1 Development and Application of Genomic Models for Large-Crop Plant Genomes

Robert M.D. Koebner and Rajeev K. Varshney

CONTENTS

1.1 INTRODUCTION

Plant genomes vary enormously in size. A part of this variation is generated by polyploidy, which is ubiquitous in the plant kingdom; however, even between closely related, ostensibly diploid species, it can still vary by an order of magnitude. A notable, but not atypical example is the contrast between rice (1 C DNA content of 0.50 pg, equivalent to 450 Mbp) and barley (5.55 pg, 5300 Mbp). The gene content of these two species is thought to be rather similar, numbering something under 40,000, depending on the gene prediction program employed [1]. Thus, much of the difference in DNA content is made up of nongenic DNA—in particular, retrotransposons.

When large-scale genome sequencing became possible in the 1990s, the large size of the majority of the leading crop genomes was technically and financially prohibitive. This prompted the plant research community to identify species (in

particular *Arabidopsis thaliana*) with more tractably sized genomes as genomic models. Technical improvements in the efficiency of sequencing achieved the finishing of the *Arabidopsis* genome by 2000 (4 years ahead of schedule) and the sequence was released with some fanfare in *Nature* [2].

At the time, *Arabidopsis* represented one of the first eukaryotes to be sequenced fully (along with *Saccharomyces cerevisiae*, human, and *Caenorhabditis elegans*). Its protein-encoding gene content has been estimated to be about 25,000 [2]. In the meantime, the genome sequence of *Arabidopsis* has been joined by those of a bewildering and ever growing list of eukaryotic and prokaryotic organisms numbering over 300 as of December 2005 (http://www.genomesonline.org). Of the 40 fully sequenced eukaryotic genomes, 25 belong to simple organisms (protozoans and fungi), 7 are vertebrates, 3 are insects, 2 are nematodes, and 3 are plants (of which 2 are the *indica* and *japonica* subspecies of rice).

The divergence of the monocot from the dicot clade is an ancient event, currently dated using molecular clock methods applied to the chloroplast genome at 140 to 150 MYA during the late Jurassic to early Cretaceous periods [3]. Independent estimates based on mitochondrial sequences have placed it somewhat earlier, at 170 to 235 MYA [4]. Dating of the time of speciation within each clade has been attempted by applying molecular clock methodology to repetitive sequences such as retrotransposons, but sequence homology in this class of element between clades is insufficient to use this method to date the monocot–dicot divergence.

Thus, it was recognized at an early stage that the *Arabidopsis* genome sequence would probably be of only partial relevance to monocot genomes. With a genome size about three times larger than that of *Arabidopsis*, rice was rapidly identified as the donor of a suitable model monocot genome. Before completion of the rice genome sequence, it became apparent that only a poor level of commonality in gene order existed between *Arabidopsis* and rice [5], thereby justifying post hoc the need for a separate model for the two major plant clades.

Nevertheless, the two genomes do retain some similarity as a result of common descent. Although some 85% of predicted *Arabidopsis* proteins were found to share significant homology with those of rice, about a tenth of them show a strong level of conservation [6]; in addition, most monocot–dicot homologues maintain exon order as expected. Perhaps most surprisingly, in many homologues, intron number, position, and even relative size show a remarkable level of conservation [7]. Despite the apparent disparity in gene number between the two models (25,000 vs. 40,000), it has recently been claimed that only a few hundred, or at most a few thousand, rice genes appear to lack close homologues in *Arabidopsis* [1].

The infrastructure and efficiency of whole genome sequencing is now at a point at which it has become much more realistic to undertake on a large scale. Current crop species targets include oat, *Brassica* spp., orange, coffee, barley, soybean, cotton, ryegrass, alfalfa, tomato, banana, bean, poplar, castor oil, sorghum, and maize. A growing number of other species has been targeted for sequencing of the gene space (ESTs or similar). If these trends continue, it is likely that within 10 years, most of the major crop genomes will have been fully sequenced. In the meantime, species that are nodal in crop phylogenies may be chosen to serve to

generate a network of submodels; a particular example of this lies behind the current proposal to sequence the grass *Brachypodium distachyon.*

This chapter attempts to take stock of model genomes' contribution to understanding of the genomes of crop species to date. Perhaps other contributors to this volume will show the lasting value that model species biology has made to crop improvement.

