Arabidopsis to Rice. Applying Knowledge from a Weed to Enhance Our Understanding of a Crop Species[1]

W.A. Rensink and C. Robin Buell*

The Institute for Genomic Research, Rockville, Maryland 20850

Although Arabidopsis is well established as the premiere model species in plant biology, rice (*Oryza sativa*) is moving up fast as the second-best model organism. In addition to the availability of large sets of genetic, molecular, and genomic resources, two features make rice attractive as a model species: it represents the taxonomically distinct monocots and is a crop species. Plant structural genomics was pioneered on a genome-scale in Arabidopsis and the lessons learned from these efforts were not lost on rice. Indeed, the sequence and annotation of the rice genome has been greatly accelerated by method improvements made in Arabidopsis. For example, the value of full-length cDNA clones and deep expressed sequence tag resources, obtained in Arabidopsis primarily after release of the complete genome, has been recognized by the rice genomics community. For rice >250,000 expressed sequence tags and 28,000 full-length cDNA sequences are available prior to the completion of the genome sequence. With respect to tools for Arabidopsis functional genomics, deep sequence-tagged lines, inexpensive spotted oligonucleotide arrays, and a near-complete whole genome Affymetrix array are publicly available. The development of similar functional genomics resources for rice is in progress that for the most part has been more streamlined based on lessons learned from Arabidopsis. Genomic resource development has been essential to set the stage for hypothesis-driven research, and Arabidopsis continues to provide paradigms for testing in rice to assess function across taxonomic divisions and in a crop species.

ARABIDOPSIS AND RICE STRUCTURAL GENOMICS

Access to a complete, finished genome for any organism provides the basis for large-scale exploration of biology. With respect to plants, Arabidopsis has secured the historical record of being the first plant genome to be sequenced (Arabidopsis Genome Initiative, 2000) with rice (*Oryza sativa*) coming in second (Goff et al., 2002; Yu et al., 2002). The Arabidopsis genome is essentially complete with the exception of few gaps primarily at the centromeres. However, the rice genome remains as a draft sequence until the projected completion in December 2004 by the public International Rice Genome Sequencing Project (IRGSP; http://rgp.dna.affrc.go.jp/IRGSP/).

The value of both organisms as model species for plant biology is further supported by the availability of not one genome sequence, but multiple genome sequences. For Arabidopsis, the public consortium sequenced to draft level the heavily utilized Columbia accession while a private company, Cereon, sequenced the second most utilized accession, Landsberg *erecta* (L*er*; Jander et al., 2002). For rice, there have been four genome sequencing efforts; three focused on the Nipponbare cultivar from the temperate subspecies *japonica* (Sasaki and Burr, 2000; Barry 2001; Goff et al.,

2002) and one on the 93-11 variety from the tropical subspecies *indica* (Yu et al., 2002). In addition to the nuclear genomes, both the chloroplast and mitochondrial genomes of Arabidopsis and rice are publicly available (Hiratsuka et al., 1989; Unseld et al., 1997; Sato et al., 1999; Notsu et al., 2002).

Both Arabidopsis and rice genome sequencing was preceded by expressed sequence tag (EST) sequencing as this provides not only an inexpensive sampling method for the expressed fraction of a genome, but also provides a quantitative profile of expression levels in specific tissues. ESTs also have utility as the cDNA clones themselves are valuable reagents for functional genomic studies. Currently, there are approximately 200,000 Arabidopsis and approximately 266,000 rice ESTs in the dbEST division of GenBank (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

ANNOTATION OF THE ARABIDOPSIS AND RICE GENOMES

Obtaining genomic sequence is the first step in a genomics-oriented approach to biology. Following sequencing, annotation, in which the genes and other features in the genome are curated, is essential to provide researchers with tools for biological research. However, most researchers perceive annotation as a straightforward step that can be completed quickly. In fact, annotation is not trivial and is a dynamic process, improving with time and effort as it is an iterative and on-going process. For Arabidopsis, the community was able to access manual annotation of each bacterial artificial chromosome (BAC) as it was

