D1028–D1033 *Nucleic Acids Research, 2008, Vol. 36, Database issue* *Published online 17 December 2007*
doi:10.1093/nar/gkm978

The Rice Annotation Project Database (RAP-DB): 2008 update*

Received September 14, 2007; Revised October 17, 2007; Accepted October 18, 2007

ABSTRACT

The Rice Annotation Project Database (RAP-DB) was created to provide the genome sequence assembly of the International Rice Genome Sequencing Project (IRGSP), manually curated annotation of the sequence, and other genomics information that could be useful for comprehensive understanding of the rice biology. Since the last publication of the RAP-DB, the IRGSP genome has been revised and reassembled. In addition, a large number of rice-expressed sequence tags have been released, and functional genomics resources have been produced worldwide. Thus, we have thoroughly updated our genome annotation by manual curation of all the functional descriptions of rice genes. The latest version of the RAP-DB contains a variety of annotation data as follows: clone positions, structures and functions of 31 439 genes validated by cDNAs, RNA genes detected by massively parallel signature sequencing (MPSS) technology and sequence similarity, flanking sequences of mutant lines, transposable elements, etc. Other annotation data such as Gnomon can be displayed along with those of RAP for comparison. We have also developed a new keyword search system to allow the user to access useful information. The RAP-DB is available at: http://rapdb.dna.affrc.go.jp/ and http://rapdb.lab.nig.ac.jp/.

INTRODUCTION

Genome-wide studies of the major cereals, including rice, have been promoted worldwide in order to respond to the expected demand of increasing food supplies. In particular, genome sequence annotation plays a pivotal role to explore agronomically useful traits by large-scale experimental analyses, and several databases about cereal genome information have been developed (1–3). After the completion of the genome sequencing of the japonica rice cultivar Nipponbare (4), the Rice Annotation Project (RAP) was organized to create annotation data with high accuracy and reliability (5). To provide the rice genome annotation, we have created the RAP-DB (6), which is a portal site for various types of data, such as the genome assembly of the IRGSP, curated annotation of the genome and full-length cDNAs (FLcDNAs) (7) and related information beneficial to researchers of rice and other cereals.

As biological data of rice continue to increase, the RAP-DB must continue to supply the most up-to-date information. For instance, the IRGSP has released the build 4 assembly, 581 446 5′- or 3′-end sequences of FLcDNA clones have been determined, and 77 763 flanking sequence tags have been generated by 10 independent functional genomics groups (8–19). Thus, the RAP annotation was extensively revised in gene structures, functional descriptions, etc. Moreover, to facilitate user access, improvements were made in some RAP-DB functions, such as a novel database search system. Here we describe the latest version of the RAP-DB that consists of the updated genome annotation and user-friendly functionalities to access the data.

NEW DATA CONTENTS

The IRGSP genome build 4 and updated RAP data

The IRGSP genome was updated and reassembled. The rice genome sequence was determined by the map-based clone-by-clone sequencing strategy using bacterial and P1 artificial chromosome (BAC and PAC, respectively) clones (4). All of the clone sequences (January 2005 data freeze) were assembled and overlaps between neighboring clones were manually removed and the lengths of all gaps were estimated by the fiber-FISH method (4). We manually checked the positions and orders of the clones, using genetic and EST markers (20,21). This new version, build 4, contains 29 newly sequenced clones and 96 updated clones. The previous version, build 3 (June 2004 data freeze), contained 49 redundant clones that had been erroneously incorporated. These clones were discarded in build 4. As a result, the build 4 assembly makes up 95.4% of the *Oryza sativa* L. ssp. *japonica* cultivar Nipponbare genome. The genomic locations of BAC/PAC sequences can be displayed in the 'Region' and 'Details' panels of GBrowse (22) by checking the 'BAC/PAC' item (Figure 1). Users can search for the BAC/PAC clones by their accession numbers or clone names.

Our genome annotation is primarily based on evidence of expressed transcripts (5). In addition to FLcDNA sequences of rice (7), we used 581 446 5′- or 3′-end

*A complete list of authors appears at the end of this article.

