Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Plant genome sequencing — applications for crop improvement

Marie E Bolger[1], Bernd Weisshaar[2], Uwe Scholz[3], Nils Stein[3], Björn Usadel[1,4] and Klaus FX Mayer[5]

It is over 10 years since the genome sequence of the first crop was published. Since then, the number of crop genomes sequenced each year has increased steadily. The amazing pace at which genome sequences are becoming available is largely due to the improvement in sequencing technologies both in terms of cost and speed. Modern sequencing technologies allow the sequencing of multiple cultivars of smaller crop genomes at a reasonable cost. Though many of the published genomes are considered incomplete, they nevertheless have proved a valuable tool to understand important crop traits such as fruit ripening, grain traits and flowering time adaptation.

Addresses
[1] RWTH Aachen University, IBMG Institute for Botany and Molecular Genetics, Aachen, Germany
[2] CeBiTec, Department of Biology, Bielefeld University, Bielefeld, Germany
[3] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Stadt Seeland (OT) Gatersleben, Germany
[4] IBG-2: Plant Sciences, Institute for Bio- and Geosciences, Forschungszentrum Jülich, Jülich, Germany
[5] MIPS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany

Corresponding author: Usadel, Björn (usadel@bio1.rwth-aachen.de)

Current Opinion in Biotechnology 2014, **26**:31–37

This review comes from a themed issue on **Food biotechnology**

Edited by **Mattheos AG Koffas** and **Jan Marienhagen**

For a complete overview see the Issue and the Editorial

Available online 21st September 2013

0958-1669/$ – see front matter,

http://dx.doi.org/10.1016/j.copbio.2013.08.019

Introduction

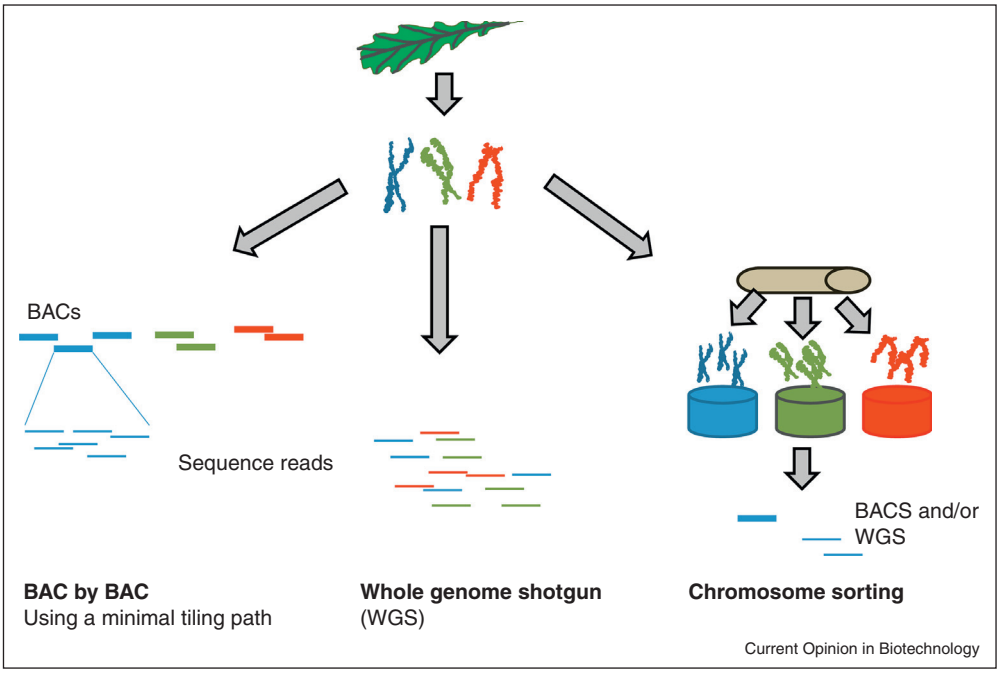
Sequencing the Arabidopsis model plant genome in 2000 [1] was a major milestone not only for plant research but also for genome sequencing. It was among the earliest genomes from multicellular organisms to be completed, and was sequenced by a large multinational consortium to cope with this daunting effort. This hitherto unprecedented resource invigorated and accelerated plant research. The approach chosen relied on overlapping bacterial artificial chromosomes (BAC) clones that represent a minimal tiling path to cover each chromosome arm (Figure 1 left panel). The BAC sequences were individually assembled and arranged according to the physical map, creating a very high quality genome sequence. The prohibitively high effort and time associated with this approach limited its applicability to a few genomes. Nevertheless, only two years later, the approximately three times larger genome of the first crop plant, rice [2,3], was elucidated using the same BAC based approach.