released to the public. When the genome was complete, the consortium created pseudomolecules (virtual contigs) for each of the five chromosomes along with the annotation of the entire genome (Arabidopsis Genome Initiative, 2000). Since 2000, the genome has been reannotated by The Institute for Genomic Research (TIGR; Haas et al., 2003; Wortman et al., 2003), resulting in refinement of the gene models, improvement of gene name assignments, and identification of new gene models. In comparison to the 25,498 genes described initially (Arabidopsis Genome Initiative, 2000), the latest release of the TIGR ATH1 genome (Release Version 5) has 26,207 predicted genes and 3,786 pseudogenes that include the transposable-element related gene models (C. Town, personal communication; http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml; Table I). Further annotation features such as gene ontologies, expression patterns, mutants, and literature curations are available for Arabidopsis, primarily at The Arabidopsis Information Resource (Rhee et al., 2003).

Compared to Arabidopsis, annotation of the rice genome is in its infancy. The IRGSP has developed a model similar to the Arabidopsis Genome Initiative with the annotation for finished BACs released as the sequence is deposited in public databases. However, the large size of the rice genome, coupled with the presence of large amounts of unfinished genome sequence data, has resulted in the need for whole genome annotation databases in which biologists can access and retrieve annotation for the entire genome, both for finished and unfinished BACs. In rice, the most current estimate of protein coding gene models is approximately 45,000 with another approximately 11,000 gene models encoding transposable element related genes (http://www.tigr.org/tdb/e2k1/osa1/

pseudomolecules/info.shtml). As with Arabidopsis, the quality of annotation data will continue to improve as the rice genome nears completion and dedicated annotation projects accelerate their activities. One advantage to rice genome annotators will be the ability to use Arabidopsis annotation as a reference and assign function to rice genes by comparison to Arabidopsis sequences. Indeed, for a large number of genes in the predicted rice proteome homologs can be found in Arabidopsis (Sasaki et al., 2002; Rice Chromosome 10 Sequencing Consortium, 2003). However, it should be noted that the simple sequence-based comparisons described to date between Arabidopsis and rice most likely are incomplete due to the unfinished rice genome sequence and/or the preliminary nature of rice genome annotation. For both Arabidopsis and rice, another layer of functional annotation can be obtained through comparison of knockout phenotypes, EST databases, and expression data sets although experimental validation will be required to confirm this putative annotation.

There are two methods for annotation: manual curation in which each gene model is inspected by a human annotator and automated annotation in which gene models and associated information are determined solely through computational methods. There are advantages and disadvantages to each method and either method, or a combination of both methods (semiautomated), can be appropriate depending on the genome size and the needs of the research community. One critical feature of genome annotation is the refinement of the gene model structure, i.e. intron-exon boundaries and untranslated regions. Annotators construct the gene model based on two main data types: output from ab initio gene finders and alignment of the sequences from various databases such as EST and

Table I. *Features of the Arabidopsis and rice genomes*

	Arabidopsis	Rice
Chromosome no.	5	12
Nuclear genome size	119 Mb	358 Mb (430 Mb)[a]
Number of genes	26,207 and 3,786[b]	Approximately 45,000 and approximately 11,000[c]
Chloroplast genome size	154,478 bp	134,525 bp
Number of genes	87 protein, 45 RNA genes	108 protein coding, 49 RNA genes
Mitochondrial genome size	366,924 bp	490,520 bp
Number of genes	117 protein, 33 RNA	56 protein, 25 RNA
No. ESTs	196,988[d]	266,004[d]

[a]The rice genome is estimated at 430 Mb (Arumuganathan and Earle, 1991) and the nonredundant sequence available to date from the IRGSP is 358 Mb (http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml). [b]Number of genes in Arabidopsis from Release 5 of the TIGR ATH1 database (http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml) is 26,207 protein coding genes and 3,786 pseudogenes, which includes the transposable-element related gene models. [c]Number of gene models in rice from Release 1 of the TIGR OSA1 database (http://www.tigr.org/tdb/e2k1/osa1/pseudomolecules/info.shtml) includes approximately 45,000 protein coding gene models and approximately 11,000 transposable element related gene models. [d]Number of ESTs from Release 011604 of the GenBank dbEST database (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html). NA, not available.

protein datasets. Often, there is conflicting information presented to the annotator who then has to make the best judgment with respect to the gene model structure. However, overriding ESTs, protein homology, and ab initio gene finder output for use in annotation are full-length cDNA sequences. Large collections of full-length cDNA sequences are available for both Arabidopsis and rice (Haas et al., 2002; Seki et al., 2002; Kikuchi et al., 2003), yet neither collection represents all the genes within the genome, presenting a challenge for annotators. A complementary approach to obtain gene structure information was used in Arabidopsis by hybridizing mRNA populations from various tissues to a set of high-density oligonucleotide arrays that span the entire genome (Yamada et al., 2003). This has proven to be a high throughput method to identify the ORFeome.

