Meeting report

Finding the functional gems in plant genomes

Alexandra M Jones and Murray Grant

Address: Department of Agricultural Sciences, Imperial College London, Wye, Kent TN25 5AH, UK.

Correspondence: Murray Grant. E-mail: M.Grant@imperial.ac.uk

Published: 5 November 2003

Genome ***Biology*** 2003, **4:**350

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/12/350

A report on the joint second Plant Genomics European Meeting (Plant GEMs) and fourth Genomic *Arabidopsis* Resource Network (GARNet) meeting, York, UK, 3-6 September 2003.

The combined Plant GEMs and GARNet meeting included talks on technology development, genome evolution and service provision, in addition to fundamental assignment of gene function. The two meetings evolved from a common desire to provide efficient development and exploitation of the national and trans-European platforms for plant functional genomics that have been established by programs in Germany, France, The Netherlands and the UK. Combining the meetings provided a forum for information and technology transfer and the opportunity to establish cross-community collaborations.

The 2002 Nobel laureate Sydney Brenner (Salk Institute, La Jolla, USA) gave the keynote presentation on 'inverse genetics'. Within little more than a decade, the global *Arabidopsis* community has developed an enviable collection of genomic resources. One might wonder how much more there might be had Brenner shunned "those things that wriggled" during a brief encounter with *Arabidopsis* in 1962. In contrast to classical genetics that starts with an interesting phenotype and works back towards its genetic basis, inverse genetics works from a known genetic background to make predictions about the resultant phenotype. 'Transgenomic relocation' - moving regulatory elements across vast evolutionary distance such as from mice to teleost fish - makes it possible to search for elements that control the same function among otherwise randomized non-coding sequences. By finding the jewels in the 'junk DNA', one can begin to unravel the code of regulatory elements that work in concert to define multicellular organisms.

In contrast, Rob Martienssen (Cold Spring Harbor Laboratory, New York, USA) says he "cares about junk". The junk in question encodes silent transposons that constitute the bulk of highly condensed chromatin - known as heterochromatin - in *Arabidopsis*. Immobile, silent transposons are distinguished from active transposons and genes by being methylated on both DNA and histones. Martienssen has used microarrays of overlapping genomic DNA fragments covering whole chromosomes to examine DNA methylation in various mutant backgrounds. Surprisingly, although the *decrease in DNA methylation* (*ddm1*) mutant affects histone modification, the corresponding gene does not encode a methyltransferase but rather a chromatin-remodeling enzyme of the SWI/SNF family that causes changes of the methylated residue of histone H3 from lysine 4 to lysine 9 in heterochromatin. Martienssen showed that the presence of a small interfering RNA (siRNA) correlated with methylation of histones bound to transposons on lysine 9, and he proposed a mechanism by which the cell can discriminate between silent transposons and active regions by the presence of specific siRNAs. Because of redundancies between the multiple genes in the *Arabidopsis* RNA interference (RNAi) pathway, Martienssen's group extended their studies using fission yeast, in which the centromeric repeats resemble transposons. Transcripts from these regions appear to be targets of RNAi that guide heterochromatin formation and centromere function.

Mutagenesis using the T-DNA vector derived from *Agrobacterium* has proved a powerful tool in reverse genetics in plants; approximately 85% of predicted *Arabidopsis* genes have now been mutagenized by T-DNA insertion. But the non-random nature of T-DNA insertion demands alternative technologies to fill the gaps. Mike Bevan (John Innes Centre, Norwich, UK) described a GARNet program that is generating sequences of the transposable-element insertion sites in promoter-trap, enhancer-trap and gene-trap lines.

Transposons have a complementary insertional bias to T-DNA and can provide a launch platform to facilitate transposon remobilization to linked target genes. An alternative approach for loss-of-function studies is gene silencing. Pierre Hilson (Ghent University, Belgium) described progress in the EU-funded Agrikola Project to generate around 28,000 gene-specific double-stranded hairpin RNAi constructs in T-DNA binary vectors. This resource promises to provide a universal utility for any transformable *Arabidopsis* ecotype and the possibility of downregulating expression of genes that would have lethal knockout phenotypes.

Introducing a subject on which many biologists are sceptical, Przemyslaw Prusinkiewicz (University of Calgary, Canada) succeeded in making mathematical modeling not only accessible but fascinating. Modeling of biological systems has several advantages: forcing explicit expression of assumptions, testing comprehension through accurate reconstruction of the original system, and analyzing complex traits that may emerge from simple developmental rules in a non-intuitive fashion. Prusinkiewicz has extended the application of parametric Lindenmayer systems (L-systems), introduced by Aristid Lindenmayer in 1968 to model the development of multicellular life, to computer-generated models of plant form, with the goal of relating genetic regulatory mechanisms to the resulting phenotypes. His models reduce plant growth to a set of simple rules involving switches and thresholds that govern the addition of 'building blocks' - apex, internode, leaf and flower. In a proof of concept, Prusinkiewicz entertained the audience with models mimicking the action of *Arabidopsis* mutants such as *leafy* and physiological traits such as the switch between vegetative growth and floral development. Most compelling was the model simulating auxin flow in two and three dimensions, creating an uncanny resemblance to reality in the resulting 'virtual plants'.