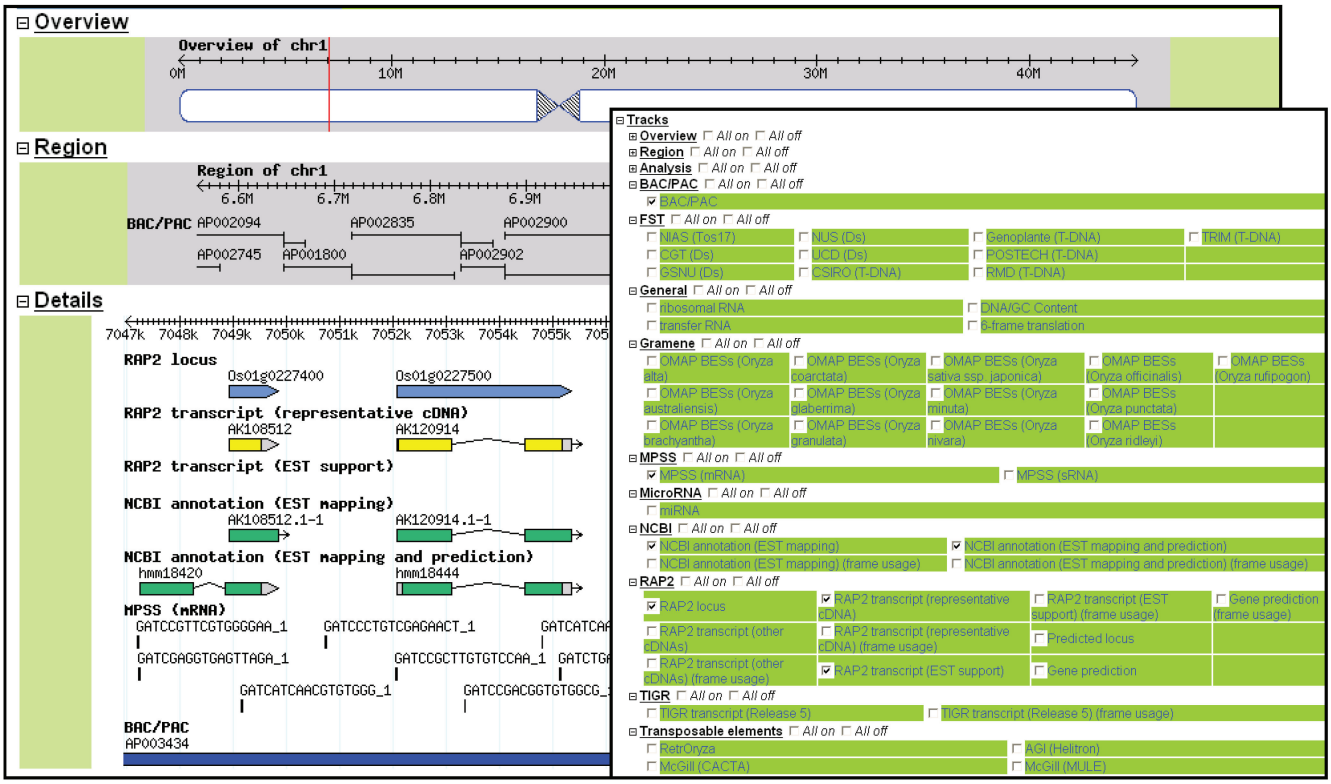

Figure 1. Schematic view of the annotation browser and items that can be selected.

sequences derived from rice FLcDNA clones that were registered in the International Nucleotide Sequence Databases (accession numbers CI000001-CI778739). All of the cDNA sequences were aligned to the genome by the method previously described (5). Please note that we defined a locus as a region covered by overlapping cDNAs and that different loci overlap only when the loci are nested or contained in an intronic region. Protein-coding genes were predicted by Fgenesh, GENSCAN and GLocate, and a single gene structure was determined for a locus by a modified version of Combiner (5). To validate these *ab initio* predictions, we also employed 380 812 rice expressed sequence tags (ESTs) and more than 2 million mRNAs and ESTs of non-rice plants (*Hordeum vulgare*, *Sorghum bicolor*, *Saccharum officinarum*, *Triticum aestivum*, *Zea mays*). We determined 31 439 loci that were supported by evidence of expression, and 30 192 of which showed the potential of coding for protein (Table 1) (6). Functional descriptions of these loci were produced by automated methods. If the descriptions were updated since the previous annotation, they were manually curated by the method previously described using our custom-made curation system (5). The curated functions of the open reading frames (ORFs), which were defined as the interval between the start and stop codons, were classified into five categories according to their level of sequence similarity (Table 2). The probable protein products of 8226 loci had functions identified or inferred by BLASTX searches against UniProt Knowledgebase (Categories I and II). In addition, 13 632 loci possessed functional domain(s) detected by InterProScan (Category III). We also examined 1247 transcripts in which no coding potential was suggested, and found 176 putative non-protein-coding RNAs by the method previously described (5). In the RAP-DB, the loci and transcripts are linked to a page of detailed description including the level of evidence, InterPro domains, Gene Ontology annotations and other information so that researchers can easily access these useful resources.