Regardless of rice or Arabidopsis, annotation is, and will continue to be, an iterative process. There are two main problems inherent to BAC-by-BAC or gene-by-gene manual annotation. First, each gene model or BAC is examined at one point in time and the gene models annotated early in the pipeline are stale or out-of-date compared to the gene models annotated at the end of the pipeline. Second, having the entire genome available at the time of annotation enables construction of paralogous families as all models for closely related genes are available and constructed simultaneously. Thus, annotating a large genome such as rice must utilize automated and semiautomated annotation methods in order to take advantage of the continually, newly available experimental evidence such as full-length cDNAs. In addition, annotating gene families rather than individual gene models, an approach adopted during the reannotation of the Arabidopsis genome, will be essential in rice.

COMPARATIVE GENOMICS

Plant biologists have rapidly incorporated comparative genomics into their research programs. With respect to Arabidopsis, comparison of the Columbia and L*er* genomes resulted in the availability of a high-density sequence-based polymorphism map, allowing positional cloning efforts to be greatly accelerated as the most common mapping population is Columbia × L*er* (Jander et al., 2002). In addition, it is becoming clear that natural variation between accessions is a valuable resource for identifying gene function. Analysis of natural variation allows for the detection of gene function for which no mutants can be isolated due to either a subtle phenotype or lethality of the mutation, the isolation of naturally occurring mutants, and the isolation of new alleles (Alonso-Blanco and Koornneef, 2000). As new genomic technologies become less expensive and higher through-put in nature, polymorphisms between accessions can be detected on a genome-wide scale, strongly accelerating this field of research (Borevitz and Nordborg, 2003). Another level of comparative genomics will involve the avail-

ability of whole shotgun sequence data from *Brassica oleracae* (http://www.tigr.org/tdb/e2k1/bog1/; http://nucleus.cshl.org/genseq/comp_genomics/index.html) that will allow for more evolutionary distant comparisons to be made.

As a member of the Poaceae family, rice is closely related to other cereals such as wheat, maize, barley, sorghum, oats, and sugarcane. Not only is there a high degree of conservation of phenotypic features across this family, synteny is conserved across the cereal genomes (for review, see Gale and Devos, 1998). With the availability of genome sequence for rice, researchers have been able to expand on synteny studies in the cereals from the macro scale reported previously to a more micro scale. Although it is clear that synteny between cereal genomes is not as absolute as previously reported, local regions of collinearity will be of immense use in positional cloning efforts in larger cereal genomes and in the identification of agronomic traits of interest. The availability of genomic sequence from two subspecies of rice provides a deep resource for evolution and adaptation studies. Clearly, the two subspecies are highly conserved (Feng et al., 2002; Yu et al., 2002) even though they are adapted to temperate (*japonica*) and tropical (*indica*) climates. However, the absence of complete genome sequence for at least one of the subspecies complicates highly detailed analyses and with the nearing completion of the *japonica* rice genome sequence, these avenues of investigation can be studied in more detail.

GENERATION OF GENOME-SCALE RESOURCES FOR FUNCTIONAL GENOMICS

The rationale for sequencing a plant genome is to obtain comprehensive information to understand plant biology with sequencing and annotation as the first steps in this process. Depending on an individual's background or the status of tools and/or reagents within a genome or community, functional genomics has many definitions. Here, we wish to be broad and consider any technique or approach that identifies gene function and/or the role of a gene in plant biology to be functional genomics. As entire books could be devoted to functional genomics in just rice or Arabidopsis, we will focus on large-scale resources and/or tools available to the respective research communities (Table II) that highlight the advances and status of functional genomics research. We will then highlight a few case studies in which Arabidopsis and rice have been directly compared and further our understanding of plant biology and the differential features of monocots and dicots.

