Plant Molecular Biology (2005) 59:53–62
DOI 10.1007/s11103-004-6237-x

The *Oryza* Map Alignment Project: the golden path to unlocking the genetic potential of wild rice species

Rod A. Wing[1,*], Jetty S.S. Ammiraju[1], Meizhong Luo[1], HyeRan Kim[1], Yeisoo Yu[1], Dave Kudrna[1], Jose L. Goicoechea[1], Wenming Wang[1], Will Nelson[2], Kiran Rao[2], Darshan Brar[6], Dave J. Mackill[6], Bin Han[7], Cari Soderlund[2], Lincoln Stein[5], Phillip SanMiguel[4] and Scott Jackson[3]

[1]*Arizona Genomics Institute, The University of Arizona, Tucson, AZ (*author for correspondence; e-mail rwing@ag.arizona.edu);* [2]*Arizona Genomics Computational Laboratory, The University of Arizona, Tucson, AZ;* [3]*Department of Agronomy, Purdue University, West Lafayette, IN;* [4]*Genomics Core Facility, Purdue University, West Lafayette, IN;* [5]*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York;* [6]*International Rice Research Institute, Los Baños, The Philippines;* [7]*National Center for Gene Research, Shanghai, China*

Received 21 September 2004; accepted in revised form 12 November 2004

Key words: BAC-end sequencing, comparative genomics, physical mapping, rice, wild species

Abstract

The wild species of the genus *Oryza* offer enormous potential to make a significant impact on agricultural productivity of the cultivated rice species *Oryza sativa* and *Oryza glaberrima*. To unlock the genetic potential of wild rice we have initiated a project entitled the 'Oryza Map Alignment Project' (OMAP) with the ultimate goal of constructing and aligning BAC/STC based physical maps of 11 wild and one cultivated rice species to the International Rice Genome Sequencing Project's finished reference genome – *O. sativa* ssp. *japonica* c. v. Nipponbare. The 11 wild rice species comprise nine different genome types and include six diploid genomes (AA, BB, CC, EE, FF and GG) and four tetrapliod genomes (BBCC, CCDD, HHKK and HHJJ) with broad geographical distribution and ecological adaptation. In this paper we describe our strategy to construct robust physical maps of all 12 rice species with an emphasis on the AA diploid *O. nivara* – thought to be the progenitor of modern cultivated rice.

Introduction

Rice (*Oryza sativa* L.) is the most important human food crop in the world. The agronomic importance of rice, its shared evolutionary history with major cereal crops, and small genome size have led to the generation of two draft genome sequences (Goff *et al.*, 2002; Yu *et al.*, 2002) and a high-quality finished genome sequence by the International Rice Genome Sequencing Project (Sasaki *et al.*, 2002; Yu *et al.*, 2002; The Rice Chromosome 10 Sequencing Consortia, 2003; IRGSP 2004 unpublished). The highly accurate and public IRGSP sequence now serves as a unifying research platform for a complete functional characterization of the rice genome. Such an analysis will investigate the rice transcriptome, proteome and metabolome, with the goal of understanding the biological function of all rice genes (40–50, 000) and applying that information to improve rice production and quality. This comprehensive analysis will utilize a variety of techniques and resources from expression and genome tiling arrays to collections of tagged mutant populations developed in elite cultivars grown around the world.

Comparative genomics between the cereal genomes and within the genus *Oryza* will also

play a critical role in our understanding of the rice genome (Ahn *et al*., 1993; Ahn and Tanksley, 1993; Huang and Kochert, 1994; Jena *et al*., 1994; Bennetzen and Ma, 2003; Han and Xue, 2003; Ma and Bennetzen, 2004; Salse *et al*., 2004). By comparing genome organization, genes and intergenic regions between cereal species, one can identify regions of the genome that are highly conserved or rapidly evolving. Such regions are expected to yield key insights into genome evolution, speciation and domestication. The study of conserved noncoding sequences (CNSs) between cereal genomes will also increase our ability to understand and isolate cis-regulatory elements required for precise developmental and temporal gene expression (Kaplinsky *et al*., 2002).

The genus *Oryza* is composed of two cultivated *(O. sativa* and *O. glaberrima)* and 21 wild species (Khush, 1997; Vaughan *et al*., 2003). Based on recent phylogenetic data, Ge *et al*. (1999) proposed that *Porteresia coarctata* should be included in the genus as the 24th *Oryza* species. Cultivated rice is classified as an AA genome diploid and has six wild AA genome relatives. The remaining 15 wild species are classified into nine other genome types that include both diploid and tetraploid species. Figure 1 shows a proposed phylogenetic tree of the genus *Oryza* as described by Ge *et al*. (1999) based on the analysis of two nuclear and one chloroplast gene. The wild rice species offer a largely untapped resource of agriculturally important genes that have the potential to solve many of the problems in rice production that we face today such as yield, drought and salt tolerance and disease and insect resistance.