Comparison with Gnomon's annotation

Although cDNA-based annotation such as RAP generally has high accuracy, different annotation methods produce different results. In fact, a comparison of human genes annotated by several projects showed marked variation in their genomic structures (23). To validate our annotation, we compared the gene positions of RAP with those of Gnomon (Figure 1), which is an integrative annotation pipeline developed by National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genome/guide/gnomon.html). Gnomon combines *ab initio* predictions with sequence homology. We found that 32 664 FLcDNAs including redundant sequences were mapped to the build 4 assembly by the RAP method and 33 937 by the Gnomon pipeline. Our comparison of the exon positions revealed that 24 836 (76.0%) of the genes determined by RAP had the identical exon–intron structures to those by Gnomon. Furthermore, 31 433 (96.2%) of the RAP and Gnomon

Table 1. Statistics of rice genes

Number of expressed loci	31 439
Protein-coding loci with FLcDNAs	25 012
Non-protein-coding loci with FLcDNAs	1247
Ab initio predictions with evidence of expression	5180
Ab initio predictions without evidence of expression	22 022

Table 2. Classification of ORFs

Category[a]	Definition	Number of ORFs
I	Identical to known rice protein	664
II	Similar to known protein	7562
III	InterPro domain-containing protein	13 632
IV	Conserved hypothetical protein	6954
V	Hypothetical protein	1380

[a]ORFs were classified as previously described (5).

genes overlapped with each other in their genomic positions. The inconsistency of the cDNA-mappings between RAP and Gnomon was due largely to the differences of 5′ or 3′ end alignments. This can be accounted for by poor sequence quality, such as contamination with a vector sequence. Both methods, therefore, present highly similar results. Since these annotation pipelines are independent and displayed similar gene structures, cDNA-based genes provided by both methods should be reliable. However, for the computationally predicted genes without FLcDNA evidence, only 13 123 (48.2%) of 27 202 RAP genes unsupported by cDNAs were covered by Gnomon's genes. Since the current gene-finding methods inevitably generate a large number of erroneous predictions, these hypothetical genes should be validated by cDNAs in future.

MPSS and small interfering RNA (siRNA)-producing genomic regions

Although FLcDNAs are regarded as the best evidence of expressed genes, it is laborious to determine a large number of full-length transcripts. The MPSS method is a sophisticated technique for the global identification of RNA molecules (24). Since small RNA MPSS signatures of rice are currently available (25), they were mapped to the IRGSP build 4 genome (Figure 1). A total of 2 953 855 small RNA signatures that were derived from untreated flower, seedling and stem tissues were sequenced and 284 301 distinct signatures were identified from these three libraries (25). Among these distinct signatures, 204 136 matched to the IRGSP genome with numbers of hits per signature ranging from 1 to 9122. When we compared the loci determined by RAP with the MPSS signatures that mapped uniquely to a single location of the genome, we found that 68.7% of the RAP loci were supported by the MPSS signatures. This proportion is higher than that estimated by a comparison between a genome-wide tiling array and rice genes determined by another project (64.8% of non-transposons) (26).

To annotate siRNA-producing genomic regions, we grouped small RNA signatures into clusters if adjacent signatures were located within 500 bp of each other. With this strategy, 159 410 clusters were identified on the genome; the largest cluster has 15 193 signatures in a 75 375 bp region. Since the heterochromatic siRNAs are known to form relatively dense clusters, we used a cutoff value of 10 signatures per cluster to identify siRNA-producing regions. Dense clusters were observed not only in the centromeric regions but also in the pericentromeric regions. Approximately one-third (56 371) of the clusters have more than 10 signatures per cluster, indicative of the high complexity of heterochromatic siRNAs in rice.

