A 1.7 Mb Sequence Analysis of Rice Chromosome 1

Atsuko Idonuma aido@staff.or.jp
Hideki Nagasaki nagasaki@staff.or.jp
Masatoshi Masukawa masukawa@staff.or.jp
Manami Negishi negishi@staff.or.jp
Koji Arikawa arikawa@staff.or.jp
Yoshiyuki Mukai mukai@staff.or.jp
Isamu Ohta ohta@staff.or.jp
Baltazar A. Antonio antonio@staff.or.jp
Katsumi Sakata ksakata@abr.affrc.go.jp
Takuji Sasaki tsasaki@abr.affrc.go.jp

Rice Genome Research Program, NIAR/STAFF 446-1, Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki, 305-0854, Japan

Keywords: rice, sequence analysis, sequence-ready map, annotation

1 Introduction

Whole genome sequencing produces the fundamental knowledge on the structure and function of all genes in an organism, and provides the essential tools for manipulation of biologically important genes. Recently, a completed sequence of the entire chromosome of eukaryote has been reported for human [1] and Arabidopsis [2]. Rice is also a major target for whole genome sequencing because it is a valuable source of food to an ever-expanding world population. Its nearly 430 Mb genome is the smallest in size among cereal crops and could therefore provide valuable information for understanding many agricultural crops. In this report, we will focus on the first step towards the goal of sequencing the entire rice genome.

2 Genome Sequencing with PAC Clones

The International Rice Genome Sequencing Project (IRGSP) was initiated in 1998 in order to accelerate the rice genome sequencing effort. A sequence-ready map construction approach was adopted over whole-genome shotgun for the following reasons: it allows (1) efficient sharing of resources without overlap among collaborators, (2) confirmation of locus and clone order with available genetic and physical maps while sequencing, (3) immediate feedback to solve various sequencing errors, and (4) early completion of a target region.

As part of the IRGSP, the Rice Genome Research Program (RGP) is sequencing chromosomes 1 and 6. A high-density linkage map has been integrated with a yeast artificial chromosome (YAC) physical map of about 70% genome coverage. These are then used to construct a sequence-ready map with genomic clones of P1-derived artificial chromosome (PACs) by PCR screening using STS markers on genetic map and EST markers mapped onto YAC physical map. The selected PAC inserts are subcloned into pUC18 to facilitate shotgun-based sequencing and assembly.

3 Annotation Protocol and Publication

The sequence analysis of a 1.7 Mb region on the short arm of chromosome 1 has been recently completed (Figure 1). This region covers about 3% of chromosome 1 which is estimated to be about 52 Mb in total length. The genome sequence was analyzed with GENSCAN prediction programs for gene identification, splice site determination with SplicePredictor, and alignment with ESTs and proteins with BLASTN2 and BLASTX2. Then it was compared with the EST database of RGP and the protein database NRP from the MAFF (Ministry of Agriculture, Forestry and Fisheries) DNA Bank. The long terminal repeats (LTR) at both ends of retrotransposons were determined using an original program. These results were integrated by an annotation-plotting tool and manually edited. Protein similarities of the identified coding sequence were searched against NRP with BLASTP2. From the results of annotation, this region was found to be gene-rich with one gene for every 5 kb sequence.

All annotated sequence are released to the public domain through DDBJ/EMBL/GenBank and accessible through our database INE (INtegrated rice genome Explorer) [3] on our website at http://www.dna.affrc.go.jp:82/.

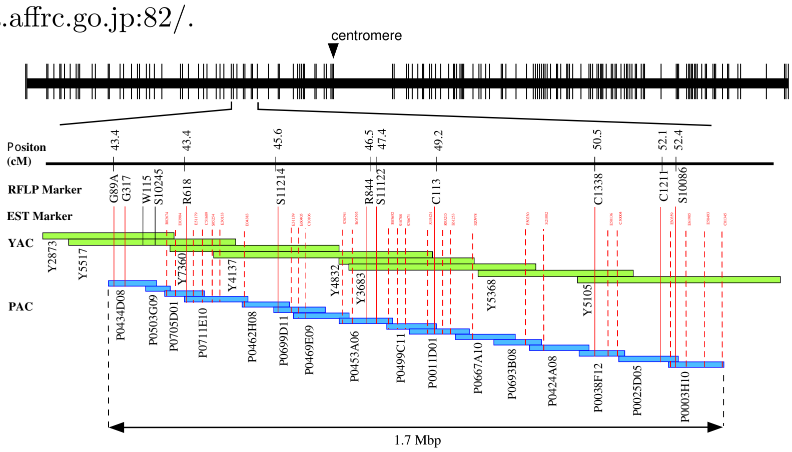

Figure 1: A 1.7 Mb contiguous region consisting of 16 PAC clones. These PACs were anchored on the short arm of chromosome 1 corresponding to about 10 cM of the genetic map (top). The DNA markers on the genetic map as well as the EST markers on the YAC physical map used for anchoring the PACs are also shown.

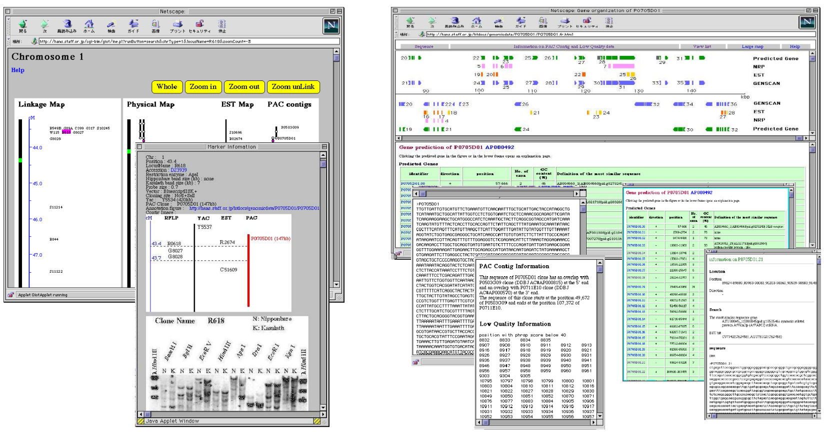

Figure 2: The annotated sequence is integrated with the genetic map, YAC physical map and EST map which can be accessed through INE. The predicted genes and the result of BLAST and GENSCAN, the complete sequence of the clone, information on overlap and Phrap assembly quality, and other details on predicted genes are presented on the annotation site for each sequenced PAC.

References

[1] Dunham, I., *et al.*, The DNA sequence of human chromosome 22, *Nature*, 402:489–495, 1999.

[2] Lin, X., *et al.*, Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*, *Nature*, 402:761–768, 1999.

[3] Sakata, K., *et al.*, INE: a rice genome database with an integrated map view, *NAR*, 28, 2000.