To better understand the wild species of rice and take advantage of the IRGSP genome sequence, we have embarked on an ambitious comparative genomics program entitled the '*Oryza* Map Alignment Project' (OMAP). The long term objective of OMAP is to create a genome-level closed experimental system for the genus *Oryza* that can be used as a research platform to study evolution, development, genome organization, polyploidy, domestication, gene regulatory networks and crop improvement. The specific objectives of OMAP are to: (1) construct deep-coverage large-insert BAC libraries from 11 wild and one cultivated African *Oryza* species (*O. glaberrima*) *Oryza* species; (2) fingerprint and end-sequence the clones from all 12 BAC libraries; (3) construct physical maps for all 12 *Oryza* species and align them to the IRGSP genome sequence; and (4) perform a detailed reconstruction of rice chromosomes 1, 3 and 10 across all 12 *Oryza* species.

In this paper, we present our current progress for OMAP and some early glimpses into the results we are finding.

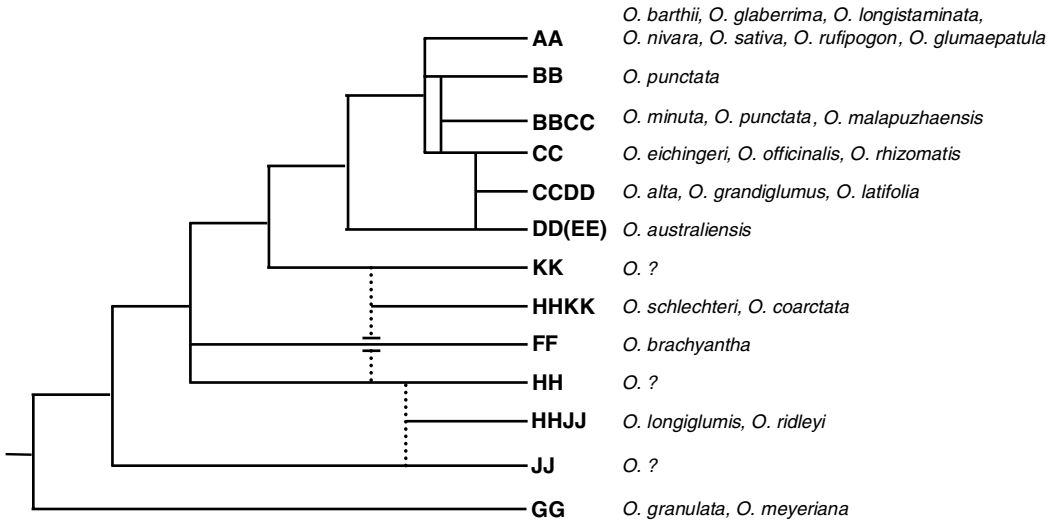

Figure 1. Phylogenetic tree of the genus *Oryza* (Modified from Ge et al. 1999).

Results

Development of the OMAP BAC library resource

Wild rice accessions were obtained from: (1) the International Rice Research Institute (IRRI) Los Baños, The Philippines; (2) the National Institute of Genetics, Mishima, Japan; and (3) Cornell University, Ithaca, New York (Table 1). Our major criteria for the selection of these wild rice accessions were that each one was robust and sufficient seed was available for distribution to the community, and that each contained potentially useful agronomic traits.

High molecular weight DNA was obtained from young seedlings for the AA genome species *O. nivara, O. rufipogon* and *O. glaberrima*. In contrast, because no inbred single-seed decent material was available for the remaining wild species, we prepared DNA from single plants that were clonally propagated at IRRI. Efforts are now underway to generate inbred seed from these wild species and seed should be available from the IRRI seed bank within 2–3 years.

Deep-coverage large-insert BAC libraries were developed for all 12 OMAP species using standard procedures developed in our laboratory over the past 10 years (Table 1) (Luo and Wing, 2003). All libraries were quality tested for insert size and depth of coverage and were found to represent at least 10 genome equivalents with average insert sizes ranging between 123 kb (*O. coarctata*) and 161 kb (*O. nivara*) (Jetty and Wing, unpublished). All OMAP BAC libraries were deposited in the Arizona Genomic Institute's BAC/EST Resource Center for public distribution (www.genome.arizona.edu).

Development of wild species FPC/STC physical maps

Once a wild species BAC library is constructed, clones representing about 10 genome equivalents are fingerprinted and end sequenced. The fingerprints are assembled into contigs based on shared bands between clones using the software FPC (Soderlund *et al*., 2000). The FPC contigs can then be aligned to the IRGSP genome sequence using the BAC end sequences.