Identification of microRNA (miRNA) genes

miRNAs are single-stranded RNAs that are composed of ~21 nt. They are known to play important roles in eukaryotic gene regulation (27). We annotated rice miRNA genes by using a data set compiled in the miRBase database, release 9.1 (28). To detect miRNA gene candidates, we employed criteria that have been adopted for other species (29). The miRNAs of miRBase were mapped to the IRGSP genome (Figure 1), if both 5′- and 3′-flanking regions could form a stem–loop structure, which is an important feature to distinguish between true and false predictions. We successfully identified 239 miRNA genes that belonged to 61 families defined in miRBase. The miRNAs detected can be displayed in the microRNA track of GBrowse. We found that 20 of these miRNA gene families were conserved in *Arabidopsis* and poplar, whereas 41 families, many of which were single-copy genes, were specific to rice. It is noteworthy that in some cases FLcDNAs with no coding potential had been cloned over predicted miRNA regions. These might be precursors of miRNAs that could be processed to the functional form of ~21 nt. Some miRNA candidates were mapped to regions in which transposable elements (TEs) were enriched. The functions of these candidates should be examined by experimentation.

Other new data and functions

The genome sequencing of rice was expected to facilitate large-scale analyses of gene functions. Mutant resources, for functional genomics studies, have been produced by several groups. To provide easy access to such resources, we integrated the mutant information created by 10 independent groups (8–19). All the flanking sequences that were tagged by *Tos17*, T-DNA and *Ds* were compared with the rice genome so that the positions of genes disrupted by different methods were simultaneously displayed in the RAP-DB (Figure 1). These flanking sequences have been linked to the web pages of the mutant providers.

More than 30% of the rice genome consists of TEs (4). A genome-wide analysis suggests that rice TEs have

played several roles during the genome evolution (30). To annotate TEs, we first transferred the *Mutator*-like elements (MULE) positions of the build 2 assembly, determined by IRGSP, to build 4 (4). In addition, CACTA and *Helitron* elements were newly surveyed and detected in build 4. LTR-retrotransposons were identified by the method of RetrOryza (31).

To assist user access to the RAP-DB, the keyword search functionality has been improved. Users can specify a section of annotation and genomic positions to be searched. In addition, since there are other annotation activities of the rice genome, such as Osa1 and BGI-RIS (3,32), a converter of gene identifiers is provided. The Os code, which is the locus identifier of the IRGSP/RAP annotation, can be converted to the LOC_Os identifier of Osa1 (3), and vice versa. This conversion system can deal with multiple identifiers separated by spaces or commas.

FUTURE DIRECTIONS

We have developed the RAP-DB as an integrative database of the IRGSP genome in which we aim to collect information relevant to bioinformatics and to functional genomics, breeding, etc. We plan to add data for molecular markers, genetic maps, orthology to *Arabidopsis* genes, EC numbers and some other results of data analysis. Since a large number of the RAP loci contain alternative splicing variants, an identification number will be assigned to each variant. The annotation of the RAP loci, such as electronically assigned Gene Ontology annotations, will be provided to other data resources. New, high-throughput DNA-sequencing technologies are being developed and it is expected that the number of rice species and cultivar genome sequences will rapidly grow. These new sequences will be incorporated into the RAP-DB by comparison to the Nipponbare reference genome. A large amount of sequence data from variant species and cultivars may increase the difficulty of finding desired information. We, therefore, plan to further improve the database search system.

ACKNOWLEDGEMENTS

The authors thank Pankaj Jaiswal, Chengzhi Liang and Sharon Wei for the information about the positions of OMAP BAC ends, and Kumiko Suzuki and Chieko Kobayashi for their technical assistance. This work was supported by a grant from the Special Coordination Funds for Promoting Science and Technology of the Ministry of Education, Culture, Sports, Science and Technology of Japan, by a grant for the NIAS Genebank Project, and by a grant for the Project ANR OsmiR NT05-3 42996. Funding to pay the Open Access publication charges for this article was provided by National Institute of Agrobiological Sciences.

Conflict of interest statement. None declared.