Polyploidy is highly prevalent in angiosperms and impedes comparative genomic studies, as genomes are duplicated and gene order shuffled. Andrew Paterson (University of Georgia, Athens, USA) described how his lab has reconstructed gene phylogenies using homologous expressed sequence tags (ESTs) to determine when duplication events occurred in angiosperms. Assuming that genes from different species clustering with duplicated genes from *Arabidopsis* denote a duplication event that occurred before divergence of the species, they identified three main duplication events - α, β and γ. The α duplication separated the *Arabidopsis* and *Brassica* genera from other dicotyledons, whereas the β duplication was common to all dicotyledons analyzed. The γ duplication represented the most ancient event, weakly discriminating between grasses and gymnosperms. Clearly, understanding polyploidization events will assist the use of comparative genomics in plant breeding.

Thomas Mitchell-Olds (Max Planck Institute of Chemical Ecology, Jena, Germany) provided further insight into evolution of plant genome size. Most of the *Arabidopsis* genus has eight chromosomes in the haploid genome so the most parsimonious explanation for the five chromosomes of *A. thaliana* is genome reduction. Comparing the size of the whole genome of *A. thaliana* with related species and genera suggests an approximately 40% reduction in *A. thaliana* genome size during the past 5 million years. Comparison of 300 orthologous genomic regions from two relatives, *A. petraea* and *Boechera drummondii*, indicates a genome-wide tendency towards smaller orthologous regions in *A. thaliana*. Examination of polymorphisms in 12 *A. thaliana* ecotypes offered no support for the hypothesis that genome contraction offers a selective advantage, however, although some evidence suggested a mechanistic bias favoring larger sizes for deletions relative to insertions. If this is true, two pertinent questions are why this phenomenon is not more widely observed and why *A. thaliana,* alone of the genus, underwent such a dramatic reduction in genome size. Perhaps the answer lies in the evolved self-compatibility of *A. thaliana* and the resulting rise in homozygosity.

In an intriguing departure from the norm, Nicholas Harberd (John Innes Centre) described biochemical studies using barley that aim to avoid the complications arising from the genetic redundancy of *Arabidopsis*. DELLA, a putative transcription factor, represses plant growth in response to gibberellic acid (GA). Barley has a single *DELLA*-like gene, *SLENDER*, compared with the five *DELLA*-like genes in *Arabidopsis*. Using proteasome inhibitors, Harberd's group has shown a requirement for the proteasome in the GA-induced disappearance of SLENDER and DELLA. Thus, GA signaling is another of the growing number of regulatory pathways shown to utilize targeted degradation of key regulators. Harberd has explored the interaction of auxin and ethylene with control of root growth by GA through a combination of genetics and transgenics. He presented some compelling evidence that DELLA proteins are key integrators of multiple signaling inputs regulating plant growth.

The problems of mining large datasets were acknowledged by most speakers in the metabolomics session. Concerns were raised over the variation observed within the same cultivar grown at different times even in controlled conditions. Royston Goodacre (University of Manchester Institute of Science and Technology, UK) described mining of metabolite-profiling data using explanatory machine-learning algorithms. These evolutionary algorithms generate 'pools' of mathematical functions following simple rules. The functions are propagated and mutated on the basis of their success at solving the set problem - here, discriminating between datasets. The process is iterated until a robust solution is found, and this is tested against other known datasets. The strength of this approach lies in its reductive power; a 'fingerprint' of thousands of metabolites can be reduced to a few characteristic metabolites or ratios, which may then be used to understand more fully the chemical

nature of the difference between these datasets (for example, a wild-type and a mutant plant).

Brenner drew the analogy between having a genome sequence and landing on the moon: landing was easy - it was getting back that was hard. The meeting highlighted concerted efforts underway to 'get back from the moon' - to achieve the challenging goal of functional annotation of all the *Arabidopsis* genes. This knowledge will be invaluable for mining the crop-plant genomes now being sequenced. Functional genomics is still in its infancy, however, and major advances in technologies are likely to push the field forward rapidly. But some key biological approaches do not lend themselves to high-throughput analysis. Although consideration is now being given to analyzing multifunctional proteins and the role of splice variants, studying the integration of proteins into multimeric complexes and their temporal composition and cellular location remain a real challenge. Hopefully these issues will be addressed at Plant GEMs 3 in Lyon, France, next year.