Traditionally, BAC fingerprinting and end sequencing was very time consuming and expensive (Mao *et al*., 2000; Chen *et al*., 2002). We have now optimized our methods so that we can produce one physical map in about 1–3 months depending on genome size. Our OMAP objectives are to complete all physical mapping and end sequencing in a 2 year period and requires us to generate approximately 1, 000, 000 fingerprints and 2, 000, 000 BAC end sequences by September 2005.

To fingerprint a BAC library, we use a modification of the SNaPshot fingerprinting method described by Luo *et al*. (2003). Briefly, BAC DNA is isolated using a semiautomated 96-well alka-

Table 1. OMAP BAC library summary*.

Oryza species	Genome type	Accession number	Accession source	# Clones	Avg Insert Size (Kb)	Genome Size (Mb)
O. glaberrima	AA	96717	Africa	55296	140	809
O. nivara	AA	W0106	India	55296	161	760
O. rufipogon	AA	105491	Malaysis	64512	134	760
O. punctata	BB	105690	Africa	36864	142	539
O. officinalis	CC	100896	Thailand	92160	141	1201
O. minuta	BBCC	101141	Philippines	129024	125	1691
O. alta	CCDD	105143	S. America	92160	133	1000
O. australiensis	EE	100882	Australia	92160	153	1054
O. brachyantha	FF	101232	Africa	36864	131	343
O. granulata	GG	102118	Thailand	73728	134	907
O. ridleyi	HHJJ	100821	Thailand	129024	127	1568
O. coarctata	HHKK	104502	Bangladesh	147456	123	1568

* BAC libraries, hybridization filters can be ordered from the AGI BAC/EST resource center (www.genome.arizona.edu). Genome sizes were retrieved from Angiosperm DNA C-value database; release 3.1,September 2001 (Bennett and Leitch, 2001: www.rbgkew.org.uk/cval/homepage.html)

line lysis protocol (Kim HR and Wing RA, unpublished), and then digested with 5 restriction enzymes of which 4 generate 5′ overhangs. The corresponding 3′ OH ends are then extended using a single fluorescently labeled ddNTP and DNA polymerase. The reaction products are then separated on ABI3730XL capillary electrophoresis sequencers and the labeled fragments are band called using ABI fragment analysis software. The major modification to the procedure described by Luo *et al*. (2003) was that both reactions – restriction enzyme digestion and SNaPshot labeling – are performed in a single tube. Currently production rates are about 20, 736 fingerprints/week.

To streamline our BAC end sequencing method, we converted all our protocols to a 384-well format and reduced our ABI BIGDYE sequencing chemistry to 0.5 μl/reaction – a five fold reduction from previous methods. Currently we are producing 23, 040 STCs/week equaling 14.4 Mb of high quality sequence.

To date we have fingerprinted and end sequenced 3 OMAP BAC libraries: *O. brachyantha* (OB-Ba); *O. nivara* (OR_BBa); and *O. rufipogon* (OR_CBa) (Table 2). As can be seen, we are achieving a 78–91% success rate for fingerprinting. Table 2 also shows the number of contigs found for standard FPC builds. Preliminary data show that the number of contigs for the three projects yield between 428 and 806 without any manual editing.

Table 3 shows the results for BAC end sequencing for the same three libraries. The average success rates are between 91 and 97% meaning that the majority of BAC end sequences are paired. The average read length is between 638 and 686 bases resulting in over 162 Mb of new wild rice sequence deposited in Genbank over the last 2 months.

These physical maps are now being refined using a variety of assembly parameters followed by end merging and contig alignment to the IRGSP genome sequence. All OMAP FPC maps are available on the internet using webFPC (www.genome.arizona) and BAC end sequence alignments on Gramene (www.gramene.org) (Ware *et al*., 2002).

Reconstruction of the O. nivara chromosome 3

The *O. sativa* ssp. *japonica* cv Nipponbare chromosome 3 was recently finished by the US Rice Chromosome 3 Sequencing Consortia (2005, submitted). Cytologically, chromosome 3 is the 2nd largest rice chromosome, measuring 56.41 μm (or ~52.4 Mb) and is one of the most euchromatic (Cheng *et al*., 2001). Genetically, chromosome 3 is 170 cM in length (Harushima *et al*., 1998) and has 27 morphological mutants. In addition, over 133 agronomic genes/traits/QTLs (Oryzabase.org.) and 963 cDNAs (Wu *et al*., 2002) have been found associated with chromosome 3 (Oryzabase.org). The consortia sequenced ~36.1 Mb of chromosome 3 and identified 6237 new genes (US Rice Chromosome 3 Sequencing Consortia, unpublished).

Although *Oryza* separated from maize and sorghum ~50 million years ago (MYA) and from wheat and barley ~40 MYA their common evolutionary history can be traced by the co-linear order of genetic markers across their chromosomes (Moore *et al*., 1995). This is particularly true for the short arm of chromosome 3 which shows large

Table 2. OMAP SNaPshot/FPC fingerprinting summary.