REVERSE GENETICS

Reverse genetics has and will continue to be a powerful tool to identify gene function. In Arabidopsis, an

Table II. *Resources available for large scale functional genomics in Arabidopsis and rice*

	Arabidopsis	Rice
Reverse genetics	http://signal.salk.edu/cgi-bin/tdnaexpress	http://tos.nias.affrc.go.jp/~miyao/pub/tos17/
	http://tilling.fhcrc.org:9366/	http://www.pi.csiro.au/fgrttpub/
	http://www.nadii.com/pages/collaborations/ garlic_files/GarlicDescription.html (SAIL lines will be available through ABRC in May 2004)	
	http://enhancertraps.bio.upenn.edu/	
Molecular reagents		
ESTs	http://arabidopsis.org/servlets/Order?state= catalog	http://rgp.dna.affrc.go.jp/Cloneaccess.html
Full-length cDNA collection	http://pfgweb.gsc.riken.go.jp/projects/ raflcdna.html	http://cdna01.dna.affrc.go.jp/cDNA/
Publicly available gene expression profiling platforms		
Oligonucleotide arrays	http://www.ag.arizona.edu/microarray/	http://www.ricearray.org
Agilent arrays	http://www.chem.agilent.com/Scripts/ PDS.asp?lPage=9892	http://www.agilent.com/about/newsroom/ presrel/2003/06nov2003a.html
Affymetrix arrays	http://www.affymetrix.com/products/arrays/ specific/arab.affx	

exquisite set of tagged lines is available to the community. The most refined set is the collection of T-DNA tagged SALK lines available from the Ecker lab (Alonso et al., 2003). Within this collection are >225,000 tagged lines, approximately 88,000 of which have been sequenced revealing insertions into 21,700 genes. Clearly, the ability to search a sequence database for a mutant line in the gene of interest is a powerful tool for Arabidopsis researchers. Other collections of Arabidopsis lines have been developed and provide additional reservoirs for reverse genetics (Young et al., 2001; Sessions et al., 2002; Till et al., 2003). Similar reverse genetics approaches have been used for rice but are not as advanced. Similar to Arabidopsis, T-DNA tagged populations are being generated in rice and flanking sequences have been analyzed enabling database searches for tagged genes (An et al., 2003; Chen et al., 2003; Sha et al., 2004). In addition, the Ac/DS gene and enhancer trapping system for insertional mutagenesis has been applied to rice (Upadhyaya et al., 2002; Kolesnik et al., 2004). One feature of rice that differs from Arabidopsis is the abundant presence within the genome of retrotransposons. Exploiting the activation of the native Tos17 retrotransposon upon tissue culturing, Miyao et al. (2003) report generation of 47,196 Tos17 lines that collectively contain an estimated 500,000 insertions. Large scale sequencing of Tos17 insertion sites is under way, with a searchable database available for screening (http://tos.nias.affrc.go.jp/miyao/pub/tos17/).

EXPRESSION PROFILING PLATFORMS

One of the more information-rich functional genomic data types is provided by expression profiling through microarrays. Although the technology is still evolving, three platforms have established themselves

in the research community: spotted cDNA arrays, spotted oligonucleotide arrays, and direct synthesis of the oligonucleotides on the slide such as Affymetrix (http://www.affymetrix.com) and Agilent (http://www.chem.agilent.com). The Arabidopsis community quickly embraced gene expression profiling resulting in a substantial number of publications that have not only documented expression levels in multiple tissues, developmental stages, and stresses (see below) but also identified conserved promoter regions among coregulated genes (Hudson and Quail, 2003). Although all three expression profiling platforms are available in Arabidopsis (Table II), the spotted oligonucleotide array and Affymetrix chips are the methods of choice for gene expression profiling in Arabidopsis. In the last few years, only a limited number of research groups have had access to rice spotted cDNA and Affymetrix arrays, resulting in a minimal number of publications on expression profiling. Recently, a high-density array is now available to the public (http://www.agilent.com/about/newsroom/presrel/2003/06nov2003a.html) with a second array to be available in mid-2004 (http://www.ricearray.org). After learning from the Arabidopsis experiences, the oligonucleotide-based arrays are the preferred platform as these have proven to be cost-effective, provide high specificity, and avoid the problems of clone-tracking and PCR amplification inherent to the spotted cDNA or amplicon arrays.

