

MEETING REPORT

Plant genomics: from weed to wheat

Kerstin Kaufmann^{1*} and Wolfgang Busch^{2,*}

Abstract

A report on the first 'Plant Genomics Congress' meeting, held in London, UK, 12-13 May 2013.

Introduction

In the past decade, genomics research has enabled enormous progress in our understanding of plant genomes with regard to their structure, function and evolution. The rise and continuous improvement of next-generation sequencing (NGS) techniques has allowed researchers to shift focus from (usually small) model plant genomes to larger and more complex genomes of crop plants. Currently, about 34 flowering plant species have been sequenced and are publicly available, and new genomes, including recently the first genome of a gymnosperm, are being published every month. At the same time, novel technological challenges have emerged, particularly in computational data analysis, data distribution and storage. The premiere of the Plant Genomics Congress in the well-connected London-Heathrow Marriott Hotel brought together scientists and technology providers from different disciplines, while it maintained a personal atmosphere with about 240 participants. The conference focused on NGS technologies, plant genomic case studies, bioinformatics and functional genomics. Here, we highlight some of the exciting work presented at this conference.

Longer, deeper, cheaper - the opportunities and challenges of sequencing-enabled science

The breathtaking pace of development in sequencing technology opens new horizons, but simultaneously leads to new challenges. These issues were reflected by many talks and were addressed specifically in panel discussions. While there seemed to be a consensus among the panelists and audience that DNA sequences are never too long or too cheap, panelist Michel Morgante

(University of Udine, Italy) stated that science is currently way behind technology, a fact clearly illustrated by the missing heritability problem and many other bottlenecks that are currently faced. Some other obstacles and potential solutions were discussed in various sessions concerning the challenges caused by the flood of sequencing data. While many of those issues can be addressed using sophisticated informatics infrastructure and process workflows, as demonstrated in talks from bioinformatics unit leaders of multiple institutions, it is still a matter of debate which data should be stored long-term and what the best technology for storage is. Most importantly, questions about funding of costs for data archiving and distribution are not yet fully resolved in many institutions, countries and funding schemes.

Another important challenge of sequencing-enabled science is the development and application of appropriate analyses methods. As the sequencing of mRNA-derived cDNA libraries (RNA-seq) becomes more popular for transcriptome analyses, comprehensive in-depth evaluation of the power of RNA-seq will become very important. Etienne Delannoy (URGV, INRA, France) presented his work on comparing a state-of-the-art two-color microarray with RNA-seq data. Most interestingly, while the correlation between the reads of different sequencing runs was better between RNA-seq replicates than between microarrays, the power of detecting differentially expressed genes was lower than that of the microarrays, due to limitations of the currently used test statistics.

Other ongoing challenges in RNA-seq include NGS library preparation protocols. Tamas Dalmay (University of East Anglia, UK) showed how preference of RNA ligase for certain hairpin structures can generate an adapter-dependent bias in the generation of small RNA-seq libraries. Notably, a signature of this bias is present in many previously published microRNA datasets. Fortunately, simple modifications in adapter sequences can overcome this problem, and thereby allow a more sensitive and quantitative detection of small RNAs.

Genomic giants - tackling the daunting wheat genome

The gigantic, hexaploid 17 Gb genome of wheat is still a daunting challenge for sequencing technologies, but Michael Bevan (John Innes Centre, UK) described how a

*Correspondence: kkaufman@uni-potsdam.de; wolfgang.busch@gmi.oeaw.ac.at
¹Institute for Biochemistry and Biology, Potsdam University, 14476 Potsdam-Golm, Germany

²Gregor Mendel Institute, Austrian Academy of Sciences, Vienna 1030, Austria

first step towards elucidating the wheat gene content has been made by combining 454 shotgun sequencing, sequencing of purified chromosome arms and high-resolution synteny maps, as published last year. Odd-Arne Olsen (UMB, Norway) reported further progress made by the International Wheat Genome Sequencing consortium based on the sequencing of dissected chromosome arms. A survey sequence of chromosome 7B that includes a gene catalog, virtual order of genes and markers will be released this year, while a BAC anchored and ordered sequence is the goal in the longer term.