Project	Genome	#clones	% Success	# Contigs	# Singletons
O. nivara	AA	51, 056	91	677	2356
O. rufipogon	AA	33, 023	91	806	1305
O. brachyantha	FF	25, 216	78	428	1805

Table 3. OMAP BAC-end sequencing summary*.

Project	Genome	# Reads	% Success	Avg HQ bp	Mb Sequenced
O. nivara	AA	110, 589	96	656 bp	~70 Mb
O. rufipogon	AA	73, 717	97	686 bp	~51 Mb
O. brachyantha	FF	70, 860	91	638 bp	~41 Mb

* All sequence data has been deposited in Genbank.

stretches of genetic marker co-linearity with maize chromosomes 1 and 9, sorghum linkage group L, and barley and wheat chromosomes 4L. Such conserved synteny across the cereals suggests that rice chromosome 3 will be a good model to study chromosome evolution.

Using our preliminary *O. nivara* FPC map and corresponding BAC ends we attempted to reconstruct the *O. nivara* rice chromosome 3. Paired BAC end sequences were mapped to the IRGSP chromosome 3 pseudomolecule using MegaBLAST with cutoff 1 e-200, and then associated with their corresponding FPC contig. The FPC contigs were then aligned to the *japonica* chromosome 3 which is graphically represented in Figure 2 using SyBr (Synteny Browser: Nelson and Soderlund, unpublished).

As can be seen, there is a tremendous amount of synteny between the cultivated and wild (*O. nivara*) chromosome as would be expected for the AA genome diploids. In this analysis, we were able to align 76 *O. nivara* contigs to the *japonica* chromosome 3. We did not detect any gross rearrangements between these species with this preliminary analysis. What can not be seen in Figure 2 is that the contig alignments tend to be somewhat ambiguous around the telomere and centromere gaps which are highly repetitive.

Insertions and deletions in genomes play a critical role in evolution. To obtain a glimpse of the level of insertions and deletions between the *O. sativa* and *O. nivara* genomes, we analyzed 1 Mb of finished rice chromosome 3 and identified *O. nivara* BAC end pairs that map to sequence of greater than 200 kb in length. Since the *O. nivara* BAC clones have an average insert size of 161 kb, pairs of BAC ends mapping to regions greater than 200 kb in the *japonica* genome could indicate the presence of an insertion in the *japonica* genome relative to the *O. nivara* genome. We identified 57 overlapping BACs from contig 236 that map between 200 and 400 kb in length (data not shown). This indicates that in the corresponding region of the *japonica* genome has undergone an

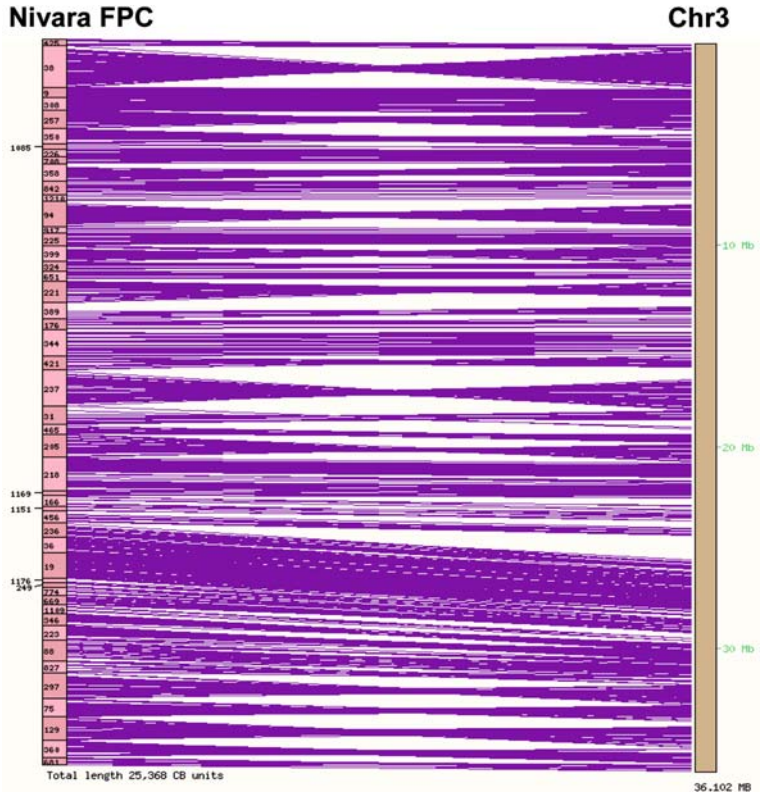

Figure 2. Alignment of *O. nivara* BAC FPC/STC contig map of chromosome 3 with Nipponbare rice chromosome 3 pseudomolecule using paired BAC end reads. Note: The apparent inversions are not real inversions but reflect the random orientation of FPC contigs that form during FPC assembly.

insertion event(s) of ~200–250 kb or a deletion(s) of same size sequence in the *O. nivara* genome. We plan to sequence *O. nivara* BACs from regions like these to compare the sequences between the two genomes. Such an analysis would provide a deeper understanding of the nature, timing, mechanisms and specificity of what events caused such recent sequence evolution in these two closely related species which may serve a general model for evolution of genomes within the genus *Oryza*.

