

MEETING REPORT

Low-cost sequencing opens new insights into diverse plant genomes

Kenneth D Birnbaum*

Abstract

A report on the Plant Genomes and Biotechnology: From Genes to Networks meeting, held at Cold Spring Harbor Laboratory, 30 November to 3 December 2011.

Introduction

In plants, many different species have a direct social impact through such problems as food production, bio-fuels and ecological services. In the past 20 years, genetic and genomic research have largely focused on a few strategically chosen models but, for many economically and environmentally important plant species, few genetic and genomic resources existed. The Plant Genomes and Biotechnology meeting at Cold Spring Harbor featured many examples of the use of next-generation sequencing (NGS) and other technologies to advance the development of emerging models that address specific problems, ranging from food security in Africa to a rare glimpse at the biochemical warfare between plants and the pests that eat them. In addition, the meeting provided an early look at an impending explosion of new data, emerging from NGS and other SNP technologies, that associate genetic (and epigenetic) variation to phenotypic variation in *Arabidopsis*, maize and rice. The ability to create inbred lines and easily share seed stocks makes plants a potential exemplar for large-scale genome annotation by association, with the same genotypes easily scored for multiple phenotypes. Other developments at the meeting included new efforts in 'industrial-grade' phenotyping (employing robotics, satellites and even military-style drones) and advances in techniques that have allowed the molecular dissection of the specialized cells that produce useful plant products.

Fast-track development of new models

How can new technologies rapidly accelerate plant research to help control a world-class pest? One such example is work on the parasitic genus *Striga*, which senses its hosts' strigolactones, hormones that are exuded into the soil. *Striga hermonthica*, for example, sends specialized roots known as haustorium to raid its hosts' nutrient supplies via vascular fusion, devastating subsistence crops such as sorghum in sub-Saharan Africa, and affecting the lives of nearly 100 million people. Ken Shirasu (RIKEN, Japan) presented his work on whole genome and RNA sequencing in *Striga* using next-generation technology. Using both facultative and obligate parasites in the genus, he described new research to identify genes expressed during the formation of haustorium by *Striga* in an effort to find an Achilles' heel of sorts of the pest. The sequencing work has shown that the dicot *Striga* has also incorporated into its genome up to a dozen genes from its monocot hosts via horizontal gene transfer, according to Shirasu, leading to new insights into host-pest evolution.

Pest control: lessons from basic science

Why can't Sorghum and other crops be engineered to reduce or eliminate strigolactone production to avoid parasite detection? Keynote speaker Ottoline Leyser (Cambridge University, UK) presented her work in the model plant *Arabidopsis* showing that strigolactones play essential roles in development by modulating the abundance of auxin efflux carriers (PINs) on plasma membranes. In Leyser's model, the balance of auxin sink capacity in the primary stem relative to the source capacity of the axillary bud regulates bud outgrowth. High auxin flux from the primary shoot prevents canalization of auxin in axillary buds, inhibiting vascular organization and bud outgrowth. Strigolactones reduce PIN accumulation to prevent vascular organization from buds and modulate axillary bud-mediated branching. Leyser showed computational models incorporating strigolactone's effect on polar auxin transport that could explain complex phenotypes. The models also confirm the central role of strigolactones in modulating auxin. The net result is that knocking out strigolactones leads to

*Correspondence: ken.birnbaum@nyu.edu
The Center for Genomics and Systems Biology, Department of Biology, New York University, 12 Waverly Place, New York, NY 10003 USA

high levels of branching. This role of strigolactones in branching is also conserved in sorghum and other monocots, as Shirasu confirmed. Thus, basic research in *Arabidopsis* shows that the strigolactone pathway is too fundamental to development to be the target for manipulation, a poignant example of how basic research in powerful model systems informs applied research. These realities necessitated Shirasu's alternative approach to combat *Striga* using genomics.

Leyser also presented recent work using natural variants to investigate the role of plasticity by examining variation in shoot branching under different nitrogen treatments. She described two strategies: low-plasticity, low-branching genotypes with rapid life cycles (so called 'James Dean' strategies) versus a high-plasticity set of genotypes with longer life cycles (so-called 'wait-and-see' strategies). However, after vernalization, some 'wait-and-see' varieties switched to the 'James Dean' strategy, apparently conditioning their strategy based on a memory of winter and the likely environmental conditions to follow. This was a keen example of a growing number of cases that document the ability of a plant to integrate many types of information to shape its body plan, a phenomenon that Leyser called 'the plasticity of plasticity.'

von Humbolt goes proteomic

The sensitivity of plant development to environment underscores the importance of examining plants in their native surroundings. Here again, new technologies have opened insights into the role of secondary metabolites in an ecological context. Keynote speaker Ian Baldwin (Max Planck Institute for Chemical Ecology, Germany) presented a body of work that began with metabolic profiling of wild tobacco (*Nicotiana attenuata*) subjected to herbivory. The ensuing research was a tale of partnership and betrayal in which pollinators turn pests, whose predators, in turn, are chemically summoned to devour them. Baldwin's group used gene silencing technology to knock down specific secondary metabolite synthesis or responses in *N. attenuata* and put plants back into their native habitats in Utah. The next step in the protocol, according to Baldwin, was perhaps the most essential: the low-tech art of field observation.