Poplar was the next crop genome sequenced and employed the emerging whole genome shotgun (WGS) strategy [4]. Here, the genome is randomly broken down into smaller pieces which are then sequenced and subsequently assembled (Figure 1 middle panel). Whilst this strategy significantly lowers preparation time and cost, the data are more difficult to assemble and typically results in a more fragmented genome sequence. The analysis of further crop genomes employed both strategies, balancing the various disadvantages and advantages. In 2007, the grapevine genome sequence was published independently by two different groups, both based on the WGS strategy. Whilst one team focused on an inbred line [5], the other analysed a heterozygous genotype used for actual wine production [6]. These data contributed to evidence suggesting that during centuries of vegetative multiplication and exchange of cultivars, a large number of synonyms and homonyms were generated. These may or may not be genetically identical, or may show clonal relationship with individual, phenotypically highly relevant mutations. To distinguish them, it is necessary to analyse multiple genomes from a single species. Given the commercial importance of grapevines, the genome sequence is exploited to understand grape development, identify genetic factors affecting wine quality and produce genotypes that withstand the enormous amount of fungicide required to allow harvest, and all that whilst maintaining the individual quality of grapevine cultivars.

Next generation sequencing comes to help — but sometimes not

Although WGS reduced the time and effort required, genome sequence generation was still expensive and time-consuming due to the high cost of Sanger sequencing. The adaptation of Next Generation Sequencing (NGS) improved the output/cost ratio of genome sequencing dramatically. NGS technology covers a broad range of 'post Sanger' approaches, which sequence multiple DNA fragments in parallel to yield a much larger number of sequence reads, but generally of shorter length and lower quality. The first NGS technology, 454, was based on pyrosequencing and was initially suited mainly to bacterial-sized genomes due to low data yield and relatively short reads. Subsequent improvements in the

Figure 1



Sequencing and assembly strategies. For BAC by BAC sequencing (left panel), the genome is split into a minimal tiling path consisting of BACs which are then sequenced. In WGS (middle panel) the whole genome is sheared, sequenced and assembled. A relatively new technique, chromosome sorting (right panel), is used to reduce the genomic complexity. The purified chromosomes can then be used for BAC by BAC or WGS sequencing.

technology enabled 454 sequencing to become applicable also to more complex genomes in conjunction with Sanger sequencing or alone.

Another early NGS technology was the sequencing by synthesis based approach offered by Illumina. This approach generated much greater data yields than 454, albeit at the cost of even shorter reads. Since its introduction both read count and in particular read length improved dramatically. Illumina sequencing was adopted to sequence the cucumber [7], where it was combined with conventional Sanger sequencing in a hybrid approach, showing that this technology is a feasible approach to plant genome sequencing.

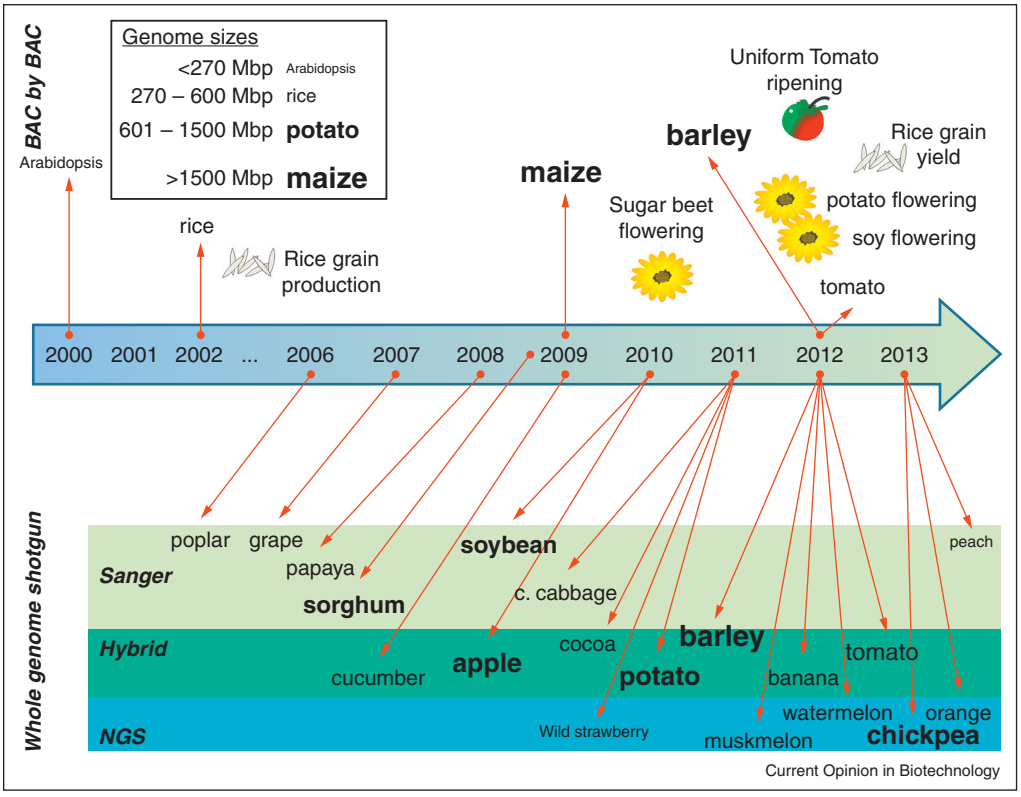
The application of NGS to plant genomes then became an increasingly strong trend. 454, which was applied in combination with Sanger sequencing for the apple genome [8], replaced Sanger sequencing as the primary data source for the Cocoa genome [9] and muskmelon [10], although Sanger sequenced BAC-end data were still required to acquire long-distance structural information. The development of moderate length (3–20 kbp) paired-end library preparation techniques also helped compensate for the shorter read length of NGS compared to Sanger.

