Twenty-First Century Plant Biology: Impacts of the Arabidopsis Genome on Plant Biology and Agriculture

C. Robin Buell and Robert L. Last*

Department of Plant Biology (C.R.B., R.L.L.) and Department of Biochemistry and Molecular Biology (R.L.L.), Michigan State University, East Lansing, Michigan 48824–1319

HOW THE SEQUENCE PROPELLED ARABIDOPSIS RESEARCH

In the late 1980s and early 1990s, Arabidopsis (*Arabidopsis thaliana*) rapidly emerged as a model system for plant biology due to the ease of genetic approaches to study development, physiology, and gene function coupled with a highly collaborative community. Arabidopsis's position as the premiere model plant system was cemented in the early 1990s when the EST project commenced at Michigan State University (Newman et al., 1994). This was the first such effort in plants and it seeded production of large-scale sequence and functional genomics resources for Arabidopsis that to this day are unmatched in any other plant species. While the EST project propelled Arabidopsis research into the genomics era, it was the foresight of leaders in the plant biology community to "...identify all of the genes by any means, and to determine the complete sequence of the Arabidopsis genome before the year 2000" (http://www.arabidopsis.org/portals/masc/Long_range_plan_1990.pdf). The resulting coordinated international genome sequence effort, the Arabidopsis Genome Initiative, laid the foundations for transition of Arabidopsis from a genetic to a genomic model system. This culminated in the publication of the high-quality genome sequence of the Columbia (Col-0) accession (Arabidopsis Genome Initiative, 2000). While there are gaps in highly repetitive low-complexity regions of what is now termed the reference genome, the quality of the Col-0 genome sequence is, and for the foreseeable future will remain, the highest quality plant genome sequence. When coupled with its manually curated annotation, the sequence data makes Arabidopsis the gold standard for all plant genomes.

Starting in 1998, Cereon Genomics LLC (a wholly owned subsidiary of Monsanto Co.) produced a low-depth shotgun sequence of the Landsberg *erecta* genome using Sanger sequencing technology (Jander et al., 2002), providing an early look at the genes of a flowering plant. For the community, access to both the Col-0 and Landsberg *erecta* genome sequences revealed an unprecedented number of polymorphisms between these landraces, enabling more than 100 published map-based cloning studies. To our knowledge, it also provided the community with the first example of how high-throughput sequencing of an accession or cultivar related to a high-quality reference sequence makes the initial sequence even more valuable (Rounsley and Last, 2010).

Next-generation sequencing technologies permit high-throughput sequencing at greatly reduced costs and sequencing of accessions is now feasible at an enormous scale. The plan to analyze genome-wide sequence variation in 1,001 accessions of Arabidopsis (Weigel and Mott, 2009) is well under way with nearly 100 completed genomes as of early 2010 (http://www.1001genomes.org/). This pioneering effort will result in a deep understanding of genetic variability within a species and permit analysis of genome evolution and population biology. This illustrates how Arabidopsis research continues to be at the forefront of defining approaches that will revolutionize our understanding of economically important plant processes and plant species.

Gene expression states are affected by DNA and chromatin protein modification. While standard genome DNA sequence alone cannot directly report these decorations, these epigenetic states can be assessed at the single nucleotide level using next-generation sequencing approaches, and descriptive studies on the Arabidopsis epigenome are emerging (Lister et al., 2008). This information enables studies of biological problems that are unique to plants such as epigenetic imprinting in the triploid endosperm and the diploid embryo of the seed. Examination of methylation states revealed imprinting arose through targeted methylation of genic regions due to nearby transposable element insertion (Gehring et al., 2009). While the Arabidopsis genome has relatively few transposable elements, other genomes such as maize (*Zea mays*) and wheat (*Triticum aestivum*) have large numbers of these and other repetitive sequences and it will be fascinating to examine the molecular details of imprinting in these agronomic species.

