

Meeting report

A cornucopia of genomes

Andrew H Paterson

Address: Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30602, USA. Email: paterson@uga.edu

Published: 28 March 2006

Genome Biology 2006, **7**:311 (doi:10.1186/gb-2006-7-3-311)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2006/7/3/311>

© 2006 BioMed Central Ltd

A report on the Plant and Animal Genome XIV Conference, San Diego, USA, 14-18 January 2006.

The 14th annual Plant and Animal Genome conference held recently in San Diego highlighted the challenges facing researchers who attempt to annotate and interpret the burgeoning numbers of plant and animal genome sequences. These include the genomes of the world's leading crops and provide valuable models for the study of genetics, evolution and development. More than 80 workshops addressed emerging results and opportunities, as well as technological developments, in a host of plant, animal and microbial genomes. Two recurring themes of the meeting were the continuing 'siliconization' of plant and animal biology and the rapid progress being made in understanding the mechanisms of epigenetics and its biological roles.

Genomes and their analysis

Progress in sequencing plant genomes was highlighted by Aristotle Patrinos (US Department of Energy (DOE), Washington DC, USA) who announced that the Joint Genome Institute will be sequencing the soybean (*Glycine max*) genome, to add to its current whole-genome sequencing projects for sorghum (*Sorghum bicolor*), *Arabidopsis lyrata* and *Capsella rubella* (close relatives of the model plant *Arabidopsis thaliana*), and *Mimulus guttatus* (monkey flower), and its participation in the maize (*Zea mays*) genome project. Patrinos briefly outlined DOE systems biology approaches to its missions, in particular a 10-15-year goal of its 'Genomes To Life' program to generate a microbial sequence, produce all proteins and molecular tags, identify multiprotein complexes, generate regulatory networks, identify metabolic capabilities and engineer control strategies - "all in a few days".

Patrinos's description of the informatics challenge as a "tsunami looming over genome projects" was further elaborated on by Kimmen Sjolander (University of California,

Berkeley, USA), who noted that only 3% of gene annotations have empirical support. In addition to mechanical and/or technical errors, domain shuffling and gene duplication play an important role in generating annotation errors. Sjolander described the web-accessible tools within the Universal Proteome Explorer [<http://phylogenomics.berkeley.edu/UniversalProteome/index.php>], which is a freely available resource for evolutionary and phylogenetic analysis. These tools combine phylogenomic approaches with protein-structure prediction to elucidate correlations between protein structure and molecular function, and they prioritize experimentally verified information over *in silico* inferences. Extensive resources for the human and *Arabidopsis* proteomes are already in place.

Reaping the benefits of a genome sequence in a better understanding of biological complexity requires the functional characterization of the proteome, in particular the physical interactions among proteins and how these change when perturbed by disease. With the goal of a comprehensive interactome map for humans, David Hill (Dana-Farber Cancer Institute, Boston, USA) described stringent, high-throughput yeast two-hybrid analyses of pairwise interactions among 8,100 open reading frames (ORFs) cloned in Gateway vectors, which allow rapid, efficient transfer of sequences between cloning and expression systems. The 2,800 interactions found had a verification rate of about 78%, including a number with strong support from the literature. The resulting interaction maps reveal patterns such as the coevolution of interacting proteins, and also predict the identity of unknown genes that are linked in the network to genes known to have roles in key processes.

A longstanding challenge to plant and animal scientists is the dissection of the genetic basis of complex traits. This is now becoming more achievable through integrative approaches, as noted by Ariel Darvasi (Hebrew University of Jerusalem, Israel). He pointed to the huge number of haplotype patterns for inbred mouse strains as a powerful tool for widely applicable *in silico* approaches to pinpointing the

locations of quantitative trait loci and searching for functional single-nucleotide polymorphisms (SNPs) that may be directly involved in the trait. Darvasi described a test case that started with a gene mapped at a marginal lod (log-odds) score of 2.5 to a single chromosome. This was further localized by association genetics to a 12 Mb interval containing 10,893 SNPs. By additional crosses and analysis of the inferred functional consequences of SNPs, two SNPs were eventually identified that also show evidence of differential expression in association with the trait.

Networking insights

An alternative and complementary approach to the genetic dissection of complex traits is the analysis of complex networks of cellular processes using a 'parts list' obtained by genome annotation together with biochemical knowledge to assemble a metabolic reconstruction of an organism *in silico*. Analysis of metabolic fluxes through the network permits predictions about the phenotypic consequences of genetic differences. Bernhard Palsson (University of California, San Diego, USA) reported early results from microbe reconstructions, including the identification of areas of metabolism that are inadequately characterized, and gene-expression patterns that suggest an impact of the three-dimensional organization of the genome - for example, *Escherichia coli* appears to have six expression domains in its chromosome. Long-term cultures of more than 60 days suggest the predictive value of network reconstructions in the adaptive evolution of *E. coli*. Palsson also referred to an ongoing reconstruction for the human genome, in which larger network size makes for exponentially more possible functional states.

