

Meeting report

'Horizontal' plant biology on the rise

Yves Van de Peer

Address: Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology (VIB), Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. E-mail: Yves.vandeppeer@psb.ugent.be

Published: 21 December 2004

Genome Biology 2004, **6**:302

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/6/1/302>

© 2004 BioMed Central Ltd

A report on the Plant Genomics European Meeting (Plant-GEMS2004), Lyon, France, 22-25 September 2004.

The annual meetings on plant genomics, of which Plant-GEMS2004 was the third, are now among the most important plant meetings in Europe. This year, almost 600 scientists from more than 30 different countries participated, and the meeting was supported by the national programs in plant genomics in France, Germany, the UK and the Netherlands, and by the French, German, Spanish and British research ministries. This report focuses in particular on the strengths and expectations of comparative genomics in plants, an area that is only now starting to be fully exploited.

Comparative genomics is often praised as an extremely powerful way of discovering novel biological features. A well-known example of its power is the identification of conserved elements, such as *cis*-acting regulatory elements, in distantly related genomes: because of their conservation over long periods of time, such elements must have some important function. Another merit of comparative genomics is expected to be its ability to uncover the transfer of structural and functional information from one genome to another. This assumption is based on the observation that, although chromosomal rearrangements can be extensive, the genomes of different species still exhibit a certain degree of colinearity. Keynote speaker Steve Tanksley (Cornell University, Ithaca, USA) argued that only through comparative and integrative approaches will the mechanisms of evolution and adaptation be revealed, and he stressed the importance of moving from 'vertical' biology within a single species to 'horizontal' biology across species. Currently, the genomes of at least 10 plant species are being fully or partially sequenced. They have been selected to complement the two model plants whose genome sequence has already been determined, namely *Arabidopsis thaliana* (thale cress) and *Oryza sativa* (rice). Tanksley also reported on the Solanaceae

Genome Initiative, which is studying the genomes of tomatoes, potatoes and their relatives. One aim is to have a draft of the tomato genome by the end of 2006. Other questions to be tackled are how a common set of genes and proteins gave rise to such a wide range of morphologically and ecologically distinct species in the Solanaceae, and how a deeper understanding of the genetic basis of diversity can be harnessed to better meet the nutritional needs of society in an environmentally friendly way.

Comparing crops

A new genome sequence is that of poplar - officially released only the day before the meeting. The poplar genome is approximately 500 megabase-pairs (Mbp), divided between 19 chromosomes. Very preliminary analyses report more than 40,000 genes. Stefan Jansson (Umeå Plant Science Centre, Sweden), a member of the Poplar Genome Assembly and Annotation Committee, discussed the added value of the poplar genome for the plant community. For a long time, poplar has been developed as a model tree for genomics, to allow study of tree-specific traits, such as wood formation, longevity, seasonal changes and the juvenility/maturity transition. The poplar genome will also be of great value for studies on natural variation, ecology and population biology, because in all these aspects poplar is very different from *Arabidopsis*. On the other hand, from a phylogenetic point of view, poplar is relatively close to *Arabidopsis*, much closer at least than *Arabidopsis* is to rice. The poplar and *Arabidopsis* lineages diverged approximately 100 million years ago, and expectations are that detailed comparison of the two genomes will uncover many novel functional sites.

Maize is one of the most important crops and was domesticated from teosinte, a group of Central and South American grasses, in Mexico more than 7,000 years ago. Alain Charcosset (Station of Plant Genetics, Gif-sur-Yvette, France) presented a detailed historical analysis indicating that maize was introduced not once but twice into Europe: first to

southern Europe by Christopher Columbus, and again at the beginning of the sixteenth century by the Spanish or French. Klaus Mayer (Munich Information Center for Protein Sequences, Munich, Germany) discussed one of the maize genome initiatives, and the bioinformatics involved, in which the ends of approximately 475,000 maize bacterial artificial chromosome (BAC) clones have been sequenced, giving a cumulative length of 307 Mb of sequence, covering about one-eighth of the maize genome. Approximately 60% of this is formed of repeat sequences, whereas genic regions occupy about 7.5%.

Although the ancestor of maize was tetraploid, fewer than half of maize genes appear to be present in two orthologous copies, indicating that the maize genome has undergone significant gene loss since the duplication event. On the other hand, the number of tandem duplicates is unusually high. Preliminary estimates, to be treated with caution, predict more than 50,000 genes in the maize genome, which is more than in any other organism sequenced so far. Apart from having many genes, the maize genome is also very variable, as discussed by Peter Bradbury (Cornell University), who pleaded for this diversity to be exploited to improve maize performance. Making use of the natural variation of maize has major advantages over transgenesis, as it does not require transformation and also avoids political problems.

Catherine Feuillet (University of Zurich, Switzerland) showed that, despite major differences in genome size (mainly attributable to transposable elements), chromosome number and ploidy, gene order is generally well conserved among the cereals, which all shared a common ancestor approximately 70 million years ago. An example of how information on colinearity between genomes can be successfully applied was presented by Beat Keller (University of Zurich), who identified quantitative trait loci (QTLs) in wheat for resistance against leaf rust (*Puccinia triticina*) and the blotch fungus *Stagonospora nodorum*. The isolation of resistance QTLs is of great importance for developing molecular tools for breeding resistant crops. Keller reported that by using microsatellite and expressed sequence tag (EST) markers derived from wheat physical mapping projects, the genetic map in the QTL target region has been improved significantly, and a region spanning 7.6 centimorgans (cM) containing the leaf-rust resistance locus has been defined on chromosome 7DS (wheat is a hexaploid, as reflected in the chromosome naming).