IMPACTS OF THE ARABIDOPSIS GENOME SEQUENCE ON FUNCTIONAL BIOLOGY

The high-quality Col-0 ecotype genome sequence revolutionized the study of nearly all areas of basic and applied plant science. Suddenly experimentalists no longer needed to clone and sequence genes and wonder whether there were uncharacterized structurally related genes undiscovered elsewhere in the ge-

* Corresponding author; e-mail lastr@msu.edu.
www.plantphysiol.org/cgi/doi/10.1104/pp.110.159541

nome. Evolutionary biologists began to peer back in evolutionary time to observe remnants of ancient genome duplications (Vision et al., 2000; Bowers et al., 2003) and deduce acquisition of new functions or gene loss over evolutionary time (Blanc and Wolfe, 2004). Extreme examples include enormous gene families such as disease resistance R genes and metabolic enzymes genes such as cytochromes P450.

The sequence had a tsunami-strength ripple effect on the study of gene function, enabling production of a variety of very useful reagents for studying large parts of the transcriptome and proteome. For example, the AtGenExpress (http://www.weigelworld.org/resources/microarray/AtGenExpress) resource includes a large set of transcript profiling experiments from many Arabidopsis tissue types and treatments using a single gene expression platform, the ATH1 Affymetrix GeneChip (Redman et al., 2004). This and other large mRNA expression datasets led to the development of a series of user friendly and powerful analysis tools such as Genevestigator and the Botanical Array Resource (see Lu and Last, 2008 for information on these and other Arabidopsis functional genomics resources). These resources allow plant biologists to look at developmental or stress-regulated expression of specific genes and larger scale regulatory networks.

Sequence-indexed insertional mutants are available for most protein coding genes of Arabidopsis, and these are arguably the most widely useful reverse genetics tools to emerge from the completed Col-0 genome sequence (Sessions et al., 2002; Alonso et al., 2003; O'Malley and Ecker, 2010). There are hundreds of thousands of T-DNA and transposon mutants with flanking DNA sequence data, including tens of thousands of homozygous lines. These can be ordered from the international seed stock centers (Arabidopsis Biological Resource Center and Nottingham Arabidopsis Stock Centre), and once their genotype is verified, the lines can be used to assess phenotypes associated with a few or thousands of mutants (Ajjawi et al., 2010).

Forward genetics continues to be among the most useful approaches to dissect complex biological processes and gene function annotation. Despite innovations in mutant screening, marker discovery, and genotyping, map-based cloning continues to be relatively slow and expensive, even in Arabidopsis. The ability to sequence and assemble genomes at low cost is changing the way that forward genetics is being done, and there is an increased interest in whole genome association mapping as an approach to identify alleles that influence a wide variety of phenotypes. Discovery of genes from mutagenesis screening is also becoming streamlined and the long-awaited genetic mapping by sequencing the whole genome of a mutant is becoming a reality (Schneeberger et al., 2009).

These and other genomics approaches as well as data and hypotheses established in Arabidopsis were quickly embraced by plant scientists working in economically important plants. Analysis of evolutionary conservation of coding sequences (Wu et al., 2006) has been hugely useful to the research and crop genetics communities. The broadly applicable reverse genetics tools TILLING (McCallum et al., 2000) and ecotilling (Comai et al., 2004) were developed in Arabidopsis using the Col-0 sequence and have been widely deployed in many species of plants and animals (for recent examples, see Minoia et al., 2010 and Stephenson et al., 2010). Given the cost and technical challenges associated with transgenic crop improvement, TILLING holds great promise for improvement of other commercially important plants.

An independent measure of the pervasive impact of the Arabidopsis genome project on applied plant science is seen by the extensive adoption of the weed in corporate research and development portfolios. Starting in the late 1990s, large functional genomics programs were developed in the private sector ranging from startups such as Paradigm Genetics (Boyes et al., 2001), Mendel Biotechnology (Riechmann et al., 2000), and Ceres (Haas et al., 2002) to large Ag-Biotech companies including BASF (Trethewey, 2001), Monsanto (Jander et al., 2003; Valentin et al., 2006), and Syngenta (Sessions et al., 2002). These companies developed and incorporated Arabidopsis genomics technologies through internal research and collaborations. By providing access to the Col-0 DNA sequence, the nonprofit research community received invaluable information on DNA polymorphisms, full-length cDNA sequences, and sequence-indexed T-DNA insertion mutants from these companies.

