system, random overexpression of the *Arabidopsis* full-length cDNA is possible. Considering gene redundancy in rice and the availability of a large number of full-length cDNAs (Kikuchi et al 2003), we have begun producing FOX lines of rice in which 15,000 independent rice full-length cDNA are overexpressed under the control of the ubiquitin promoter (H. Ichikawa et al, unpublished results).

## References


Agrawal GK, Yamazaki M, Kobayashi M, Hirochika R, Miyao A, Hirochika H. 2001. Screening of the rice viviparous mutants generated by endogenous retrotransposon *Tos17* insertion: tagging of a zeaxanthin epoxidase gene and a novel *OsTATC* gene. Plant Physiol. 125:1248-1257.

Hirochika H. 2001. Contribution of the *Tos17* retrotransposon to rice functional genomics. Curr. Opin. Plant Biol. 4:118-122.

Hirochika H, Guiderdoni E, An G, Hsing Y, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, Sundaresan V, Leung H. 2004. Rice mutant resources for gene discovery. Plant Mol. Biol. 54:325-334.

Jeong DH, An S, Kang HG, Moon S, Han JJ, Park S, Lee HS, An K, An G. 2002. T-DNA insertional mutagenesis for activation tagging in rice. Plant Physiol. 130:1636-1644.

Kikuchi S et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 301:376-379.

Liljegren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF. 2000. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. Nature 404:766-770.

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, 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.

Sato Y, Sentoku N, Miura Y, Hirochika H, Kitano H, Matsuoka M. 1999. Loss-of-function mutations in the rice homeobox gene *OSH15* affect the architecture of internodes resulting in dwarf plants. EMBO J. 18:992-1002.

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner AR, Young K, Taylor NE, Henikoff JG, Comai L, Henikoff S. 2003. Large-scale discovery of induced point-mutations with high-throughput TILLING. Genome Res. 13:524-530.


## Notes


*Author's address:* Molecular Genetics Department, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan, e-mail: hirohiko@nias.affrc.go.jp.


# Toward genome-wide transcriptional analysis in rice using MAS oligonucleotide tiling-path microarrays


Lei Li, Xiangfeng Wang, Xueyong Li, Ning Su, Viktor Stolc, Bin Han, Jiayang Li, Yongbiao Xue, Jun Wang, and Xing Wang Deng



As the international efforts to completely sequence the rice genome are nearly completed, an immediate challenge and opportunity for the plant community is to comprehensively and accurately annotate the rice genome. This will lay a solid foundation for rice functional genomics and proteomics. We report here our current strategy to use integrative approaches centered on whole-genome nucleotide tiling-path microarray and array-guided full-length cDNA analysis for conducting genome-wide transcriptional analysis. Preliminary results suggest that these experimental approaches are useful in verifying predicted rice gene models and in identifying a significant number of new transcriptional units that were simply missed from current annotation. The pilot results further indicate that array data-guided RT-PCR cloning of full-length cDNA is emerging as an effective protocol to provide the much-needed experimental verification for those rice genes that lacked previous cDNA support. Thus, success of our effort described herein is likely to result in a more accurate annotation of all rice genes in the genome.


## Current status of rice genome annotation

Estimation of the total gene number in rice based upon the initial draft sequences of *Oryza. sativa* L. subsp. *japonica* and *indica* ranged widely from 30,000 to 60,000 (Goff et al 2002, Yu et al 2002). Finished sequences of chromosomes 1, 4, and 10 allowed a much fine-tuned estimation that placed the total gene number of rice between 57,000 and 62,500 (Feng et al 2002, Sasaki et al 2002, The Rice Chromosome 10 Sequencing Consortium 2003). Although the fast-accumulating finished sequences will ultimately come to a conclusion for the total gene number estimation, it should be noted that about half of those annotated gene models lack experimental support. Further, between one-third and one-half of the predicted genes appear to have no recognizable homology in *Arabidopsis* and features that exhibit striking deviations from experimentally verified genes in other species. Therefore, experimental efforts complementary to computer-based genome annotation are needed to verify predicted genes and to discover new genes in rice.