Network-based approaches are also providing new insights into a particularly important dimension of microbial adaptive evolution - the coevolution of crops and their pathogens and parasites. Jonathan Jones (The Sainsbury Laboratory, Norwich, UK) offered a "grand unified theory for plant resistance", which is based on an intricate exchange of activators and suppressors between host and pathogen that results in a dynamic sine-curve-like fluctuation of defense strength. Jones described how work in his and other labs shows that plant hormones play key roles in two such exchanges, suggesting that interactions between growth signaling and defense signaling appear to be even more complex than feared. The upregulation of auxin as part of one pathogen's effort to reduce the host plant's alert status is countered by activation of a microRNA that downregulates key mRNAs in the auxin response. In the second example, the gibberellin produced during foolish seedling disease is associated with the degradation of the host plant's DELLA proteins, which are involved in the defense against bacteria that produce rots.

In large eukaryotic genomes rich in repetitive DNA, the epigenetic regulation of gene expression imposes another

dimension on cellular networks. Rob Martienssen (Cold Spring Harbor Laboratory, New York, USA) reported the investigation of a heterochromatic region of *A. thaliana* chromosome 4 using a genomic tiling microarray to study methylation. Specific methylation of repeats, avoiding nearby genes, was accomplished by the chromatin-remodeling ATPase DDM1 guided by small interfering RNAs (siRNAs), a hallmark of RNA interference (RNAi). The microarrays revealed differences in methylation between wild-type and *ddm1* mutants. Several lines of evidence suggested that a role of tandem repeats in 'junk' DNA may be to generate large quantities of siRNA. Varying degrees of methylation were also found within genes in the region, and this methylation was highly polymorphic among ecotypes (varieties adapted to different habitats). Martienssen therefore postulated that a major component of natural variation could be such epigenetic variation in gene methylation status, which might eventually be translated into permanent genetic variation as a result of transamination and associated point mutations.

Genomic encyclopedias

Even *A. thaliana*, arguably the best characterized of the angiosperm genomes, is still in the 'parts-list' phase. Joe Ecker (Salk Institute, San Diego, USA) reported advances in the development of tools that will contribute to the production of an encyclopedia of functional elements for this botanical model. These include the expected availability of more than half of *A. thaliana* genes as ORFs, their transition to Gateway clones, the planned development of homozygous lines for each of two insertional mutations in most genes, and the targeted identification of insertional mutations into genes missed in present collections. Ecker reiterated the possible importance of the 'methylome' in natural variation, noting the unannotated transcripts found in *met1* mutants. He also pointed to progress by Detlef Weigel (Max Planck Institute for Developmental Biology, Tübingen, Germany) and Magnus Nordborg (University of Southern California, Los Angeles, USA) in capturing the 'haplosome' of *Ara-bidopsis*, using a whole-genome microarray wafer to fully resequence the genomes of 20 accessions, discovering many SNPs that appear to cause drastic point mutations and also thousands of deletions.

The discovery of the tetraploid ancestry of *A. thaliana* is one of the bigger surprises that came out of the sequencing of its genome. Michael Freeling (University of California, Berkeley, USA) asserted that duplication of an ancestral genome estimated at 22,000 non-transposon-related genes, followed by massive loss of many of the duplicated genes, has left a net increase of about 5,000 genes. Preferential retention of specific classes of genes, such as those for transcription factors, suggests that such duplication/fractionation cycles may increase morphological complexity. Non-random distributions in the inferred locations of 'missing pages' (duplicated

genes that have been lost) from the ancestral ‘encyclopedia’ suggested a role of epigenetic factors in gene loss. Although only weakly correlated with gene retention, Freeling pointed to 14,940 conserved nucleotide sequences, which represent an underexplored feature of plant genomes. This implicates conserved nucleotide sequences in ‘first responder’ genes (genes that rapidly respond to a variety of exogenous stimuli) and suggesting that some may tend to form secondary structures.

In his talk, Patrinos referred to the Human Genome Project as “making the impossible routine”, which equally well summarizes the progress reported at the conference over a wide range of organisms and activities. Over this meeting’s 14-year history, agricultural genomics has made the transition from having genetic linkage maps for a few plants and animals to its present state with nearly complete sets of knockout mutations for model plants and the capacity to tackle the sequencing of several gigabase-sized genomes at the same time. Agricultural genomics now seems poised to play a central role in addressing global challenges ranging from feeding the world’s poor to providing fossil fuel alternatives and mitigating global warming.

Acknowledgements

I thank Michael Gale for helpful comments on the draft version of this report.