In one example, a caterpillar herbivore, *Manduca sexta*, was shown to secrete fatty acid-amino conjugates (FACs) in its saliva that trigger plant defenses. In addition to toxins and digestibility factors, FACs induce the plant to release a set of 146 volatile compounds. Using genetic manipulation (and observation), the group discovered that one of these compounds attracts carnivorous insects that prey on *M. sexta*, which is guided to the precise location of its prey by another set of compounds released from plant-FAC interactions. Baldwin also showed a new system in which RNA interference in *M. sexta* is induced

by insect feeding on plants expressing hairpin constructs, a technology that could have important implications for crop protection. He invoked the approach of the 19th Century Biogeographer Alexander von Humbolt, who attempted to unify different scientific disciplines in explanation of natural phenomena. Baldwin described his vision of a new breed of field scientists with an interdisciplinary perspective 'like von Humbolt with spectrophotometer in one hand and a microarray in the other.'

In another view of the plant-pest-symbiont relationship in *Medicago*, Blake Meyers (University of Delaware, USA) showed that small-RNA sequencing revealed a group of *trans*-acting small interfering RNAs (siRNAs) that appear to regulate a large family of defense-related proteins, NB-LRRs. Such a large-scale targeting of genes represents a new potential mode of regulation in small RNAs and provides potential insights into the control of symbiosis in *Medicago*.

Popgen takes on functional genomics

Another group of researchers focused on using NGS to explore plant diversity at a finer scale. Magnus Nordborg (Gregor Mendel Institute, Austria) provided an overview of progress in the genetic characterization of natural variation in *Arabidopsis*. With the combined output of the international 1001 Genomes sequencing project and other SNP data from hybridization technologies, Nordborg estimated that the *Arabidopsis* community will have high-density SNP and other polymorphisms for about 2,000 natural inbred lines within the next year. His group showed some early results in which associations were mapped to combined phenotypes, where multiple traits borrow mapping strength from each other and reduce noise. Alluding to published work, Nordborg provided some words of caution about instances in which genome-wide association mapping can be positively misleading. Nevertheless, Nordborg said such cases are relatively infrequent; he envisioned that genome-wide association studies will soon provide the plant community with a general resource in which each variable position in the *Arabidopsis* genome is annotated with associations for many phenotypes.

Edward Buckler (Cornell University, USA) described a similar scale project in maize, looking at 100 lines for 50 million segregating variable regions. Buckler said early analysis of the domesticated maize genome compared with that of its wild relative teosinte showed that about 500 to 600 genes were the targets of domestication. Surprisingly, there appear to be an equal number of genes targeted between tropical and temperate maize varieties - underscoring the importance of local adaptation. Buckler said the costs of genotyping are declining rapidly, with current costs at about \$9 a sample, which presents both an opportunity and a massive data problem. He

envisioned that 30,000 samples will soon be genotyped at 200 million variable positions, with a looming '12 trillion data point' storage and access challenge. Nonetheless, like Nordborg, Buckler foresees the pending genotype-to-phenotype resources as a game-changing development for genetics. In the case of maize, such a dense set of markers and phenotype associations should provide predictive power to cut down on one of the biggest bottlenecks in breeding programs: waiting for crops to mature to score their phenotype.

In another potential advance for plant breeding, Simon Chan (UC Davis, USA) showed his work on creating haploid plants using perturbation of the gene encoding the centromere-specific histone (*CENH3*). He showed his work on creating haploid plants using perturbation of the gene encoding the centromere-specific histone (*CENH3*). Chan showed that haploid plants can be rapidly generated, with aberrant-parent chromosomes efficiently lost after fertilization and diploids recovered in the next generation. He showed a remarkable set of pilot experiments in which he and collaborators created completely homozygous recombinant inbred mapping populations within two generations, a potential fast track for quantitative trait locus analysis in crops and other plants.