REFERENCES

1. Jaiswal,P., Ni,J., Yap,I., Ware,D., Spooner,W., Youens-Clark,K., Ren,L., Liang,C., Zhao,W. *et al.* (2006) Gramene: a bird's eye view of cereal genomes. *Nucleic Acids Res.*, **34**, D717–D723.
2. Droc,G., Ruiz,M., Larmande,P., Pereira,A., Piffanelli,P., Morel,J.B., Dievart,A., Courtois,B., Guiderdoni,E. *et al.* (2006) OryGenesDB: a database for rice reverse genetics. *Nucleic Acids Res.*, **34**, D736–D740.
3. Ouyang,S., Zhu,W., Hamilton,J., Lin,H., Campbell,M., Childs,K., Thibaud-Nissen,F., Malek,R.L., Lee,Y. *et al.* (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Res.*, **35**, D883–D887.
4. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.
5. The Rice Annotation Project (2007) Curated genome annotation of *Oryza sativa* ssp. *japonica* and comparative genome analysis with *Arabidopsis thaliana*. *Genome Res.*, **17**, 175–183.
6. Ohyanagi,H., Tanaka,T., Sakai,H., Shigemoto,Y., Yamaguchi,K., Habara,T., Fujii,Y., Antonio,B.A., Nagamura,Y. T. *et al.* (2006) The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. *japonica* genome information. *Nucleic Acids Res.*, **34**, D741–D744.
7. Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H. *et al.* (2003) Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. *Science*, **301**, 376–379.
8. Miyao,A., Tanaka,K., Murata,K., Sawaki,H., Takeda,S., Abe,K., Shinozuka,Y., Onosato,K. and Hirochika,H. (2003) Target site specificity of the *Tos17* retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell*, **15**, 1771–1780.
9. Eamens,A.L., Blanchard,C.L., Dennis,E.S. and Upadhyaya,N.M. (2004) A bidirectional gene trap construct suitable for T-DNA and *Ds*-mediated insertional mutagenesis in rice (*Oryza sativa* L.). *Plant Biotechnol. J.*, **2**, 367–380.
10. Sallaud,C., Gay,C., Larmande,P., Bes,M., Piffanelli,P., Piegu,B., Droc,G., Regad,F., Bourgeois,E. *et al.* (2004) High-throughput T-DNA insertion mutagenesis in rice: a first step towards *in silico* reverse genetics. *Plant J.*, **39**, 450–464.
11. Jeon,J.S., Lee,S., Jung,K.H., Jun,S.H., Jeong,D.H., Lee,J., Kim,C., Jang,S., Yang,K. *et al.* (2000) T-DNA insertional mutagenesis for functional genomics in rice. *Plant J.*, **22**, 561–570.
12. Kim,C.M., Piao,H.L., Park,S.J., Chon,N.S., Je,B.I., Sun,B., Park,S.H., Park,J.Y., Lee,E.J. *et al.* (2004) Rapid, large-scale generation of *Ds* transposant lines and analysis of the *Ds* insertion sites in rice. *Plant J.*, **39**, 252–263.
13. Kolesnik,T., Szeverenyi,I., Bachmann,D., Kumar,C.S., Jiang,S., Ramamoorthy,R., Cai,M., Ma,Z.G., Sundaresan,V. *et al.* (2004) Establishing an efficient *Ac*/*Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. *Plant J.*, **37**, 301–314.
14. Hsing,Y.I., Chern,C.G., Fan,M.J., Lu,P.C., Chen,K.T., Lo,S.F., Sun,P.K., Ho,S.L., Lee,K.W. *et al.* (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. *Plant Mol. Biol.*, **63**, 351–364.
15. van Enckevort,L.J., Droc,G., Piffanelli,P., Greco,R., Gagneur,C., Weber,C., Gonzalez,V.M., Cabot,P., Fornara,F. *et al.* (2005) EU-OSTID: a collection of transposon insertional mutants for functional genomics in rice. *Plant Mol. Biol.*, **59**, 99–110.
16. Zhang,J., Li,C., Wu,C., Xiong,L., Chen,G., Zhang,Q. and Wang,S. (2006) RMD: a rice mutant database for functional analysis of the rice genome. *Nucleic Acids Res.*, **34**, D745–D748.
17. An,S., Park,S., Jeong,D.H., Lee,D.Y., Kang,H.G., Yu,J.H., Hur,J., Kim,S.R., Kim,Y.H. *et al.* (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. *Plant Physiol.*, **133**, 2040–2047.
18. Ryu,C.H., You,J.H., Kang,H.G., Hur,J., Kim,Y.H., Han,M.J., An,K., Chung,B.C., Lee,C.H. *et al.* (2004) Generation of T-DNA tagging lines with a bidirectional gene trap vector and the establishment of an insertion-site database. *Plant Mol. Biol.*, **54**, 489–502.
19. Jeong,D.H., An,S., Park,S., Kang,H.G., Park,G.G., Kim,S.R., Sim,J., Kim,Y.O., Kim,M.K. *et al.* (2006) Generation of a