Structural analysis of the Hd1 orthologous region in the O. nivara genome

To obtain a more detailed understanding of the syntenic relationships between the IRGSP genome sequence and OMAP species under investigation, we have initiated a study to sequence five orthologous regions across all 12 OMAP species.

One region on rice chromosome 6 contains the heading date locus, *Hd1,* which is an extremely important agronomic trait for rice production and one that has undergone intense selection during the process of domestication. *Hd1* has been shown to play a major role in flowering time in rice and was recently cloned and shown to be a homologue of the Arabidopsis CONSTANS gene (Putterill *et al*., 1995; Yano *et al*., 2000). It is proposed that *Hd1* functions in the promotion of heading under short-day conditions and in inhibition under long-day conditions.

To identify the *Hd1* orthologous region in *O. nivara,* we screened the BAC library with an *Hd1* gene specific probe and two genes located within 20 kb on either side of the *Hd1* gene resulting in the identification of 26 positive BACs. Twenty-two of these BACs were located in a single FPC contig # 189 and the remaining 4 did not have fingerprints. To further confirm that this contig represented the *Hd1* orthologous region, we mapped 51 of the possible 52 BAC-ends to 833, 443 bp of finished IRGSP rice genome sequence from the *Hd1* region of chromosome 6. Figure 3 shows FPC contig 189 and the unam-

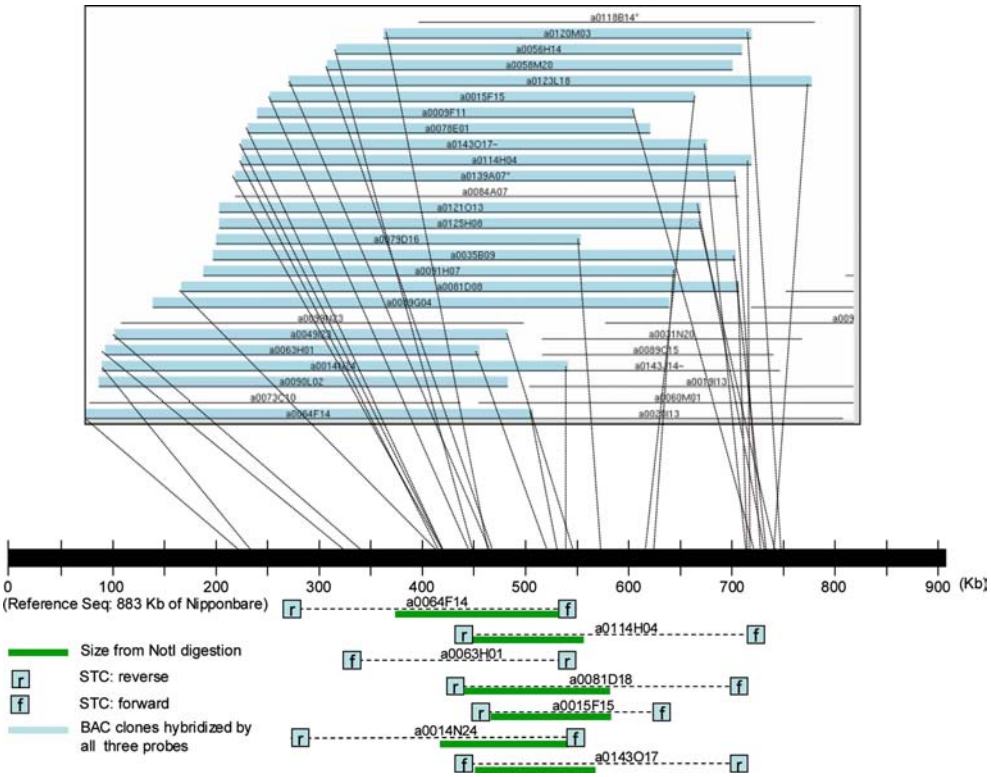

Figure 3. *O. nivara* contig # 189 is orthologous to the *O. sativa* ssp. *japonica* chromosome 6 region containing *Hd1*. Clones shaded in blue were shown to hybridize with an *Hd1* probe. Vertical lines represent BAC-ends that could be aligned to the IRGSP genome sequence.

biguous mapping of 37 of the BAC-end sequences demonstrating that this region is indeed orthologous to the *O. sativa* ssp. *japonica Hd1* locus.