ARABIDOPSIS AS THE GENOMICS ENTRY SPECIES FOR PLANT BIOLOGY

The value of access to the Arabidopsis genome sequence and downstream functional genomics resource was readily apparent to the greater plant community and genome sequencing, along with resequencing, epigenetics, and functional genomics, was adapted rapidly in a number of other species. Following Arabidopsis, the rice (*Oryza sativa*) genome was sequenced, and as of Spring 2010, genome sequences are publicly available for nearly two dozen plant species. Given the innovations in sequencing technologies and shotgun sequence assembly algorithms, we can expect similar large-scale whole genome sequencing initiatives in many more species of agronomic, evolutionary, and ecological importance within 5 to 10 years. However, it is important to note that creation of gold standard reference genomes like Arabidopsis is still very expensive. As a result, few plant species have, or will likely achieve, this level of quality in the near future. Currently, rice has a high-quality reference genome (The International Rice Genome Sequencing Project, 2005) and curated genome annotation (Ouyang et al., 2007). In contrast, all other plant genome sequences exist primarily as uncurated draft genome sequences with inherent errors in the sequence, assembly, and annotation.

Small genome size and simple transformation methods were among the attributes central to Arabidopsis being chosen by the community to be the first plant genome sequenced and for creation of the most advanced set of functional resources. However, the rapid increase in numbers of crop species and other small-sized genome species with genome sequences leads to the question of whether Arabidopsis will continue to play such a dominant role in plant genomics. Arabidopsis will certainly dominate the plant genomics field in the near term due to the ease of method development and hypothesis testing. Indeed, the assessment of population diversity in plants was initiated in Arabidopsis first using single feature polymorphism detection, hybridization-based resequencing, and now with the 1,001 genomes project. All of these were then applied to other species.

However, Arabidopsis will not always be the first or most desirable organism for tool development or basic genomics research. In fact, other clades such as the Poaceae have played leading roles in comparative plant genomics, where genome sequences are available from four species. Another example is the increasing ability to employ versatile genetic materials and genomics information in crop species research thanks to the diversity of germplasm, including breeding lines. For example, in maize, diversity panels and a set of nested association mapping lines were developed prior to availability of the genome sequence and are enabling study of topics as diverse as domestication mechanisms (Tian et al., 2009) and genetic variation that influences grain nutritional quality (Yan et al., 2010).

The open and collaborative nature of the Arabidopsis community, along with the high value placed by scientists and international funding agencies on creation and sharing of resources, is an often overlooked key to the success of Arabidopsis research. Publically available resources abound, including stock centers for depositing, archiving, and retrieving germplasm, and molecular clone resources that are unsurpassed in plants. In conjunction with the high-quality genome sequence and annotation, Arabidopsis will certainly be a centerpiece in all future plant research. Indeed, the transitive nature of automated annotation methods in which functional annotation is inferred electronically from sequence similarity are the primary, if not sole mechanisms, of functional annotation in genome projects, highlights why continuing to characterize gene function in Arabidopsis will be critical to all plant systems.

IS THE PLANT SCIENCES PIPELINE FROM INNOVATION TO APPLICATION AT RISK?

An important sign of a healthy international research and development environment is innovation and excellence at all stages of the pipeline, from cutting edge research on the most fundamental processes to delivery of economic value to society. One reason why curiosity-driven research is so critical to long-term technology advancement is that it is impossible to accurately predict where the next transformative discoveries will be made. For instance, the recombinant DNA revolution has its roots in fundamental research in bacterial and phage genetics and much of our current understanding of vertebrate development was spawned by work on developmental genetics in the fruitfly (*Drosophila melanogaster*). Similarly unpredictable roads led to key innovations in plants, including crop plant transformation (Shah et al., 1986) from studies of the tumor-forming soil bacterium *Agrobacterium tumefaciens* (Van Larebeke et al., 1975; Chilton et al., 1977), and unexpected results from engineering petunia (*Petunia hybrida*) flower color (Napoli et al., 1990) led to a revolution in our understanding of how small noncoding RNAs influence eukaryotic gene expression.

Funding opportunities for translational plant sciences are increasing while basic model-organism research diminishes, putting the future of plant research at risk. A consensus is emerging among many plant biologists that the current government funding programs have decreasing room for research on fundamental biological processes in crop or model plants. As exemplified in this article, Arabidopsis continues to have a key role in understanding plant biology in all species. It is imperative that all funding agencies continue to financially support that cutting edge in Arabidopsis and other reference and model organisms to the maximal extent that befits their individual missions. The research community must continually educate policy makers about the importance of all aspects of plant biology, from eureka moments to the farm gate.