1.1.1 Dicot Models

1.1.1.1 *Arabidopsis thaliana* (Thale Cress)

Arabidopsis is by far the most well developed of the crop plant models. In addition to its completed genome sequence, it is easily transformable and enjoys a huge range of genetic (mutants, mapping populations, ecotypes) and genomic (cloned genes, libraries, arrays, markers, etc.) resources and an ever expanding database relating phenotype to genotype. The closest crop relatives to *Arabidopsis* are the three diploid *Brassica* species *rapa*, *nigra*, and *oleracea* that carry, respectively, the A, B, and C genomes as described in reference 8. Although all of these represent rather minor crop species, the major contributor of *Brassica* spp. to agriculture is *B. napus* (oilseed rape or canola), which is an AC allotetraploid formed from the combination *B. oleracea* × *B. rapa.*

The lineages of *Arabidopsis* and *Brassica* are thought to have diverged from one another between 14 and 20 MYA [4]; this divergence has included a number of distinct polyploidization events because the present-day diploid *Brassica* spp. carry multiple paralogueues of chromosomal segments collinear with the *Arabidopsis* genome. This copy number is most commonly three, so the inference is that the diploids must have evolved from a hexaploid ancestor [9,10]. Copy number is frequently less than three, varying in 4× *B. napus* from four to seven [10]. Within the triplicated paralogues, a common pattern of interspersed gene loss is emerging, with the result that each paralogue typically carries a slightly different spectrum of the full gene set presumably present on the progenitor segment [11].

AU: confusing. You say "frequently less than three" and then "varying from four to seven" (which are >3.

A further complication is that *Arabidopsis*, as revealed from its genome sequence, is a cryptic polyploid, carrying a sufficient number of large segmental duplications for an evolutionary history of at least four different large-scale duplication events to have been proposed [12]. Overall, an estimated 74 translocations, fusions, deletions, or inversions separate the genomes of *Arabidopsis* and *B. napus* [10], of which about one half are common to A and C genomes in present-day oilseed rape.

1.1.1.2 *Lotus japonicus* (Trefoil) and *Medicago truncatula* (Barrel Medic)

The Fabaceae, one of the largest families of flowering plants with 650 genera and over 18,000 species, is distinguished from other dicot families by its symbiotic relationship with nitrogen-fixing *Rhizobium*. The economic and nutritional importance of nitrogen fixation has been sufficient to justify targeting a model representative, and two competitive species are currently being pursued. *Medicago truncatula*

has some importance in its own right as a forage crop in Australia. It has a small diploid genome (1 C DNA 0.48 pg) and a rapid generation time, and it is self-fertile, transformable, and a prolific seed producer. *Lotus japonicus* is a short-life-cycle, perennial wild legume that also has a small genome size (1 C DNA 0.48 pg).

The genomes of both species are currently being sequenced (see, respectively, http://www.medicago.org and http://www.kazusa.or.jp/lotus/index.html). The two sequences show a high degree of similarity to one another [13]. Collinearity between *M. truncatula* and pea at the level of coarse genetic maps appears to be encouragingly high [14], although there is significant sequence divergence between those of *Lotus* and the major legume crop species soybean [13]. In a computational approach, *Lotus*, *Medicago*, and *Glycine* unigenes were BLASTed against nonlegume unigene sets and the rice genome sequence to define legume-specific gene motifs; this delivered some 2500 such contigs, of which less than 3% showed any homology to any previously identified legume genes [15]. Such results underline the utility of a model legume to define sequences specific to this group of agriculturally important crop species.

1.1.1.3 *Populus trichocarpa* (Poplar or Black Cottonwood)

Conventional genetic approaches in trees are limited by the large size, long generation interval, and outcrossing mating system of most species. The need for a tree model reflects the importance of many traits that are not shared by an herbaceous annual plant such as *Arabidopsis*. Important among these are wood formation, longevity, seasonal growth, and hardiness. The genus *Populus* consists of 30 to 40 species, 4 of which have significant commercial importance. Selection and hybridization programs in poplars began in North America in the 1960s, and the most commonly exploited crosses have involved *P. trichocarpa*, *P. deltoides*, *P. nigra*, *P. grandidentata*, *P. alba*, *P. tremuloides*, and *P. tremula.*

Because the genomic resources of *P. trichocarpa* were the most developed at the time that genome sequencing was proposed, this species became the accepted tree model. It was chosen as the first tree for genome sequencing largely because of its modest genome size (0.6 pg)—about 40 times smaller than that of pine, the most important of all forestry species. It also has a number of other advantages over potential alternative tree species specifically related to its rapid juvenile growth, which allows for phenotypic assessments to be made relatively quickly; its well established transformation and regeneration protocols; and the pre-existence of a body of genetic mapping, which includes placement and tagging of a number of quantitative trait loci (QTL). The final draft sequence was scheduled for release in early 2005, but is still awaited at the time of writing. Current status is updated on http://genome.jgi-psf.org/Poptr1/Poptr1.home.html.