CASE STUDIES

In the case studies below, we do not intend to provide a comprehensive review on the status of Arabidopsis and rice functional genomics. Instead, we selected recent publications that we feel illustrate the themes of research in these two model species.

Genome-Wide Comparison of Gene Families

One of the first and most informative studies that can be made upon the availability of genome sequence is a genome-wide comparison of gene families, both within and between species. We present four examples of gene families, P-Type ATPase ion pumps (Baxter et al., 2003), CONSTANS-like genes (Griffiths et al., 2003), cryptochromes (Matsumoto et al., 2003), and calcium-sensing gene families (Kolukisaoglu et al., 2004) that have been examined in both rice and Arabidopsis and shed light on conserved genes and pathway components in these two model species. With respect to P-type ATPase ion pumps, Arabidopsis has the highest number of genes (46) reported in any organism with rice having a similar number (43). Both species have representatives in all 5 major subfamilies of P-type ATPase ion pumps, indicating that monocots and dicots have evolved a similar large pool of P-type ATPAse ion pumps (Baxter et al., 2003). The CO (CONSTANS) gene of Arabidopsis has an important role in the regulation of flowering by photoperiod and is a member of a large gene family (17 members), whereas in rice only 16 gene family members could be identified. Although both species have similar gene numbers and contain family members from all 3 major CO-gene family classes, structural differences exist between the rice and Arabidopsis orthologs (Griffiths et al., 2003). Orthologs of the two Arabidopsis blue-light-receptor cryptochromes (AtCRY1 and AtCry2) are also present in rice (OsCRY1 and OsCry2). Cryptochromes regulate seedling deetiolation, entrainment of the circadian clock, and day length-sensitive timing of flowering and have similar functions in rice and Arabidopsis. Interestingly, cryptochromes have been detected in both the rice nucleus and the cytoplasm yet only in the Arabidopsis nucleus, suggesting additional functions for these proteins in rice (Matsumoto et al., 2003). In calcium signaling, a similar number of calcium-binding proteins such as calcineurin B-like proteins and their target kinases the calcineurin B-like-interacting protein kinases are present in Arabidopsis and rice, suggesting a similar level of complexity within this signaling network (Kolukisaoglu et al., 2004). Collectively, these examples demonstrate that comparative analyses can provide not only functional information regarding the number of genes in the reciprocal species but also identify conservation or divergence of pathways.

Flowering Pathways in Arabidopsis and Rice

Arabidopsis and rice differ greatly in flowering time as Arabidopsis is a long day plant whereas rice is a short day plant. In Arabidopsis, molecular and genetic approaches were used to identify the components and pathways that regulate floral induction (for review, see Mouradov et al., 2002). In rice, the components of the flowering pathway were identified by mapping quantitative trait loci for photoperiod sensi-

tivity (Yano et al., 2000; Takahashi et al., 2001). Comparative analyses with the rice genome sequence revealed that a majority of the Arabidopsis components are present in rice (Izawa et al., 2003). In addition, functional analyses in rice demonstrated that the key regulatory genes for flowering time are conserved between Arabidopsis and rice (Hayama et al., 2003). However, the function of a central transcriptional regulator is reversed between Arabidopsis and rice, demonstrating that distinct photoperiodic responses can be conferred by the same genetic pathway (for review, see Cremer and Coupland, 2003; Izawa et al., 2003). The discovery of a similar pathway in these two species that represents a major developmental difference illustrates that similar function cannot be inferred from sequence similarity alone and that functional studies involving knockout mutants, coupled with interchange of the signaling pathway components, will be needed.

Responses to Abiotic Stress

Abiotic stresses such as salinity, low temperature, and drought tolerance are important aspects of plant research as abiotic stress is a substantial limitation to further increases in global crop production. Once again, efforts in Arabidopsis have accelerated studies in rice. At the level of transcriptional control, the dehydration-responsive element binding/C-repeat transcription factors that control expression of many stress-inducible genes in Arabidopsis are present in rice (Dubouzet et al., 2003). These genes are also functionally interchangeable as overexpression of OsDREB1A in Arabidopsis resulted in the overexpression of a subset of the targets of the AtDERB1A transcription factor, indicating similar function in these two species (Dubouzet et al., 2003). The function of the rice OsMyb4 transcription factor that is inducible by low temperature was functionally characterized by overexpression in Arabidopsis (Vannini et al., 2004). In addition to comparative genomics and transgenic approaches, microarrays have been extensively used to identify genes involved in abiotic stress responses (Hazen et al., 2003). A comparison of expression data from rice microarrays to expression data of similar experiments in Arabidopsis revealed that of 73 genes differentially expressed in rice, 51 had been associated with stress responses in Arabidopsis (Rabbani et al., 2003). This indicates that there is a substantial degree of overlap that can be leveraged by researchers.

