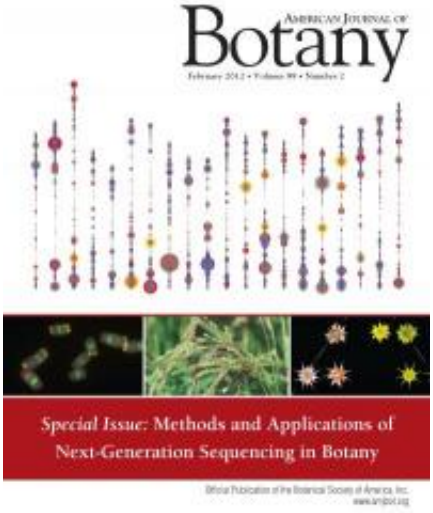
Analyzing complex plant genomes with the newest next-generation DNA sequencing techniques

February 27 2012



A visualization of relative gene expression among three species of the legume genus Glycine: an allotetraploid (blue) and its diploid progenitors (red and yellow). Per-cell expression level is represented by the surface area of the circles, one circle for each gene model × species combination, and mapped onto 44,326 gene models across the 20 chromosomes of the closely related reference genome of *Glycine max* (soybean). Image credits: Glycine gene expression (courtesy of Daniel Ilut); Root tip metaphase chromosomes of *Tragopogon miscellus* (courtesy of Richard Buggs); rice sample from the IRRI gene bank (courtesy of Susan R. McCouch); *Tragopogon* triangle (Reprinted from Cold Spring Harbor Symposia; see article by Buggs et al., Fig. 2, for full credit line.) Credit: Cover design: Adrianna Sutton (http://www.adriannasutton.com/).

Genomes are catalogs of hereditary information that determine whether an organism becomes a plant, animal, fungus or microbe, and whether the organism is adapted to its surroundings. Determining the sequence of DNA within genomes is crucial to human medicine, crop genetics, biotechnology, forensic science, threatened species management, and evolutionary studies. The last 5 years have witnessed tremendous advances in DNA sequencing technologies, and it is now possible to sequence millions of fragments of DNA in a single analysis, and at a fraction of their previous cost. These "next-generation" methods are spurring a revolution in plant biology by providing powerful tools to examine previously-unimagined questions, in any plant of interest.

Richard Cronn and colleagues (from the USDA Forest Service, Oregon State University, Brigham Young University, and Linfield College) have published an overview of newly developed, up-and-coming DNA sequencing techniques as one of a series of articles in a Special Issue on Methods and Applications of Next-Generation Sequencing in Botany in the *American Journal of Botany*. In their article, Cronn and co-authors summarize "targeted enrichment" strategies that can be used to obtain specific DNA sequences from complex plant genomes. Articles in the Special Issue provide a detailed snapshot of how "next-generation" sequencing is transforming plant biology.

"Plant genomes range from simple to exceptionally complex," noted Cronn. "Combining next-generation sequencing with targeted enrichment allows plant scientists to reduce the complexity of plant genomes and focus on specific genes or unique regions that are easy to analyze."

Modern next-generation sequencers are capable of sequencing all nucleotides (the G, A, T, and C components) contained within simple genomes, and from the genomes of well-studied organisms like humans. Direct genome sequencing is less tractable for many plants, whose

simple appearance belies their genomic complexity. As complex as we are, humans possess about as many genes (22,000) as diminutive mosses, and less than half the number of genes found in alfalfa or apple. Moreover, our genome of 3.2 billion nucleotides is dwarfed by many plants, such as pines and coastal redwood (which are 10 times larger). Finally, plants have an additional genome within each cell that encodes the genes required to carry out photosynthesis inside the chloroplast.

"At the moment, these sequencers are less than optimal for comparative plant biologists who focus on select regions of the genome from hundreds of samples," adds David Spooner, one of the editors for the special issue.

The enrichment methods described by the authors fill this much-needed gap, and are based on time-tested methodologies, including: (1) PCR-based enrichment, (2) hybridization-based enrichment, (3) restriction enzyme-based enrichment, and (4) enrichment of expressed gene sequences. The advantages and disadvantages of each technique are explored, and suggestions are offered for their optimal use.

In their study, the authors conclude that traditional PCR-based methods offer a cost-effective strategy for accessing small genomic targets in the range of 50,000 nucleotides or less. Larger genomic targets are most efficiently enriched using hybridization- and transcriptome-based methods. The article is accompanied by open-access electronic appendices that provide example protocols for hybridization-enrichment techniques, and detailed cost estimates for each targeted sequencing method.

"The use of these methods is sure to hasten the pace of discovery in all aspects of plant biology—crop production, ecosystem health, and our understanding of plant diversity in the present and the distant evolutionary past," adds Cronn.

More information: Cronn, Richard, Brian J. Knaus, Aaron Liston, Peter J. Maughan, Matthew Parks, John V. Syring, and Joshua Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99(2): 383-396. DOI: 10.3732/ajb.1100356