1.1.2 Monocot Models

1.1.2.1 *Oryza sativa* (Rice)

Rice is the pre-eminent monocot model and is uniquely a model and a crop. The particular importance of this duality lies in the much greater potential that this allows

for transferring phenotype, as well as genotype, from model to crop. Rice is a tropical species and thus more likely to share pathogens and/or abiotic stresses with its tropical crop relatives such as the millets (and, to a lesser extent, maize and sorghum) than with its important temperate small-grain and pasture-grass relatives (wheat, barley, rye, oat, and ryegrass). Nevertheless, shared morphology and crop architecture among all cereal species do allow many phenotypic connections to be made. The dicot models, in contrast, are far removed in their crop morphology, making such transfers much less predictive.

The grasses belong to the Poaceae, which evolved from a common ancestor some 50 to 60 MYA; together, they provide an estimated 60% of global human calorific intake. The family includes at least 10,000 species, classified into 650 genera [16,17]. The crop species within the family fall into the three subfamilies Pooideae (which includes the temperate cereals and ryegrass), Panicoideae (maize, sorghum, millets, sugar cane) and Bambusoideae (rice). Until the development of generic DNA technology, primarily in the 1990s, genetic research in each grass crop was conducted in isolation from that in the others. Before this time there was no secure way of verifying what had already been suspected for some time: that because these species were related by (albeit distant) descent, they were likely to share genetic content and, at least at a basic level, genetic mechanisms.

The first demonstration of what is now referred to as "comparative genetics" was carried out in the Solanaceae, where common RFLP linkage relationships in tomato and potato were uncovered using DNA probes developed from a tomato template [18]. The concept spread quickly to the Poaceae, and numerous cross-species comparisons began to appear in the literature during the early to mid 1990s [19–21]. These led to the construction of partial consensus maps linking maize with sorghum [22] and wheat with barley and rye [23]. A synthesis of these maps was generated by relating them all to that of the rice genome [24]. The concept of "synteny" elaborated by these cross-species comparisons of gene order reflects conservation over evolutionary time at the macroscale. Whether this was extendable to the microscale was questionable, given the large variation in genome size between individual Poaceae species.

The outcome of sequence-based comparisons in selective syntenic regions is that although gene structure and sequence is extremely well conserved between taxa, intergenic regions are highly divergent, even at the level of genotypes within a taxon [25]. Much of this intra- and interspecific divergence is generated by retroelement activity and, in particular, helitron-like transposons composed of multiple gene-derived fragments [26]. In addition, the increasing body of evidence generated from large-scale sequence comparisons between related taxa demonstrates how synteny is also disturbed by the presence of species-specific localized duplications and other forms of genome reorganization [27–29].

By the end of the 1990s, with the *Arabidopsis* genome project already well underway, rice became an increasingly attractive candidate for whole genome sequencing [30] in the private and public sectors. These efforts were combined to produce almost full genomic sequences of *japonica* and *indica* subspecies [6,31,32], along with a near complete compendium of full-length cDNA sequence [33]. The finished sequence currently covers about 95% of the genome, including most euchromatic

regions and 2 (out of 12) complete centromeres. Mirroring the situation in *Arabidopsis,* the genome sequence has revealed a history of polyploidization in the evolution of modern day rice, with about half of the gene content duplicated as paralogues.

1.1.2.2 *Brachypodium* spp. (False Bromes)

The false bromes are a group of noncultivated grasses, mostly regarded in agriculture as weeds rather than as beneficial plants. The perennial *B. sylvaticum* (slender false brome) and the annual *B. distachyon* (purple false brome) have been suggested as intermediate models for the temperate cereals. *B. distachyon* has been thought to have a genome size indistinguishable from that of *Arabidopsis* [34], but measurement of 1 C DNA content suggested that it is three times larger (0.36 pg; http://www.rbgkew.org.uk/cvalues/). The genome size of *B. sylvaticum* is slightly higher still (0.48 pg), but both genomes are smaller than that of rice.