Plant Development

As representative of the monocots and dicots, rice and Arabidopsis have clear phenotypic differences, and comparison of developmental pathways between these two species will provide a foundation for understanding essential features of taxonomic differentiation in the angiosperms. The Arabidopsis

WUSCHEL and SCARECROW genes are involved in diversification of cell function and specification of cell fate. Orthologs have been identified in rice and function in similar processes (Kamiya et al., 2003a, 2003b). In rice, regulators of shoot branching, Lax Pinnacle (LAX) and Small Pinnacle (SPA), were identified from mutant populations (Komatsu et al., 2003). Although LAX encodes a basic helix-loop-helix transcription factor (bHLH) and a number of bHLH proteins can be found in Arabidopsis, the similarity of LAX in Arabidopsis is limited to the bHLH domain, illustrating the value of using rice mutants to identify features unique to grasses (Komatsu et al., 2003). Root development in Arabidopsis is well studied because of its simple architecture and the availability of molecular and genetic tools (for review, see Benfey and Scheres, 2000). As rice roots differ from Arabidopsis with respect to the anatomy of individual roots and the role of the embryonic root in development, Arabidopsis may not be a general model for root development (for review, see Hochholdinger et al., 2004). Current as well as pending functional genomics efforts should lead to the identification of rice mutants, which will aid in the identification of genes unique to rice development.

Microarray Data

Microarrays provide several layers of annotation for a genome. First, expression patterns can reveal potential functions for genes based on correlation of expression with phenotype. Second, expression profiles on a genome-wide scale enable the identification of co-regulated genes. Using clustering algorithms, genes with similar expression patterns can be grouped and inferences can be made with respect to function by extending annotation of known genes within these clusters to genes with no known function within the cluster. Third, regulatory motifs associated with co-regulated genes can be identified. This principle has been used to identify genes, as well as the underlying transcriptional network, of seed development in Arabidopsis (Girke et al., 2000; Ruuska et al., 2002). Grain filling in rice is expected to be different from Arabidopsis seed development due to the difference in seed structure, developmental process, and storage reserves. A multi-layered approach was used to not only identify grain filling genes in rice but also to determine regulatory features involved in transcriptional control. First, genes were selected based on computational approaches. Second, additional genes that had similar expression profiles during grain filing were identified from microarrays. Third, based on patterns of coregulation, regulatory motifs important in grain filling were identified (Zhu et al., 2003).

While not widely available or developed in either rice or Arabidopsis, proteomics has the potential to further identify similarities and differences between these two species as the integration of protein-protein interaction data with expression profiles can provide functional information. Such a study was performed in rice, and candidate genes in biotic stress in rice were proven to have a similar function in Arabidopsis (Cooper et al., 2003). Clearly, continued use of global expression profiling will result in generation of large datasets of functional data for rice and Arabidopsis. From the integration of these datasets, new types of annotation for these genomes will be generated.

CONCLUSIONS

Rice has clearly benefited from Arabidopsis research, both in the use of functional data and in research methodology. In parallel, Arabidopsis has benefited from the advances in rice genomics as rice currently is a robust platform for hypothesis testing. In addition, the availability of two model species, both with deep genomic resources, allows for comparative analyses and insight into evolution, adaptation, and differentiation within the angiosperms. For example, Arabidopsis and rice share a substantial number of orthologous genes, but the pathways and the underlying networks may function in an alternative fashion due to the lack of absolute 1:1 pairing of Arabidopsis-rice orthologs or an alternative function of the orthologs. Thus, an immediate challenge in rice genomics will be to provide functional data for gene family members that lack an ortholog in Arabidopsis. Simple sequence comparisons between Arabidopsis and rice can identify these similar genes, but at this time they may be incomplete due to by the unfinished rice genome sequence and preliminary nature of the rice genome annotation. It may also be that more advanced algorithms are needed to detect similarities with a high enough confidence for a gene to be called shared between Arabidopsis and rice. Thus, the completion of the rice genome sequence, the refinement of the annotation, and the integration of functional data will provide more accurate insights into how different (or similar) the two species are. For a broader understanding of plant gene function, it may be more interesting to focus on rice-specific genes (and Arabidopsis-specific genes) as these may better represent the fundamental differences between rice and Arabidopsis and, potentially, a monocot and dicot. The growing availability of mutant collections of rice and rice microarrays, coupled with the refinement of functional genomic tools in Arabidopsis, should accelerate the functional characterization of these genes within these two model species.