flanking sequence-tag database for activation-tagging lines in japonica rice. *Plant J.*, **45**, 123–132.
20. Harushima,Y., Yano,M., Shomura,A., Sato,M., Shimano,T., Kuboki,Y., Yamamoto,T., Lin,S.Y., Antonio,B.A. *et al.* (1998) A high-density rice genetic linkage map with 2275 markers using a single F2 population. *Genetics*, **148**, 479–494.
21. Wu,J., Maehara,T., Shimokawa,T., Yamamoto,S., Harada,C., Takazaki,Y., Ono,N., Mukai,Y., Koike,K. *et al.* (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites. *Plant Cell*, **14**, 525–535.
22. Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
23. Hsu,F., Kent,W.J., Clawson,H., Kuhn,R.M., Diekhans,M. and Haussler,D. (2006) The UCSC Known Genes. *Bioinformatics*, **22**, 1036–1046.
24. Meyers,B.C., Vu,T.H., Tej,S.S., Ghazal,H., Matvienko,M., Agrawal,V., Ning,J. and Haudenschild,C.D. (2004) Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. *Nat. Biotechnol.*, **22**, 1006–1011.
25. Nobuta,K., Venu,R.C., Lu,C., Belo,A., Vemaraju,K., Kulkarni,K., Wang,W., Pillay,M., Green,P.J. *et al.* (2007) An expression atlas of rice mRNAs and small RNAs. *Nat. Biotechnol.*, **25**, 473–477.
26. Li,L., Wang,X., Sasidharan,R., Stolc,V., Deng,W., He,H., Korbel,J., Chen,X., Tongprasit,W. *et al.* (2007) Global identification and characterization of transcriptionally active regions in the rice genome. *PLoS ONE*, **2**, e294.
27. Jones-Rhoades,M.W., Bartel,D.P. and Bartel,B. (2006) MicroRNAS and their regulatory roles in plants. *Annu. Rev. Plant Biol.*, **57**, 19–53.
28. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
29. Ambros,V., Bartel,B., Bartel,D.P., Burge,C.B., Carrington,J.C., Chen,X., Dreyfuss,G., Eddy,S.R., Griffiths-Jones,S. *et al.* (2003) A uniform system for microRNA annotation. *RNA*, **9**, 277–279.
30. Sakai,H., Tanaka,T. and Itoh,T. (2007) Birth and death of genes promoted by transposable elements in *Oryza sativa*. *Gene*, **392**, 59–63.
31. Chaparro,C., Guyot,R., Zuccolo,A., Piegu,B. and Panaud,O. (2007) RetrOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Res.*, **35**, D66–D70.
32. Zhao,W., Wang,J., He,X., Huang,X., Jiao,Y., Dai,M., Wei,S., Fu,J., Chen,Y. *et al.* (2004) BGI-RIS: an integrated information resource and comparative analysis workbench for rice genomics. *Nucleic Acids Res.*, **32**, D377–D382.