The value of both as genomic models for the temperate grain cereals lies in their membership within the Pooideae clade and hence their much closer relationship to wheat, barley, rye, and oats than rice enjoys. The significance of this relationship has been confirmed in two recent positional cloning projects, one in wheat [35] and the other in barley [36]. The quality of probe hybridization to and prediction of overall gene content in the target were superior in *Brachypodium* to that offered by rice [37]. Although *B. sylvaticum* has been proposed to date only as a genomic and not a biological model, *B. distachyon* does have a number of generic advantages as a functional genomic and biological model (self-fertility, in-breeding habit, short life cycle, small size [approximately 20 cm at maturity], lack of seed-head shatter, and undemanding growth requirements) [34]. At the time of writing, there is a concerted effort to develop *B. distachyon* as a fully functional genomic model.

1.2 HARNESSING MODEL GENOMES FOR CROP GENETICS AND IMPROVEMENT

The impact of model genomes on crop species has been felt mainly in their delivery of a strategy for gene isolation in the large genome crop species. This strategy relies on the maintenance of synteny, assuming that gene content in the model in a specific genomic region is more or less conserved in the target crop genome. The model-to-crop paradigm follows a combination of:

Mapping a trait to a defined genetic interval in the crop
Identifying the corresponding genomic region in the model via the use of common genic markers (because it is substantially only the gene content, not the nongenic, largely Retrotransposon-containing, repetitive content that is conserved across clades)
Identifying a potential candidate sequence in the model on the basis of a relationship between predicted gene function (derived from the annotation of the model genome) and the target trait
Validating the crop homolog of the candidate, demonstrated by allelic association and/or mutation complementation

The first major success of the model-to-crop genomic approach in the monocots came with the isolation of the "green revolution" wheat semidwarfing genes *Rht-B1* and *Rht-D1* [38]. Together, these two genes have been responsible for probably the most far-reaching and widespread change in the appearance of any crop worldwide. Their incorporation into the breeding pool has generated shorter plants that enjoy an enhanced grain yield potential, thanks to the consequent increase in harvest index, and are responsive to higher application rates of fertilizer without becoming liable to straw collapse.

The isolation of these genes predated the availability of the full rice or *Arabidopsis* genome sequence, but nevertheless relied heavily on genomic information from both model species. Critical to the success of their cloning was that the physiological nature of the semidwarf variants of wheat was similar to that of previously characterized mutants in maize and *Arabidopsis*. This allowed an approach whereby the rice orthologue of the *Arabidopsis gai* gene was identified from a rice EST collection. When this rice sequence hybridized to wheat DNA at the genomic locations of the *Rht-1* genes, the rice probe was exploited to extract the full genomic sequence of both of the wheat genes. Thereafter, the sequence and functional basis of these important semidwarf alleles were readily obtained.

Finishing the genome sequences of the models enabled the model-to-crop paradigm to be tested. A textbook illustration was provided by the recent successful cloning of the barley gene *Ppd-H1*, the major determinant of flowering time under long photoperiods [36]. Unlike the situation with *Rht*-1, the physiological model provided by *Arabidopsis* was not informative because the candidate genes provided by *Arabidopsis* did not map to the genomic location of the barley gene target. Thus, the initial step was to fine-map *Ppd-H1* in a conventional cross between parents carrying contrasting alleles, and the linked markers thereby derived then allowed for construction of a physical contig based on the presence of key marker loci on barley BACs. The gene content of the homologous region in *Brachypodium sylvaticum* helped to define the matching region in rice, and the critical barley recombinants finally identified a region in the homologous rice segment that contained only a single candidate sequence.

AU: sp. out for first use.

This rice gene, *Os-PRR*, shares significant sequence homology with *Arabidopsis At-PRR7*, which, when mutated, leads to delayed flowering under long day conditions, just as the inactive form of *Ppd-H1* does in barley. *Ppd-H1* and *At-PRR* also share temporal patterns of expression. Finally, resequencing of the critical parts of *Hv-PRR* across varieties of known allelic status at *Ppd-H1* was able to demonstrate a correlation between a functional glycine to tryptophan change in a domain of the gene that is well conserved across taxa.

A more elaborate, but essentially equivalent strategy was used to isolate the wheat gene responsible for determination of winter habit (vernalization requirement) [39]. Once again, a large mapping population, this time in the diploid wheat *Triticum monococcum*, was used to delineate a genetic interval of <0.1 cM containing the target. Sequencing of the 324 kb represented by this segment identified two genes, with no additional candidates present in the homologous segments of rice or sorghum. Both candidate genes had *Arabidopsis* homologues, but only one of them, *AP1*, is required for the transition between vegetative and reproductive phases in

Arabidopsis; the other is a floral meristem identity gene. The association between sequence variation at *Tm-AP1* and phenotype was established by demonstration of three independent deletions distinguishing the promoter sequence of spring from winter accessions.