LITERATURE CITED

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana.* Science **301:** 653–657

Alonso-Blanco C, Koornneef M (2000) Naturally occurring variation in Arabidopsis: an underexploited resource for plant genetics. Trends Plant Sci **5:** 22–29

An S, Park S, Jeong DH, Lee DY, Kang HG, Yu JH, Hur J, Kim SR, Kim YH, Lee M, et al (2003) Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol **133**: 2040–2047

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408**: 796–815

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Report **9**: 208–218

Barry GF (2001) The use of the Monsanto draft rice genome sequence in research. Plant Physiol **125**: 1164–1165

Baxter I, Tchieu J, Sussman MR, Boutry M, Palmgren MG, Gribskov M, Harper JF, Axelsen KB (2003) Genomic comparison of P-type ATPase ion pumps in Arabidopsis and rice. Plant Physiol **132**: 618–628

Benfey PN, Scheres B (2000) Root development. Curr Biol **10**: R813–R815

Borevitz JO, Nordborg M (2003) The impact of genomics on the study of natural variation in Arabidopsis. Plant Physiol **132**: 718–725

Chen S, Jin W, Wang M, Zhang F, Zhou J, Jia Q, Wu Y, Liu F, Wu P (2003) Distribution and characterization of over 1000 T-DNA tags in rice genome. Plant J **36**: 105–113

Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S, Guimil S, Dunn M, Luginbuhl P, Ellero C, et al (2003) A network of rice genes associated with stress response and seed development. Proc Natl Acad Sci USA **100**: 4945–4950

Cremer F, Coupland G (2003) Distinct photoperiodic responses are conferred by the same genetic pathway in Arabidopsis and in rice. Trends Plant Sci **8**: 405–407

Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J **33**: 751–763

Feng Q, Zhang Y, Hao P, Wang S, Fu G, Huang Y, Li Y, Zhu J, Liu Y, Hu X, et al (2002) Sequence and analysis of rice chromosome 4. Nature **420**: 316–320

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA **95**: 1971–1974

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol **124**: 1570–1581

Goff SA, Ricke D, Lan TH, Presting G, Wang R, 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

Griffiths S, Dunford RP, Coupland G, Laurie DA (2003) The evolution of CONSTANS-like gene families in barley, rice, and Arabidopsis. Plant Physiol **131**: 1855–1867

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al (2003) Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res **31**: 5654–5666

Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL (2002) Full-length messenger RNA sequences greatly improve genome annotation. Genome Biol **3**: RESEARCH0029.1-0029.12

Hayama R, Yokoi S, Tamaki S, Yano M, Shimamoto K (2003) Adaptation of photoperiodic control pathways produces short-day flowering in rice. Nature **422**: 719–722

Hazen SP, Wu Y, Kreps JA (2003) Gene expression profiling of plant responses to abiotic stress. Funct Integr Genomics **3**: 105–111

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, et al (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet **217**: 185–194

Hochholdinger F, Park WJ, Sauer M, Woll K (2004) From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci **9**: 42–48

Hudson ME, Quail PH (2003) Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol **133**: 1605–1616

Izawa T, Takahashi Y, Yano M (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and Arabidopsis. Curr Opin Plant Biol **6**: 113–120

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002)

Arabidopsis map-based cloning in the post-genome era. Plant Physiol **129**: 440–450

Kamiya N, Itoh J, Morikami A, Nagato Y, Matsuoka M (2003a) The SCARECROW gene's role in asymmetric cell divisions in rice plants. Plant J **36**: 45–54