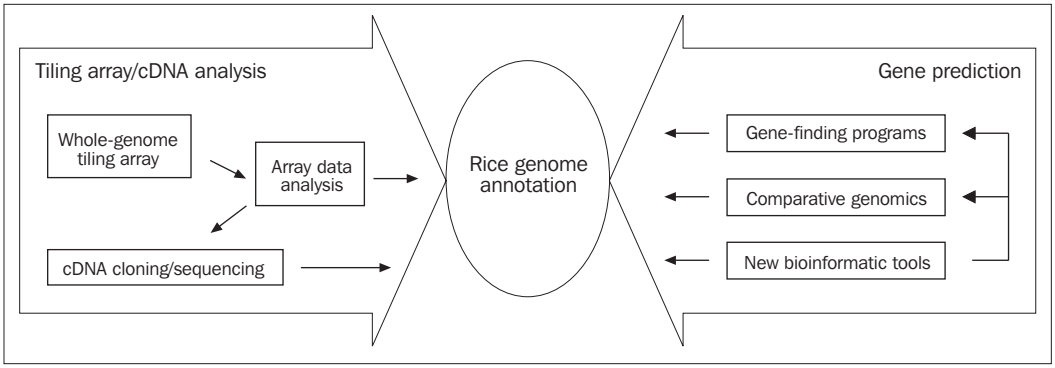


**Fig. 1. A proposed workflow for using tiling microarray to improve rice genome annotation. Current rice genome annotation efforts focus on *ab initio* gene prediction, comparative genomics, and other nonexperimental methods (block arrow on the right). Experiment-oriented approaches such as whole-genome tiling microarray and full-length cDNA analysis are emerging as complements to the computation-based methods (block arrow on the left). Arrowheads indicate the direction of information flow. See text for more details.**

## Rice whole-genome tiling-path microarray

One powerful and proven annotation approach complementary to *ab initio* gene prediction, comparative genomics, and other nonexperimental methods is whole-genome tiling-path microarray analysis coupled with array-guided cDNA cloning (Shoemaker et al 2001, Kapranov et al 2002, Yamada et al 2003). Figure 1 illustrates the strategy of our effort to enhance the rice genome annotation using an integrative approach centered on tiling microarrays. Essential to our approach is the design and development of custom tiling microarrays that represent the whole rice genome. Given the size of the rice genome (approx. 430 Mb) and the repetitive nature of our experiments, a suitable tiling microarray should offer high feature density, versatility of modification, and compatibility with our existing conventional microarray facility. These considerations led us to choose the Maskless Array Synthesizer (MAS) platform developed by NimbleGen (www.nimblegen.com, Stolc et al 2004).

In the initial phase, a minimum tiling strategy was designed to efficiently represent the genome with a minimum number of interrogating oligos employed that were strategically allocated to each potential gene such that expression of these genes can be unambiguously assayed. Implementation of this design required approximately 6.5 million pairs of interrogating oligos to tile essentially all the nonrepetitive sequences of the genome (Table 1), resulting in a resolution of a pair of 36-mer interrogating oligos every 60-bp genome sequence on average. Table 1 also shows the number of tiling oligo pairs for each of the 12 chromosomes. When synthesized at a density of 389,287 oligos per slide (Stolc et al 2004), the interrogating oligos that tile the japonica and indica genomes can be accommodated onto a set of 32 and 34 MAS arrays, respectively. It is expected that once the workflow of processing the minimum tiling arrays is streamlined, higher-resolution arrays will be designed and used to further our analysis.

**Table 1. Estimation of 36-mer oligo pairs for a minimum tiling of the rice genomes.**

| Chr. | Oligo pairs | |
|---|---|---|
| | japonica | indica |
| 1 | 758,477 | 843,677 |
| 2 | 638,666 | 678,379 |
| 3 | 618,937 | 747,025 |
| 4 | 563,693 | 597,322 |
| 5 | 456,837 | 536,721 |
| 6 | 526,611 | 551,022 |
| 7 | 487,411 | 478,130 |
| 8 | 469,369 | 514,189 |
| 9 | 367,182 | 378,664 |
| 10 | 375,141 | 419,408 |
| 11 | 424,379 | 404,151 |
| 12 | 440,214 | 390,756 |
| Total | 6,126,917 | 6,539,444 |

## A pilot tiling array hybridization experiment

To test the rice tiling array hybridization conditions, data acquisition, and analysis procedures, a pilot experiment was conducted in which two tiling arrays representing a portion of chromosome 10 were hybridized with mixed cDNA targets derived from four tissues: seedling root, seedling shoot, panicle, and suspension-cultured cells. Analysis of the array data detected expression of 77% of the reference gene models. Specifically, expression of 80.2% of the full-length cDNA matched genes (198 out of 247) and 70.9% (175 out of 247) of the genes without previous experimental support were detected. Thus, the cDNA-confirmed genes served as positive controls and demonstrated the sensitivity and feasibility of this approach. The relatively lower detection rate for the unsupported genes suggests that they are expressed at lower levels or restricted to specific cell types/developmental stages. Alterna-

tively, some of these predicted genes might be false models that do not exist *in vivo*.