The most recent example of positional cloning in a monocot crop that has relied on the availability of model genomes is the isolation of the *Ph1* locus in wheat [35]. This "gene" is responsible for the diploid-like inheritance of hexaploid wheat, and its isolation was hampered at the outset by a lack of any verifiable allelic variation. Because of this, it was not possible to generate a fine-scale genetic map as a first step to defining the target genomic region. Instead, a series of overlapping deletion mutants was generated, and phenotype (loss of diploid-like chromosome pairing) was related with loss of genic markers in the *Ph1* region, which had been derived from synteny comparisons between wheat and rice and/or *Brachypodium*.

As a result, the number of genes present in the smallest genetic interval defined was over 30, and because the effect of *Ph1* is specific to polyploids, there were no sensible leads derived from the predicted function of any of these candidates. To progress beyond this point, it was necessary to sequence a substantial tract of wheat DNA directly; the identity of the locus was finally determined through an internal comparison among the individual A, B, and D genome segments.

A reasonable level of synteny between *Arabidopsis* and *Brassica* exists, the complications of segmental duplication notwithstanding [40], and the finished *Arabidopsis* sequence has been available for longer than that of rice; however, gene isolation in *Brassica* has relied more on functional homology than on positional cloning. Thus, having established function of a gene in *Arabidopsis*, primarily by mutation/complementation, homologues in *Brassica* have been extracted from genomic or cDNA libraries and function in *Brassica* established by associating variation in phenotype with polymorphism at the RFLP or sequence level. Beyond the *Brassica* spp., high rates of sequence divergence have greatly inhibited success of orthologous cDNAs as hybridization probes against genomic DNA and restricted the applicability of the model to its immediate relatives.

1.3 PERSPECTIVE

The value of a small number of model plants in a strictly genomic context is probably ephemeral. This is primarily because large genomes are increasingly considered practical to sequence on cost or technical grounds. Within 10 years, it is likely that most of the major crops will have been sequenced, at least with respect to their gene space. At the same time, comparative genomics is showing that although gene order at the macroscale is well conserved over large taxonomic distances, the microsynteny necessary to predict sequence across species (and even, to a surprising extent, within species [25]) at the microscale is insufficient for a small number of models to be able to serve many diverse crop species. The cereals are exceptional in this respect, in that so many cereal crop species are clustered within a narrow taxonomic clade, but even for these, the models have their limitations.

The more lasting value of models will surely lie in the insights into plant biology that they will allow. Some of these will include the rapidly developing fields of

epigenetic and micro-RNA-directed gene regulation, where *Arabidopsis* is already serving as a model organism for species well beyond the plant kingdom [41,42]. Many of the more specifically plant-orientated areas of biology informed by model species are covered by other contributions to this volume.

REFERENCES

1. Bennetzen, J.L. et al., Consistent overestimation of gene number in complex plant genomes, *Curr. Opin. Plant Biol*., 7, 732, 2004.
2. *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature* 408, 796, 2000.
3. Chaw, S.M. et al., Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes, *J. Mol. Evol.*, 58, 424, 2004.
4. Yang, Y.W. et al., Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages, *J. Mol. Evol.*, 48, 597, 1999.
5. Liu, H., Sachidanandam, R., and Stein, L., Comparative genomics between rice and *Arabidopsis* shows scant collinearity in gene order, *Genome Res.*, 11, 2020, 2001.
6. Goff, S.A. et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), *Science*, 296, 92, 2002.
7. Carels, N. and Bernardi, G., Two classes of genes in plants, *Genetics*, 154, 1819, 2000.
8. U. N., Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization, *Jpn. J. Bot.*, 7, 389, 1935.
9. Lysak, M.A. et al., Chromosome triplication found across the tribe Brassiceae, *Genome Res.*, 15, 516, 2005.
10. Parkin, I.A.P. et al., Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*, *Genetics*, 171, 765, 2005.
11. Rana, D. et al., Conservation of the microstructure of genome segments in *Brassica napus* and its diploid relatives, *Plant J*., 40, 725, 2004.
12. Vision, T.J., Brown, D.G., and Tanksley, S.D., The origins of genomic duplications in *Arabidopsis*, *Science*, 290, 2114, 2000.
13. Choi, H.K. et al., Estimating genome conservation between crop and model legume species, *Proc. Nat. Acad. Sci. USA*, 101, 15289, 2004.
14. Kalo, P. et al., Comparative mapping between *Medicago sativa* and *Pisum sativum*, *Mol. Genet. Genomics*, 272, 235, 2004.
15. Graham, M.A. et al., Computational identification and characterization of novel genes from legumes, *Plant Physiol*., 135, 1179, 2004.
16. Bennetzen, J.L. and Freeling, M., Grasses as a single genetic system—genome composition, collinearity and compatibility, *Trends Genet.*, 9, 259, 1993.
17. Kellogg, E.A., Relationships of cereal crops and other grasses, *Proc. Nat. Acad. Sci. USA*, 95, 2005, 1998.
18. Bonierbale, M.W., Plaisted, R.L., and Tanksley, S.D., RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato, *Genetics*, 120, 1095, 1988.
19. Ahn, S. and Tanksley, S.D., Comparative linkage maps of the rice and maize genomes, *Proc. Nat. Acad. Sci. USA*, 90, 7980, 1993.
20. Kurata, N. et al., Conservation of genome structure between rice and wheat, *Bio-Technology*, 12, 276, 1994.