The two ESTs flanking this QTL in wheat are conserved on chromosome 6 of rice in a region that is colinear between the two cereals. In rice, the homologous ESTs define a physical region of three BACs spanning approximately 300 kilobases (kb). The colinearity between rice and wheat will now be used to isolate possibly homologous wheat ESTs for mapping in the wheat region of interest. Rice genome information has thus been used to increase the number of markers in wheat,

so as to identify QTLs and disease-resistance genes. Another example of using colinearity between genomes to identify resistance genes was given by Pere Puigdomènech (Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain), who has identified the gene that confers resistance to melon necrotic spot carmovirus in *Cucumis melo* through considering localized synteny (microsynteny) of the *Cucumis* genome with that of *Arabidopsis*.

From simplicity to complexity

Hervé Moreau (Laboratoire Arago, Banyuls-sur-Mer, France) described the forthcoming release of the complete genome (approximately 11.5 Mb) of one of the smallest free-living photosynthetic organisms, the green alga *Ostreococcus tauri*. This is a marine photosynthetic picoeukaryote with one nucleus, one chloroplast and one mitochondrion. Comparison of gene order and conservation between green algae and higher plants will be difficult, but such simplified organisms may provide important clues about complex biological processes. This genome is indeed remarkable for the minimization of many cellular and biological processes. For example, Moreau showed that *O. tauri*, which diverged from the base of the green plant lineage, has the smallest complete set of core cell-cycle genes described to date. Therefore, unicellular algae might be good model organisms for improving understanding of basic but key molecular processes. The genomes of higher plants are usually not that simple and often contain, through gene duplication, many copies of genes, forming large gene families.

Such partial or complete redundancy can seriously complicate functional genomics studies. Gerco Angenent (Plant Research International, Wageningen, The Netherlands) discussed one large gene family, namely the MADS-box genes. In *Arabidopsis* this family has more than 100 members (in *O. tauri* there is evidence for only one MADS box gene) involved in different processes such as floral organ specification and root, seed and fruit development. Although these genes are the focus of much research, the function of many of the MADS-box transcription factors they encode is still unknown, as are their interacting protein partners (most MADS-box proteins form dimers). Angenent uses screens for protein-protein interactions to unravel, at least in part, the network of protein complexes in which MADS-box proteins play a role. He also uses protein-protein interaction screens to identify orthologs in other species, which is hard to do from sequence comparison where large gene families are concerned. Protein interactions are much better conserved than sequences in proteins from different species and therefore provide more reliable evidence on orthology.

Todd Vision (University of North Carolina, Chapel Hill, USA) reported on the divergence of expression profiles between duplicated genes in *Arabidopsis thaliana*. Subtle differences in the divergence pattern were observed between

duplicates that arose through different processes, such as tandem duplications, transpositional duplication or polyploidy. Time seems to be a poor predictor for divergence expression, which had mostly occurred very soon after the duplication event. He also noted a striking asymmetry between many duplicates in the breadth and abundance of expression, a phenomenon that is difficult to explain with the current models for functional divergence of duplicated genes.

Over 5 million plant EST sequences are now publicly available, with collections of more than 5,000 sequences for over 60 plant species. As Stephen Rudd (Centre for Biotechnology, Turku, Finland) noted, these species cover most of the plant kingdom, but with a clear bias towards the monocotyledons (which include the cereals and other grasses), and the dicotyledon subclasses Rosidae and Asteridae. EST sequences can play an important role in comparative genomics even though they represent a partial view of the genome at best. The suitability of EST sequences for comparative genomics has been evaluated by comparing EST sequences to the genomic scaffolds. The average rate of sequence error is 2.2 mismatches or indels (insertions and deletions) per 100 nucleotides. The lowest-quality sequences are the oldest in terms of when they were sequenced, whereas *Arabidopsis* ecotype differences apparently have only a minor effect on sequence quality. As might be expected, the clustering of the same sequences from different sequencing experiments to build so-called unigenes dramatically improves the quality; when sequence clusters with more than three members are considered, the error rate is reduced to only 1.6 per 100 nucleotides. Rudd presented an EST sequence-analysis pipeline called openSputnik [<http://sputnik.btk.fi>], in which both patterns of domain architectures and taxonomic restriction can be visualized, providing a foundation for more directed expeditions into comparative genomics.

Jan Lohmann (Max Planck Institute for Developmental Biology, Tübingen, Germany) focused on a different aspect of expressed genes. He discussed an international effort to develop a gene-expression atlas of *Arabidopsis* designated AtGenExpress, which will provide free access to a comprehensive set of Affymetrix microarray data that covers many different experimental conditions. Lohmann discussed a large-scale analysis of expression data from approximately 80 samples, consisting of a wide range of *Arabidopsis* tissues at various developmental stages, which forms part of this major resource. One of his main conclusions was that a large proportion of the more than 20,000 *Arabidopsis* genes are expressed in at least one developmental stage; in other words, approximately 93% of the *Arabidopsis* genes are expressed during development.

In summary, this year's meeting again made clear that these are the best of times for plant biologists. Besides the huge amounts of functional genomics data being generated, the

availability of many new partial or complete plant genomes will boost the use of comparative approaches. Undoubtedly this will lead to many novel and exciting findings in the near future. Stay tuned!