Received May 19, 2010; accepted June 15, 2010; published October 6, 2010.

LITERATURE CITED

Ajjawi I, Lu Y, Savage LJ, Bell SM, Last RL (2010) Large-scale reverse genetics in Arabidopsis: case studies from the Chloroplast 2010 Project. Plant Physiol **152:** 529–540

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

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

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679–1691

Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature **422:** 433–438

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. Plant Cell **13:** 1499–1510

Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. Cell **11:** 263–271

Comai L, Young K, Till BJ, Reynolds SH, Greene EA, Codomo CA, Enns LC, Johnson JE, Burtner C, Odden AR, et al (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. Plant J **37:** 778–786

Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. Science **324:** 1447–1451

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

Jander G, Baerson SR, Hudak JA, Gonzalez KA, Gruys KJ, Last RL (2003) Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant Physiol **131:** 139–146

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

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell **133:** 523–536

Lu Y, Last RL (2008) Web-based Arabidopsis functional and structural genomics resources. *In* The Arabidopsis Book. American Society of Plant Biologists, Rockville, MD, doi/10.1199/tab.0118, http://www.aspb.org/publications/arabidopsis/

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. Plant Physiol **123:** 439–442

Minoia S, Petrozza A, D'Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bendahmane A, Carriero F (2010) A new mutant genetic resource for tomato crop improvement by TILLING technology. BMC Res Notes **3:** 69

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes in trans. Plant Cell **2:** 279–289

Newman T, de Bruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, et al (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones. Plant Physiol **106:** 1241–1255

O'Malley RC, Ecker JR (2010) Linking genotype to phenotype using the Arabidopsis unimutant collection. Plant J **61:** 928–940

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, et al (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res **35:** D883–887

Redman JC, Haas BJ, Tanimoto G, Town CD (2004) Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. Plant J **38:** 545–561

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, et al (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science **290:** 2105–2110

Rounsley SD, Last RL (2010) Shotguns and SNPs: how fast and cheap sequencing is revolutionizing plant biology. Plant J **61:** 922–927

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jorgensen JE, Weigel D, Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. Nat Methods **6:** 550–551

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

Shah DM, Horsch RB, Klee HJ, Kishore GM, Winter JA, Tumer NE, Hironaka CM, Sanders PR, Gasser CS, Aykent S, et al (1986) Engineering herbicide tolerance in transgenic plants. Science **233:** 478–481

Stephenson P, Baker D, Girin T, Perez A, Amoah S, King GJ, Ostergaard L (2010) A rich TILLING resource for studying gene function in *Brassica rapa.* BMC Plant Biol **10:** 62

The International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature **436:** 793–800

Tian F, Stevens NM, Buckler ESt (2009) Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. Proc Natl Acad Sci USA (Suppl 1) **106:** 9979–9986

Trethewey RN (2001) Gene discovery via metabolic profiling. Curr Opin Biotechnol **12:** 135–138

Valentin HE, Lincoln K, Moshiri F, Jensen PK, Qi Q, Venkatesh TV, Karunanandaa B, Baszis SR, Norris SR, Savidge B, et al (2006) The *Arabidopsis vitamin E pathway gene5-1* mutant reveals a critical role for phytol kinase in seed tocopherol biosynthesis. Plant Cell **18:** 212–224

Van Larebeke N, Genetello C, Schell J, Schilperoort RA, Hermans AK, Van Montagu M, Hernalsteens JP (1975) Acquisition of tumour-inducing ability by non-oncogenic agrobacteria as a result of plasmid transfer. Nature **255:** 742–743

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. Science **290:** 2114–2117

Weigel D, Mott R (2009) The 1001 genomes project for *Arabidopsis thaliana.* Genome Biol **10:** 107

Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics **174:** 1407–1420

Yan J, Kandianis CB, Harjes CE, Bai L, Kim EH, Yang X, Skinner DJ, Fu Z, Mitchell S, Li Q, et al (2010) Rare genetic variation at *Zea mays* crtRB1 increases beta-carotene in maize grain. Nat Genet **42:** 322–327