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; 12Department of Ecology & Evolutionary Biology, University of Chicago, USA; 13Department of Biology, University of Maryland, USA; 14Department of Ecology & Evolutionary Biology, University of Arizona, USA; 15USDA-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.
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* Genomes Evolution (OGEP) Projects which have generated a large 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, JJJK) (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 Leerisa 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 Symposia 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)

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


