 Mb	AGI, Wing	Sequencing IP

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”.

In addition to an I-OMAP status report, the analysis of the *O. glaberrima* genome. *O. glaberrima* is 2nd species of *Oryza* that was independently domesticated ~3,500 years ago in West Africa is discussed.

Conclusions

Our OMAP &I-OMAP consortia have generated a vast array of *Oryza* genomic tools and data that can now be used to help solve the 9BPQ. We anticipate a full array of 16 reference quality *Oryza* genome sequences 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

- Ammiraju JS, Fan C, Yu Y, Song X, Cranston KA, Pontaroli AC, Lu F, Sanyal A, Jiang N, Rambo T, Currie J, Collura K, Talag J, Bennetzen JL, Chen M, Jackson S and Wing RA.** 2010. Spatio-temporal patterns of genome evolution in homeologous *Adh1 - Adh2* regions from four allotetraploid species of the genus *Oryza*. *Plant J* 63:430-442.
- Ammiraju JSS, Lu F, Sanyal A, Yu Y, Song X, Jiang N, Pontaroli AC, Rambo T, Currie J, Collura K, Talag J, Fan C, Goicoechea JL, Zuccolo A, Chen J, Bennetzen JL, Chen M, Jackson S and Wing RA.** 2008. Dynamic evolution of *Oryza* genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. *Plant Cell* 20:3191-3209
- Ammiraju JSS, Luo M, Goicoechea JL, Wang W, Kudrna D, Mueller C, Talag J, Kim HR, Sisneros NB, Blackmon B, Fang E, Tomkins JB, Brar D, MacKill D, McCouch S, Kurata N, Lambert G, Galbraith DW, Arumuganathan K, Rao K, Walling JG, Gill N, Yu Y, SanMiguel P, Soderlund C, Jackson S, Wing RA.** 2006. 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* 16:140-147.
- Ammiraju JSS, Song X, Luo M, Sisneros N, Angelova A, Kudrna D, Kim HR, Yu Y, Goicoechea JL, Lorieux M, Kurata N, Brar D, Ware D, Jackson S, and Wing RA.** 2010. The *Oryza* BAC Resource: A genus-wide and genome scale tool for exploring rice genome evolution and leveraging useful genetic diversity from wild relatives. *Breed Sci* 60:536-543.
- Ammiraju JS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar D, SanMiguel P, Jiang N, Jackson S, Panaud O and Wing RA.** 2007. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. *Plant J* 52:342-351.
- Ge S, Sang T, Lu BR and Hong DY.** 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. *Proc Natl Acad Sci USA* 96: 14400–14405.
- Goff SA, Ricke D, Wang R, Presting G, Lan D, Dunn M, Glazebrook J, Varma H, Hadley D, Hutchison D, Sessions A, Martin C, Katagiri F, Oeller P, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Ho CB, Paszkowski U, Zhang S, Colbert M, Sun W, Chen L, Cooper B, Park S, Kimmerly W, Wood TC, Mao L, Quail P, Wing RA, Dean R, Yu Y, Briggs S, Zharkikh A, Shen R, Oliphant A, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T.** 2002. Draft sequence of the rice genome. *Science* 296:92-100.
- Hurwitz BL, Kudrna D, Yu Y, Sebastian A, Zuccolo A, Jackson SA, Ware D, Wing RA and Stein L.** 2010. Rice structural variation: a comparative analysis of structural variation between rice and three of its closest relatives in the genus *Oryza*. *Plant J* 63(6):990-1003.
- Kim, H., B. Hurwitz, Y. Yu, K. Collura, N. Gill, P. SanMiguel, J.C. Mullikin, C. Maher, W. Nelson, M. Wissotski, M. Braidotti, D. Kudrna, J.L. Goicoechea, L. Stein, D. Ware, S.A. Jackson, C. Soderlund, R.A. Wing.** 2008. Construction, alignment and analysis of twelve framework physical maps that represent the ten genome types of the genus *Oryza*. *Genome Biol* 9: R45.
- Lu F, Ammiraju JSS, Sanyal A, Zhang SL, Song RT, Chen JF, et al.** 2009. Comparative sequence

- analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proc Natl Acad Sci USA* 106:2071–2076.
- Nature.** 2005. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* 436(7052):793-800.
- Nature.** 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450(7167):203-218.
- Parker J.** 2011. The 9 billion- people question. In: *The Economist*. February 24, 2011.
- Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA et al.** 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res* 16(10):1262-1269.
- Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X et al.** 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296(5565):79-92.
- Zhang Q, Li J, Xue Y, Han B and Deng XW.** 2008. Rice 2020: A call for an international coordinated effort in rice functional genomics. *Mole Plant* 1(5):715-719.
- Zhang Q.** 2007. Strategies for developing green super rice. *Proc Natl Acad Sci USA* 104(42):16402-16409.
-
- Citation:** Wing RW, Chen M, Gao L, Han B, Henry R, Hsing Y, Kurata N, Oliveira, Olivier AC, Panaud O, Wang W, Brar D, Jackson S, Jena K, Long M, Machado C, Sanderson M and Ware D. 2013. The international - *Oryza* map alignment project: Development and analysis of a genus-wide comparative genomics platform to help solve the 9 billion people question. In: Muralidharan K and Siddiq EA, eds. 2013. *International Dialogue on Perception and Prospects of Designer Rice*. Society for Advancement of Rice Research, Directorate of Rice Research, Hyderabad 500030, India, pp 4-8.