Fourteen of the 37 BAC-end sequences were paired and thus could be used to compare the size of the *Hd1* regions between the cultivated and wild *O. nivara* species. Although the average insert size of the *O. nivara* BAC library was 161 kb, the 7 BAC-end pairs mapped to an average of 247 kb apart, thus indicating that the *Hd1* region in *japonica* rice was about 86 kb (~50%) larger than its progenitor *O. nivara* (Table 4). Upon closer examination of the insert sizes of these BAC clones by sizing NotI restricted BAC clones on a CHEF gel, inserts of 6 of the 7 BACs (Table 4) we found to be on average 25 kb smaller than the average for the BAC library. This analysis suggests that the *japonica Hd1* region is about 111 kb larger (80%) than in *O. nivara*.

The nature and timing of this genome expansion relative to *O. nivara* is now under investigation by performing a detailed sequence analysis of *O. nivara* BACs encompassing the *Hd1* region and will likely reveal important insight into the processes of speciation and domestication.

The Potential of filling physical gaps in the IRGSP genome sequence

One practical application of the *O. nivara* physical map is to identify contigs and individual BACs that span physical gaps in the IRGSP genome sequence. Sequence analysis of such contigs and clones could provide an estimate of the number of genes that are missing in the IRGSP genome sequence. Approximately 50 physical gaps, ranging in size from 10 to >500 kb, still exist in the IRGSP 'finished' genome sequence (T. Sasaki & J. Wu, personal communication). The primary reason for the inability to fill these gaps is due to: (1) lack of clones in the existing BAC and PAC libraries due to restriction enzyme bias during library construction; (2) the genomic regions are highly repetitive and therefore difficult to identify; and (3) the region is not clonable in *E. coli*.

The short arm of rice chromosome 3 contains two physical gaps (1 & 2) that are ~50 and ~500 kb in size as estimated by FISH (J. Jiang personnel communication). Our FPC/STC alignment analysis identified three *O. nivara* BACs with paired BAC ends from contig 236 that completely encompass the first 50 kb gap as shown in Figure 4. The fact that all three BAC end pairs share 99% sequence similarity with the *japonica* sequence flanking gap 1 suggest that sequencing of any one of these *O. nivara* BAC clones would uncover the genes missing in the 'finished' IRGSP genome sequence, with the assumption that the same number and kind of genes are present in both sister species.

Summary, conclusions and future research

The domestication of rice some 10, 000 years ago has severely limited the gene pool that breeders can utilize to improve rice. The wild species of the genus *Oryza* contain a wealth of genetic diversity that must therefore be uncovered if we are to meet the challenges of feeding the world in the 21st century.

The OMAP has been designed to conduct a detailed characterization of a single representative of each of the 10 genome types of wild rice species. The alignment of these genomes to the IRGSP genome sequence will provide a comprehensive physical framework whereby numerous genome-wide applied and basic research projects can be launched to unlock the genetic potential of these wild genomes and provide

Table 4. Size comparison between BAC-end sequence map position on the IRGSP genome sequence and BAC insert size for 7 *O. nivara* BACs orthologous to the *O. sativa* ssp. *japonica Hd1* genomic region.

BAC Clone	Size from STC (kb)	Size from not I Gel (kb)	Difference (kb)
OR_BBa0014N24	267	135	132
OR_BBa0015F15	164	125	39
OR_BBa0063H01	203	ND	ND
OR_BBa0064F14	269	161	108
OR_BBa0081D08	283	145	138
OR_BBa0114H04	273	121	152
OR_BBa0143O17	270	126	144