The epigenetic contribution to variation

Such technological advances in genotyping are leading to a more thorough accounting of the genetic causes of phenotypic variation. Indeed, Nordborg showed instances in which almost all the heritable variation for a given trait could be mapped to specific genome locations. Still, one question that looms in genotype-to-phenotype analyses is the role of epigenetic effects. Bob Schmitz (Salk Institute, USA) from the Ecker Lab showed recent results of high-throughput whole-genome bisulfite DNA sequencing in 182 different accessions. The group found 14,000 differentially methylated regions among the accessions, with the majority in transposable elements, but a significant proportion in coding regions. Approximately 50% of the epialleles in coding regions tended to be rare or often exclusive to one background. Moreover, using an association mapping approach, methylQTLs identified epialleles that are present at high allele frequencies in the population. These methylQTLs will help reveal the genetic architecture of some epialleles. For epialleles that did not have genetic basis, Schmitz posited that they come from incomplete methylation during reproduction, as evidenced by the overlap of loci affected in natural variants and *met1* or *rdm1* mutants.

Clearly, some proportion of epigenetic variation is caused by underlying genetic variation, such as transposable elements. However, it will be important to know how much phenotypic variation could be caused by purely epigenetic effects, which are potentially invisible

to the genotype map. Both Schmitz and Detlef Weigel (Max Planck Institute for Developmental Biology, Germany) described recently published work on the spontaneous accumulation of CG methylation variation after 30 generations of single-seed propagation. The take-home message was similar in both presentations: the accumulation of DNA methylation variation is perhaps 1,000-fold faster than the rate of appearance of spontaneous genetic variation. In addition, newly arising epigenetic variation is far from random. For example, Weigel reported that although only 1% of all methylated sites distinguished a random pair of lines 30 generations apart, many sites changed in more than one line. Although phenotypic effects of some methylation variants are well documented, the next question is how much of the natural variation and spontaneous DNA methylation observed in genome-wide surveys has an impact on phenotype, answers to which should be coming soon. Another important question that arises from genotype-to-phenotype mapping when 'phenotypes' are transcriptomic readouts, as noted by Weigel, is what tissue and cell types will be assayed to associate with genetic variation.

On the flip side of the genotype-to-phenotype efforts, the characterization of phenotype is also benefiting from an infusion of new technology. Ulrich Schurr (Forschungszentrum Jülich, Germany) described a battery of automated phenotyping tools for root and shoot being developed at Jülich's Institute of Plant Sciences. The pipeline included magnetic resonance imaging and positron emission tomography of roots grown in soil to characterize root architecture and function, with Schurr emphasizing the collection of high-throughput physiological measurements. They have also tracked soil nutrient distribution in fields, showing dynamic changes in local nitrogen profiles on a day-to-day basis. Schurr explained later that efforts to coordinate different types of measurements taken from unmanned drones and satellites will go toward gathering more accurate measurements of photosynthetic rates, which are notoriously inaccurate. He said that such revised measures of photosynthesis could, for example, improve the accuracy of climate models.

Cellular genomics

Another set of well-developed genomic techniques in plants has led to analysis of specialized cells and stages of maturation. Several speakers presented recent work using high-resolution transcriptomic profiles as a diagnostic tool to interpret phenotypes. Zachary Lippman (Cold Spring Harbor Laboratory, USA) used a library of developmental stage-specific profiles to show that mutants or natural variants in tomato with highly branched inflorescences appeared to arise from perturbations of developmental staging, a heterochronic effect of sorts; sympodial inflorescence meristems in backgrounds that gave rise to

more reproductive structures were 'born' at younger stages of development than less branched backgrounds. The apparent juvenilization of secondary meristems correlated with the potential to bear an inflorescence with more fruiting structures, a potentially important agronomic trait. I used the diagnostic features of cell type profiles to quantitatively analyze single cell profiles, showing that a single cell could be fine mapped using its RNA-seq profile to a small population of cells in the root. The goal of the work is to examine how the cell buffers transcriptional noise and how plant cells retain their remarkable plasticity. Siobhan Brady (UC Davis, USA) used a compendium of tissue- specific expression profiles to define a candidate set of transcription factors that together may regulate xylem fate. She then used yeast one hybrid analyses to construct a transcriptional network for xylem cells. Since many useful products or nutritive parts of the plant arise from specialized cells,

knowledge of their circuitry in these powerful cellular dissection systems in *Arabidopsis* is likely to be useful in improving plant production, much as basic research in the strigolactone pathway has informed applied research in crop protection.

Abbreviations

FAC, fatty acid-amino conjugate; NGS, next-generation sequencing; SNP, single-nucleotide polymorphism.

Acknowledgements

KDB is supported by the National Institutes of Health and the National Science Foundation.

Published: 30 January 2012

doi:10.1186/gb-2012-13-1-311

Cite this article as: Birnbaum KD: Low-cost sequencing opens new insights into diverse plant genomes. *Genome Biology* 2012, **13**:311.