A new BAC-based strategy that was introduced by Hélène Berges (INRA - CNRGV, France) promises to alleviate the issues with complex plant genomes. With such a non-gridded BAC approach combined with 454 sequencing, map-based and positional cloning of loci in wheat that confer resistance against pathogens could be realized in relatively short time spans.

The power of genome synteny in unraveling large crop genomes was pointed out by Klaus Mayer (MIPS, Helmholtz Center Munich, Germany). Chromosome survey sequence assemblies can be integrated using a 'GenomeZipper' approach that makes use of synteny to other sequenced grass genomes. The resulting gene maps can then be assigned to the different wheat sub-genomes that make up the hexaploid wheat genome. Subsequent meta-analysis and comparative analysis will shed light on the impact of polyploidization/hybridization events on gene content in bread wheat.

Tasty genomes - tomato and grape

Yielding colorful and tasty fruits, tomato and grape are tasty and colorful species that have long been subject to intensive breeding processes that have given rise to a large variety of fruit properties. Fruit quality traits in these species are of great economic importance. Mathilde Causse (INRA Avignon, France) reported association studies using 180 cherry tomato lines for 70 metabolites, revealing multiple significant associations underlying different metabolome states. To aid the mapping of causal genetic loci for fruit traits, other new genetic resources are being set up, such as a MAGIC tomato line population from eight divergent, fully sequenced parental strains for which phenotypes, metabolomes, proteomes and transcriptomes have been recorded in great detail.

Another approach to reveal the secrets of tomato taste was introduced by Richard Visser (Wageningen UR Plant Breeding, Wageningen University & Research Centre, Netherlands): the 150 Tomato Genome Project will include the sequencing of old varieties, land races and wild accessions. Touching on a common theme at the conference, the speaker also emphasized that modern crop breeding programs require the integration of different types of -omics data, thus creating new challenges for

data storage and analysis pipelines. Methods and software packages are now becoming available (and also for tetraploid species, which have more complex genetics).

Specific metabolites are of great interest in grape. In particular, polyphenols are among the strongest determinants of wine quality and have been associated with health benefits. Michel Morgante approached intra-species variation of grapevine using the PAN genome concept to differentiate the core genome, which is contained in all grape varieties, from the private, variety-specific genomes. With that concept in view, Alberto Ferrarini from the Massimo Delle Donne lab (University of Verona, Italy) identified genes that underlie high polyphenol content using Tannat, the red wine variety with the highest polyphenol content. Interestingly, 'private' (variety-specific) genes contribute much more strongly to polyphenol pathway expression than core genes. This is a strong argument for a switch to a more PAN genome-centered approach for genetic mapping, since many current 'reference genome'-centered approaches are limited to the core genome.

Conclusions

The research presented at the Plant Genomics Conference in London provided an informative snapshot of major current research directions in the field that encompass efforts to decipher the genomes of complex crop genomes, together with functional genomics approaches to understand the fundamental activities of plant genomes and their applications in plant breeding. In light of the rapid ongoing developments in sequencing technologies, it becomes ever more important to ensure that biology-driven research questions, experimental design, analytical rigor and careful data interpretation keep pace with data production. Conferences such as this provide important platforms to enable discussions and collaborations among experts from different areas in genomics, to raise the awareness of common aims and challenges, and eventually to move the field forward.

Abbreviations

NGS, next-generation sequencing.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Thomas Friese (Gregor Mendel Institute) for proofreading the manuscript.

Published: 27 June 2013

doi:10.1186/gb-2013-14-6-308

Cite this article as: Kaufmann K, Busch W: Plant genomics: from weed to wheat. *Genome Biology* 2013, **14**:308.