Kamiya N, Nagasaki H, Morikami A, Sato Y, Matsuoka M (2003b) Isolation and characterization of a rice WUSCHEL-type homeobox gene that is specifically expressed in the central cells of a quiescent center in the root apical meristem. Plant J **35**: 429–441

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 **301**: 376–379

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S (2004) Establishing an efficient Ac/Ds tagging system in rice: large-scale analysis of Ds flanking sequences. Plant J **37**: 301–314

Kolukisaoglu U, Weinl S, Blazevic D, Batistic O, Kudla J (2004) Calcium sensors and their interacting protein kinases: genomics of the Arabidopsis and rice CBL-CIPK signaling networks. Plant Physiol **134**: 43–58

Komatsu K, Maekawa M, Ujiie S, Satake Y, Furutani I, Okamoto H, Shimamoto K, Kyozuka J (2003) LAX and SPA: major regulators of shoot branching in rice. Proc Natl Acad Sci USA **100**: 11765–11770

Matsumoto N, Hirano T, Iwasaki T, Yamamoto N (2003) Functional analysis and intracellular localization of rice cryptochromes. Plant Physiol **133**: 1494–1503

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

Mouradov A, Cremer F, Coupland G (2002) Control of flowering time: interacting pathways as a basis for diversity. Plant Cell **14** (Suppl): S111–S130

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genomics **268**: 434–445

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol **133**: 1755–1767

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, et al (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Res **31**: 224–228

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

Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14**: 1191–1206

Sasaki T, Burr B (2000) International Rice Genome Sequencing Project: the effort to completely sequence the rice genome. Curr Opin Plant Biol **3**: 138–141

Sasaki T, Matsumoto T, Yamamoto K, Sakata K, Baba T, Katayose Y, Wu J, Niimura Y, Cheng Z, Nagamura Y, et al (2002) The genome sequence and structure of rice chromosome 1. Nature **420**: 312–316

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res **6**: 283–290

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, et al (2002) Functional annotation of a full-length Arabidopsis cDNA collection. Science **296**: 141–145

Sessions A, Burke E, Presting G, Aux G, McElver J, Patton D, Dietrich B, Ho P, Bacwaden J, Ko C, et al (2002) A high-throughput Arabidopsis reverse genetics system. Plant Cell **14**: 2985–2994

Sha Y, Li S, Pei Z, Luo L, Tian Y, He C (2004) Generation and flanking sequence analysis of a rice T-DNA tagged population. Theor Appl Genet **108**: 306–314

Takahashi Y, Shomura A, Sasaki T, Yano M (2001) Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the alpha subunit of protein kinase CK2. Proc Natl Acad Sci USA **98:** 7922–7927

Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, Young K, Taylor NE, et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res **13:** 524–530

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet **15:** 57–61

Upadhyaya NM, Zhou X, Zhu QH, Ramm K, Wu L, Eamens A, Sivakumar R, Kato T, Yun D, Santhoshkumar C, et al (2002) An iAC/Ds gene and enhancer trapping system for insertional mutagenisis in rice. Funct Plant Biol **29:** 547–559

Vannini C, Locatelli F, Bracale M, Magnani E, Marsoni M, Osnato M, Mattana M, Baldoni E, Coraggio I (2004) Overexpression of the rice Osmyb4 gene increases chilling and freezing tolerance of *Arabidopsis thaliana* plants. Plant J **37:** 115–127

Wortman JR, Haas BJ, Hannick LI, Smith RK Jr, Maiti R, Ronning CM,

Chan AP, Yu C, Ayele M, Whitelaw CA, et al (2003) Annotation of the Arabidopsis genome. Plant Physiol **132:** 461–468

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

Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, Baba T, Yamamoto K, Umehara Y, Nagamura Y, et al (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

Young JC, Krysan PJ, Sussman MR (2001) Efficient screening of Arabidopsis T-DNA insertion lines using degenerate primers. Plant Physiol **125:** 513–518

Yu J, Hu S, Wang J, Wong GK, 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

Zhu T, Budworth P, Chen W, Provart N, Chang HS, Guimil S, Su W, Estes B, Zou G, Wang X (2003) Transcriptional control of nutrient partitioning during rice grain filling. Plant Biotechnol J **1:** 59–70