## Integration of tiling array data to improve rice genome annotation

The hybridization data of the tiling microarrays can be used to further rice genome annotation (Fig. 1). For example, the hybridization data can provide support to or verification of predicted genes without prior experimental support, as detection of hybridization signals is strongly indicative of RNA synthesis directed by the genome segment represented by the interrogating oligos. Thus, the information generated from tiling arrays can be used to improve genome annotation by means of array-guided RT-PCR cloning and analysis of corresponding cDNAs. Likewise, the array data are also expected to reveal novel transcriptional units (i.e., those that were not included in any annotation but were detected by tiling arrays). In this regard, the array data could serve as a guidepost for cloning the corresponding cDNA by means of RT-PCR (Fig. 1). Compared with sequencing library-based full-length cDNA clones, which is considered the standard of gene annotation (Kikuchi et al 2003), array-assisted cDNA cloning and analysis will be targeted, and thus be more cost-effective and inclusive to cover the remaining portion of the rice genome that lacks corresponding expressed sequences.

## Compilation of reference genes for tiling array analysis

For the sake of simplicity, a set of BGI (Beijing Genome Institute) japonica gene models was used as the reference genes in the pilot trial (Stolc et al 2004). To effectively decode the tiling microarray hybridization data, however, reference genes are critically important, for which a comprehensive set of rice gene models is precisely anchored in the genomic sequences. A unique advantage in rice is the availability of multiple annotations such as the TIGR (The Institute for Genomic Research) japonica annotation (www.tigr.org/tdb/e2k1/osa1/) and the BGI indica and japonica annotations (http://rise.genomics.org.cn). While each annotation has its own forte and weakness, comparison and analysis of these annotations should yield a more comprehensive inventory of gene models to be evaluated by tiling arrays than any single annotation can offer. Reference gene compilation also takes advantage of the EST and full-length cDNA collections in the TIGR and the Kikuchi full-length cDNA data sets (http://cdna01.dna.affrc.go.jp/cDNA/). For example, the full-length cDNA-confirmed genes could naturally serve as positive controls to test array data and be used as training data to improve rice gene-finding algorithms (Fig. 1). Moreover, rice sequences have been subjected to extensive comparative genomic analysis with other cereals and other plant species. Closer and further inspection of those data should help to identify common and unique cereal genes and to integrate multiple annotations to provide a more comprehensive representation of the rice genome content (Fig. 1).

## Development of new bioinformatic tools to facilitate rice genome annotation

It has been realized from the pilot experiments that new computational tools need to be developed and validated to interpret tiling data and facilitate incorporation of tiling data to improve rice genome annotation. For example, algorithms to ascertain transcription need to be improved by incorporating more statistical parameters. Reliable tiling data in turn will provide better training data to improve gene prediction and other rice bioinformatic tools to enhance genome annotation (Fig. 1). The tiling data can be used in conjunction with several other public-funded rice genomics projects to maximize the detection of rice genes expressed at different developmental stages or under diverse environmental conditions. Consequently, improved genome annotation will also permit better array design to probe subtle transcriptional events such as alternative splicing, differential initiation and termination, etc. Finally, to present the tiling data in an accessible and informative form, an interactive database is required where the tiling data will be correlated with the complete genome sequence with all the annotated rice genes and other genomic features linked to cDNA/EST sequences and proteomic and structural information. Applying these approaches in rice should aid in the current genomic efforts to provide a complete and accurate rice genome annotation that holds the key to unravel the biology of the agriculturally important cereal crops.

The current research on japonica rice genome tiling array analysis in the laboratory of Xing Wang Deng at Yale University is supported by a grant from the National Science Foundation plant genome program (DBI-0421675). The collaborative effort in China was supported by the 863 rice functional genomics program from the Ministry of Science and Technology of China.

## References

Feng Q, Zhang YJ, Hao P, Wang SY, Fu G, Huang YC, Li Y, Zhu JJ, Liu YL, Hu X, et al. 2002. Sequence and analysis of rice chromosome 4. Nature 420:316-320.

Goff SA, Ricke D, Lan TH, Presting G, Wang RL, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, et al. 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100.

Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SPA, Gingeras TR. 2002. Large-scale transcriptional activity in chromosomes 21 and 22. Science 296:916-919.