AU: need complete name.

21. Devos, K.M. et al., Chromosomal rearrangements in the rye genome relative to that of wheat, *Theor. Appl. Genet.*, 85, 673, 1993.
22. Dufour, P. et al., Comparative genetic mapping between duplicated segments on maize chromosomes 3 and 8 and homoeologous regions in sorghum and sugarcane, *Theor. Appl. Genet.*, 92, 1024, 1996.
23. VanDeynze, A.E. et al., Molecular-genetic maps for group-1 chromosomes of Triticeae species and their relation to chromosomes in rice and oat, *Genome*, 38, 45, 1995.
24. Moore, G. et al., Cereal genome evolution—grasses, line up and form a circle, *Curr. Biol.*, 5, 737, 1995.
25. Brunner, S. et al., Evolution of DNA sequence nonhomologies among maize inbreds, *Plant Cell*, 17, 343, 2005.
26. Morgante, M. et al., Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize, *Nat. Genet.*, 37, 997, 2005.
27. Salse, J. et al. New *in silico* insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome, *Plant J.*, 38, 396, 2004.
28. Sorrells, M.E. et al., Comparative DNA sequence analysis of wheat and rice genomes, *Genome Res.*, 13, 1818, 2003.
29. Klein, P.E. et al., Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement, *Plant J.*, 34, 605, 2003.
30. Goff, S.A., Rice as a model for cereal genomics, *Curr. Opin. Plant Biol.*, 2, 86, 1999.
31. Yu, J. et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*), *Science*, 296, 79, 2002.
32. International Rice Genome Sequencing Project, The map-based sequence of the rice genome, *Nature*, 436, 793, 2005.
33. Rice Full-Length cDNA Consortium, Collection, mapping, and annotation of over 28,000 cDNA clones from *japonica* rice, *Science*, 301, 376, 2003.
34. Draper, J. et al., *Brachypodium distachyon*. A new model system for functional genomics in grasses, *Plant Physiol.*, 127, 1539, 2001.
35. Griffiths, S. et al., Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat, *Nature*, in press (doi: 10.1038), 2006.
36. Turner, A. et al., The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley, *Science*, 310, 1031, 2005.
37. Foote, T.N. et al., Construction and analysis of a BAC library in the grass *Brachypodium sylvaticum*: its use as a tool to bridge the gap between rice and wheat in elucidating gene content, *Funct. Integr. Genomics*, 4, 26, 2004.
38. Peng, J.R. et al., "Green revolution" genes encode mutant gibberellin response modulators, *Nature*, 400, 256, 1999.
39. Yan, L. et al., Positional cloning of the wheat vernalization gene *VRN1*, *Proc. Nat. Acad. Sci. USA*, 100, 6263, 2003.
40. Lukens, L. et al., Comparison of a *Brassica oleracea* genetic map with the genome of *Arabidopsis thaliana*, *Genetics*, 164, 359, 2003.
41. Martienssen, R.A., Doerge, R.W., and Colot, V., Epigenomic mapping in *Arabidopsis* using tiling microarrays, *Chromosome Res.*, 13, 299, 2005.
42. Lippman, Z. et al., Role of transposable elements in heterochromatin and epigenetic control, *Nature*, 430, 471, 2004.

AU: update?