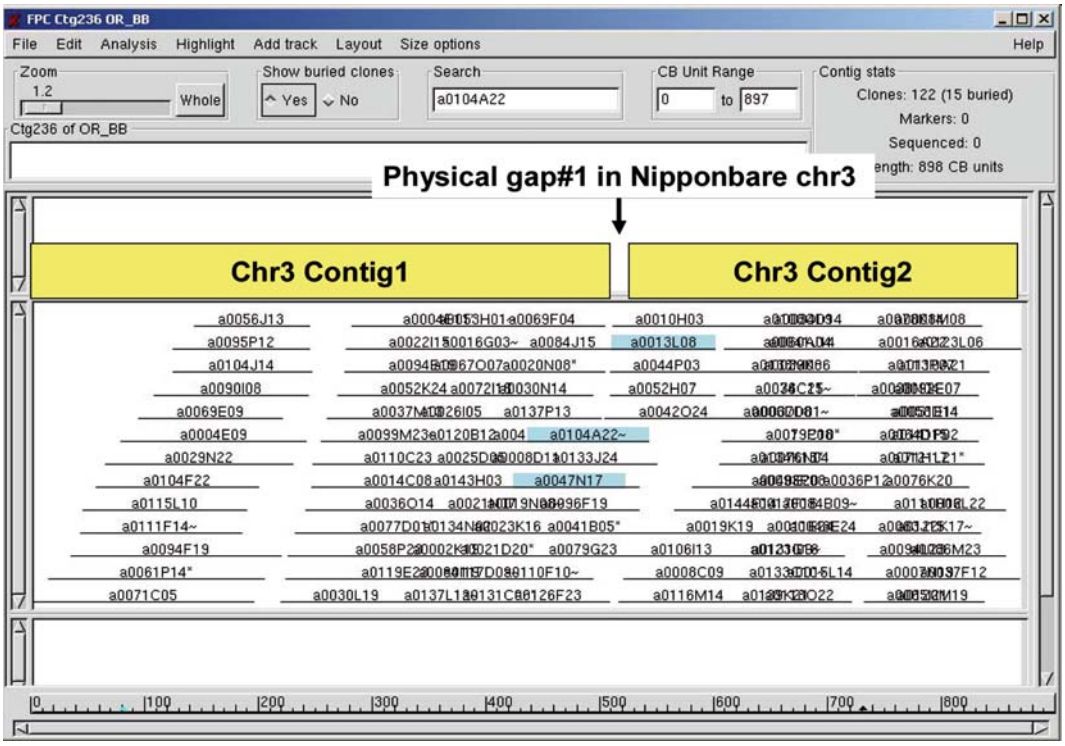

Figure 4. Gap filling with *O. nivara* contig 236.

breeders with new candidate genes and QTL for rice improvement.

To enhance the value of OMAP to plant breeding, we have established collaborations with rice geneticists in China and the United States to generate, phenotype and map BC4F2 advanced backcross populations of the three OMAP AA genome accessions we are working with. Once candidate genes and QTL are mapped we will have immediate access to these genomic regions for positional cloning.

OMAP has only just begun, but, we are already beginning to discover new insights into the genus *Oryza*. We achieved our first milestone by generating 12 wild rice BAC libraries that appear to be of sufficient genome depth and quality. The libraries are available to the scientific community and should provide a good resource for new gene discovery.

The second milestone was to develop robust protocols to be able to fingerprint approximately 1, 000, 000 BAC clones and twice that many BAC end sequences in an efficient and cost effective manner. Our production facility is now in high gear in order to complete the fingerprinting and end sequencing of all 12 wild rice genomes by September 2005.

We are presently optimizing our fingerprint assembly protocols, so that we can automatically create FPC maps that can be efficiently aligned to the IRGSP genome sequence. First we will align the AA genome diploids and establish all the protocols to align the remaining wild species. After the AA genome alignments are complete, we will move to the BB genome until we eventually align the most distantly related *Oryza* species – *O. granulata*. In this way, we can learn from previous alignments from more closely related species to make prediction as to how to assemble a more distantly related genome.

References

Ahn, S. Anderson, J.A., Sorrells, M.E. and Tanksley, S.D. 1993. Homoeologous relationships of rice, wheat and maize chromosomes. Mol. Gen. Genet. 241: 483–490.

Ahn, S. and Tanksley, S.D. 1993. Comparative linkage maps of the rice and maize genomes. Proc. Natl. Acad. Sci. USA 90: 7980–7984 .

Bennetzen, J.L. and Ma, J. 2003. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr. Opin. Plant Biol. 6: 128–133.

Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Frisch, D., Yu Y., Sun S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N.K., Tyagi, A.K., Soderlund, C., Dean, R.A. and Wing, R.A. 2002. An integrated physical and genetic map of the rice genome. Plant Cell 14: 537–545.

Cheng, Z., Buell, C.R., Wing, R.A., Gu, M. and Jiang, J. 2001. Toward a cytological characterization of the rice genome. Genome Res. 11: 2133–2141.

Ge, S., Sang, T., Lu, B.R. and Hong, D.Y. 1999. Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc. Natl. Acad. Sci. USA 96: 14400–14005.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica)*. Science 296: 92–100.

Han, B. and Xue, Y. 2003. Genome-wide intraspecific DNA-sequence variations in rice. Curr. Opin. Plant Biol. 6: 134–138.

Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S. and Sasaki, T. 1998. A high-density rice genetic linkage map with 2275 markers using a single F2 population. Genetics 148: 479–494.

Huang, H. and Kochert, G. 1994. Comparative RFLP mapping of an allotetraploid wild rice species (*Oryza latifolia)* and cultivated rice (*O. sativa)*. Plant Mol. Biol. 25: 633–638

Jena, K.K., Khush, G.S. and Kochert, G. 1994. Comparative RFLP mapping of a wild rice, *Oryza officinalis,* and cultivated rice *O. sativa*. Genome 37: 382–389.