Kikuchi S, Satoh K, Nagata T, Kawagashira N, Doi K, Kishimoto N, Yazaki J, Ishikawa M, Yamada H, Ooka H, et al. 2003. Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice. Science 300:1566-1569.

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu JZ, Niimura Y, Cheng ZK, Nagamura Y, et al. 2002. The genome sequence and structure of rice chromosome 1. Nature 420:312-316.

Shoemaker DD, Schadt EE, Armour CD, He YD, Garrett-Engele P, McDonagh PD, Loerch PM, Leonardson A, Lum PY, Cavet G, et al. 2001. Experimental annotation of the human genome using microarray technology. Nature 409:922-927.

Stolc V, Li L, Wang X, Li X, Su N, Tongprasit W, Han B, Xue Y, Li J, Snyder M, Gerstein M, Wang J, Deng X-W. 2004. Towards genome wide transcriptional unit analysis in rice using oligonucleotide tiling-path microarrays. Plant Mol. Biol. (In press.)

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

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, et al. 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. Science 302:842-846.

Yu J, Hu S, Wang J, Shu WG, 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 296:79-92.

## Notes

*Authors' addresses:* Lei Li, Xueyong Li, Ning Su, and Xing Wang Deng, Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520; Xiangfeng Wang, National Institute of Biological Sciences, Zhongguacun Life Science Park, Beijing 102206, China; Xiangfeng Wang and Jun Wang, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China; Xueyong Li and Yongbiao Xue, National Center of Crop Design, China Bioway Biotech Group Co., Ltd., Beijing 100085, China; Xiangfeng Wang and Ning Su, Peking-Yale Joint Center for Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China; Viktor Stolc, Genome Research Facility, NASA Ames Research Center, MS 239-11, Moffett Field, CA 94035; Bin Han, National Center for Gene Research, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 500 Caobao Road, Shanghai 200233, China; Jiayang Li, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Datun Road, Beijing 100101, China, e-mail: xingwang.deng@yale.edu.

# Candidate gene characterization at the *Pup1* locus: a major QTL increasing tolerance of phosphorus deficiency

Matthias Wissuwa, Kristy Gatdula, and Abdelbagi Ismail

Phosphorus (P) deficiency is a major yield-limiting factor for rice, particularly under upland or rainfed lowland conditions (Kirk et al 1998). P deficiency is frequently not due to low soil-P content in absolute terms but due to tight binding of soil-P in forms that are not readily available to the plant. The development of rice cultivars capable of using a higher portion of this fixed P already present in soils could be an attractive and cost-effective approach to increasing rice yields on P-deficient soils.

One promising step toward the development of more P-efficient cultivars was the identification of the *Pup1* locus in a QTL mapping population that was derived from rice cultivars Nipponbare (low P uptake) and Kasalath (high P uptake) (Wissuwa et al 2002). The effect of *Pup1* was confirmed using a near-isogenic line (NIL) carrying the positive *Pup1* allele from Kasalath in the Nipponbare genetic background. NIL-*Pup1* had three to four times the P uptake from a highly P-fixing volcanic ash soil than Nipponbare with equally high effects on biomass accumulation and grain yield (see Okada and Wissuwa, this volume). Further analysis suggested that roots of NIL-*Pup1* were slightly more efficient in extracting soil-bound P and that this advantage helped to maintain higher root growth rates, with additional benefits for P uptake (Wissuwa 2004).

## Candidate gene identification

Efforts at IRRI are now directed toward identifying the gene(s) at the *Pup1* locus. Further fine mapping of *Pup1* has advanced considerably and *Pup1* has now been mapped to a 195-kb interval spanning three BAC clones on chromosome 12 (Fig. 1). Gene annotation in the *Pup1* region identified 31 putative genes. Only one of those is a known gene, whereas some sequence similarities to known genes exist for an additional four genes. The remaining annotations were for hypothetical proteins (13) and transposable elements (13). None of the sequence similarities suggested an association with processes involved in P uptake or metabolism. This would suggest that *Pup1* is most likely a novel gene. It is also possible, however, that the gene is simply absent or highly distorted in Nipponbare, which would make annotations based on Nipponbare sequence data impossible.

Gene-specific primers have been developed for the putative genes located in the *Pup1* interval with the aim to investigate P-deficiency-induced expression patterns. Nipponbare and NIL-*Pup1* were grown in nutrient solution with three levels of P supply: excess P (50 µM), low P (1 µM), and zero P. RNA was isolated from roots and shoots of 5-week-old plants. RT-PCR performed on transcribed RNA samples identified two candidate genes that showed differential expression depending on genotype and P supply (Fig. 2). Candidate gene #1