LIST OF AUTHORS FOR THE RICE ANNOTATION PROJECT CONSORTIUM

Tsuyoshi Tanaka[1], Baltazar A. Antonio[1], Shoshi Kikuchi[1], Takashi Matsumoto[1], Yoshiaki Nagamura[1], Hisataka Numa[1], Hiroaki Sakai[1], Jianzhong Wu[1], Takeshi Itoh[1,2,†], Takuji Sasaki[1], Ryo Aono[3], Yasuyuki Fujii[3,4], Takuya Habara[3], Erimi Harada[3], Masako Kanno[3], Yoshihiro Kawahara[3,5], Hiroaki Kawashima[3], Hiromi Kubooka[3], Akihiro Matsuya[3], Hajime Nakaoka[3], Naomi Saichi[3], Ryoko Sanbonmatsu[3], Yoshiharu Sato[3], Yuji Shinso[3], Mami Suzuki[3], Jun-ichi Takeda[3], Motohiko Tanino[3], Fusano Todokoro[3], Kaori Yamaguchi[3], Naoyuki Yamamoto[3], Chisato Yamasaki[3], Tadashi Imanishi[2], Toshihisa Okido[6], Masahito Tada[6], Kazuho Ikeo[6], Yoshio Tateno[6], Takashi Gojobori[6], Yao-Cheng Lin[7], Fu-Jin Wei[7], Yue-ie Hsing[7], Qiang Zhao[8], Bin Han[8], Melissa R. Kramer[9], Richard W. McCombie[9], David Lonsdale[10], Claire C. O'Donovan[10], Eleanor J. Whitfield[10], Rolf Apweiler[10], Kanako O. Koyanagi[11], Jitendra P. Khurana[12], Saurabh Raghuvanshi[12], Nagendra K. Singh[13], Akhilesh K. Tyagi[12], Georg Haberer[14], Masaki Fujisawa[15], Satomi Hosokawa[15], Yukiyo Ito[15], Hiroshi Ikawa[15], Michie Shibata[15], Mayu Yamamoto[15], Richard M. Bruskiewich[16], Douglas R. Hoen[17], Thomas E. Bureau[17], Nobukazu Namiki[18], Hajime Ohyanagi[18], Yasumichi Sakai[18], Satoshi Nobushima[18], Katsumi Sakata[18], Roberto A. Barrero[6,19], Yutaka Sato[20], Alexandre Souvorov[21], Brian Smith-White[21], Tatiana Tatusova[21], Suyoung An[22], Gynheung An[22], Satoshi OOta[23], Galina Fuks[24], Joachim Messing[24], Karen R. Christie[25], Damien Lieberherr[26], HyeRan Kim[27], Andrea Zuccolo[27], Rod A. Wing[27], Kan Nobuta[28], Pamela J. Green[28], Cheng Lu[28], Blake C. Meyers[28], Cristian Chaparro[29], Benoit Piegu[29], Olivier Panaud[29] and Manuel Echeverria[29]

[1]National Institute of Agrobiological Sciences, Ibaraki 305-8602, Japan, [2]Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, Japan, [3]Japan Biological Informatics Consortium, Tokyo 135-0064, Japan, [4]Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Innovation Center Okayama for Nanobio-targeted Therapy, Okayama 700-8558, Japan, [5]Department of Biological Sciences, Tokyo Metropolitan University, Tokyo 192-0397, Japan, [6]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Research Organization of Information and Systems, Shizuoka 411-8540, Japan, [7]Institute of Botany, Academia Sinica, Taipei 11529, Taiwan, [8]Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China, [9]Cold Spring Harbor Laboratory, NY 11723, USA, [10]European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, CB10 1SD, UK, [11]Graduate School of Information Science and Technology, Hokkaido University, Hokkaido 060-0814, Japan, [12]Department of Plant Molecular Biology, University of Delhi South Campus, New Delhi 110021, India, [13]National Research Centre on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi 110012, India, [14]Institute for Bioinformatics/MIPS, GSF National Research Center for Environment and Health, D-85764 Neuherberg, Germany, [15]Institute of the Society for Techno-innovation of Agriculture, Forestry and Fisheries, Ibaraki 305-0854, Japan, [16]Crop Research Informatics Laboratory, International Rice Research Institute, Metro Manila, Philippines, [17]Department of Biology, McGill University, Quebec H3A 1B1, Canada, [18]Tsukuba Division, Mitsubishi Space Software Co., Ltd., Ibaraki 305-0032, Japan, [19]Centre for Comparative Genomics, Murdoch University, Western Australia 6150, Australia, [20]Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan, [21]National Center for Biotechnology

Information, National Institutes of Health, MD 20894, USA, [22]Pohang University of Science and Technology, Pohang 790-784, Korea, [23]RIKEN BioResource Center, RIKEN Tsukuba Institute, Ibaraki 305-0074, Japan, [24]Waksman Institute of Microbiology, Rutgers University, NJ 08854, [25]Stanford University Medical Center, CA 94305-5120, USA, [26]Swiss-Prot Group, Swiss Institute of Bioinformatics, Geneva 1206, Switzerland, [27]Arizona Genomics Institute, The University of Arizona, AZ 85721, USA, [28]University of Delaware, DE 19711, USA and [29]University of Perpignan, UMR CNRS-IRD 5096, Perpignan 66860, France

[†]To whom correspondence should be addressed. Tel/Fax: +81 29 838 7065; Email: taitoh@affrc.go.jp