Kaplinsky, N.J., Braun, D.M., Penterman, J., Goff, S.A. and Freeling, M. 2002. Utility and distribution of conserved noncoding sequences in the grasses. Proc. Natl. Acad. Sci. USA 99: 6147–6151.

Khush, G.S. 1997. Origin, dispersal, cultivation and variation of rice. Plant Mol. Biol. 35: 25–34.

Luo, M. and Wing, R.A. 2003. An improved method for plant BAC library construction. Methods Mol. Biol. 236: 3–20.

Luo, M.C., Thomas, C., You, F.M., Hsiao, J., Ouyang, S., Buell, C.R., Malandro, M., McGuire, P.E. Anderson, O.D. and Dvorak, J. 2003. High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82: 378–389.

Ma, J. and Bennetzen, J.L. 2004. Rapid recent growth and divergence of rice nuclear genomes. Proc. Natl. Acad. Sci. USA 101: 12404–12410.

Mao, L., Wood, T.C., Yu, Y., Budiman, M.A., Tomkins, J., Woo, S., Sasinowski, M., Presting, G., Frisch, D., Goff, S., Dean, R.A. and Wing, R.A. 2000. Rice transposable elements: a survey of 73,000 sequence-tagged-connectors. Genome Res. 10: 982–990.

Moore, G., Devos, K.M., Wang, Z. and Gale, M.D. 1995. Cereal genome evolution. Grasses, line up and form a circle. Curr. Biol. 5: 737–739.

Putterill, J., Robson, F., Lee, K., Simon, R. and Coupland, G. 1995. The CONSTANS gene of Arabidopsis promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. Cell 80: 847–857.

Salse, J., Piegu, B., Cooke, R. and Delseny, M. 2004. New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome. Plant J. 38: 396–409.

Sasaki, T., Matsumoto, T., Yamamoto, K., Sakata, K., Baba, T., Katayose, Y., Wu, J., Niimura, Y., Cheng, Z., Nagamura, Y., Antonio, BA., Kanamori, H., Hosokawa, S., Masukawa, M., Arikawa, K., Chiden, Y., Hayashi, M., Okamoto, M. Ando, T., Aoki, H., Arita, K., Hamada, M., Harada,C., Hijishita, S., Honda, M., Ichikawa, Y., Idonuma, A., Iijima, M., Ikeda, M., Ikeno, M., Ito, S., Ito, T., Ito, Y., Iwabuchi, A., Kamiya, K., Karasawa, W., Katagiri, S., Kikuta, A., Kobayashi, N., Kono, I., Machita, K., Maehara T., Mizuno, H., Mizubayashi, T., Mukai, Y., Nagasaki, H., Nakashima, M., Nakama, Y., Nakamichi, Y., Nakamura, M., Namiki, N., Negishi, M., Ohta, I., Ono, N., Saji, S., Sakai, K., Waki, K., Yamagata, H., Yamane, H., Yoshiki, S., Yoshihara, R., Yukawa, K., Zhong, H., Iwama, H., Endo, T., Ito, H., Hahn, J.H., Kim, H.I., Eun, M.Y., Yano, M., Jiang J. and Gojobori, T. 2002. The genome sequence and structure of rice chromosome 1. Nature 420: 312–316.

Soderlund, C., Humphray, S., Dunham, A. and French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10: 1772–1787.

The Rice Chromosome 10 Sequencing Consortia 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. Science 300: 1566–1569.

Vaughan, D.A., Morishima, H. and Kadowaki, K. 2003. Diversity in the *Oryza* genus. Curr. Opin. Plant Biol. 6: 139–146.

Ware, D., Jaiswal, P., Ni. J., Pan, X., Chang, K., Clark, K., Teytelman, L., Schmidt, S., Zhao, W., Cartinhour, S., McCouch, S. and Stein, L. 2002. Gramene: a resource for comparative grass genomics. Nucleic Acids Res. 30: 103–105.

Wu, J., Maehara, T., Shimokawa, T., Yamamoto, S., Harada, C., Takazaki, Y., Ono, N., Mukai, Y., Koike, K., Yazaki, J., Fujii, F., Shomura, A. Ando, T., Kono, I., Waki, K., Yamamoto, K., Yano, M., Matsumoto, T. and Sasaki, T. 2002. A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14: 525–535.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. and Sasaki, T. 2000. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell 12: 2473–2484.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, W., Ye, C., Tong, W.,

Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Li, J., Liu, Z., Qi, Q., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J.,Tan, J., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Zhao, W., Li, P., Chen, W., Zhang, Y., Hu, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Tao, M., Zhu, L., Yuan, L. and Yang, H. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Indica)*. Science 296: 79–92.