It was not however until the woodland strawberry [11•] that the first plant genome was sequenced using next generation sequencing alone, combing the 454, Illumina and SOLID platforms. More recently, Illumina sequencing emerged as the dominant NGS platform for genome sequencing, providing the bulk of the data for recent genomes such as Chinese cabbage [12], potato [13], banana [14], chickpea [15], orange [16], pigeonpea [17] and watermelon [18] and even the large genome of spruce [19•] (see Figure 2 for a timeline putting genome sequences into order).

Notably absent from these NGS success stories were the Triticeae which include wheat, barley and rye and are immensely important crops for animal feed and human nutrition [20]. This is largely because Triticeae genomes are notoriously difficult to access due to genome size and the underlying complexity in terms of repetitive sequences. With approximate sizes of 5 Gb, 8 Gb and 17 Gb for barley [21•], rye and bread wheat, respectively, they exceed mammalian and other crop genomes by far. In addition, bread wheat has an allohexaploid genome structure with individual subgenomes being very similar both at structural as well as at sequence level which poses an additional barrier in accessing this genome.

For barley, wheat and rye, 80–90% of the genomes consist of repetitive sequences distributed throughout the genome in highly similar copies and organized in long repetitive arrays. For the moment, this repetitive 'matrix'

Figure 2



Crop and plant genomes and their application. The figure gives the approximate timeline of when crop genomes were sequenced along with the underlying techniques (Figure 1) and sequencing strategy used. Hybrid strategies which use BAC by BAC and WGS are indicated by the placement of a genome twice. Also note that the distinction between pure NGS and Hybrid sequencing is sometimes arbitrary as many genome projects rely on previously generated Sanger sequences. In addition, some major applications are marked by symbols: Grains for an improvement in grain quality, a flower for flowering time and a tomato for a tomato ripening trait.

cannot be bridged by the read length of NGS techniques, restricting assemblies to gene containing and low copy regions and leaving intergenic and repetitive regions largely un-assembled.

Still, remarkable progress in accessing the genomes and 'gene-omes' (i.e. the structured gene component of a complex genome) has been made recently based on Triticeae chromosome sorting (Figure 1 right hand panel) [22•]. This method allows the purification of individual chromosomes which are subsequently used as a template for shotgun sequencing or the construction of BAC libraries. In case of barley, this approach has given access to the 'gene-ome' (as well as repeat content) of all seven barley chromosomes individually. Classical physical map construction and sequencing of BACs arranged in a tiling path are still required to unlock Triticeae genomes completely. A powerful shortcut to suitable genome data is the approximate ordering and positioning of genes by using synteny information from related grass (Poaceae) genomes. On the basis of the complete sequenced model of grass genomes, namely rice [2,3], *Brachypodium distachyon* [23] and sorghum [24], synteny data allow the deduction of gene order and finally an 'assembled gene-ome'. This approach has been pioneered in barley and has since been applied to ryegrass [25•], wheat [26] as well as rye (Martis *et al.*, under revision).

Sequencing your own plant genomes?

Many of the major crops have been sequenced in recent years, although quality and completeness vary. These genome sequences provide an unprecedented resource which can be exploited in numerous ways. Improvements in NGS platforms now allow plant genomes of up to about 1 Gbp to be coarsely assembled in a matter of months at a modest cost given that they are mostly homozygous and do not feature too many 'difficult' long repetitive elements. Whilst insufficient for a high quality assembly of a new species, these data are sufficient for re-sequencing a close relative of an already sequenced species, or for mapping and identification of a novel trait. With the advent of the newest personal sequencers, such as the MiSeq (Illumina Corp: 2× 300 bp read length; see

the excellent field guide for an overview of NGS system capabilities [27••]), such NGS capabilities are now turning classical labs into genome labs.

Despite the remarkable progress in sequencing technologies, highly repetitive and also heterozygous or polyploid genomes still remain an issue. Current protocols for generating long NGS libraries, which are necessary to span repetitive elements frequently found in plants, are laborious and expensive, and in addition often give mixed results. The recent acquisition of Moleculo by Illumina promises to deliver up to 10 kbp pseudo-reads which would greatly facilitate genome assembly and would be complementary to other long range sequencing. Additionally, third generation platforms are expected to provide longer reads which can span repetitive elements. Typically, these technologies do not require prior DNA amplification, thereby removing the clonal amplification bias that NGS suffers from. Currently, the only third generation sequencer available on the market is from PacBio. Though still relatively new, the PacBio realtime sequencing technology promises mean read length of several thousand bp, with the drawback of error rates much higher than for Illumina sequencing. Nevertheless, this technology already enables the assembly of bacterial genomes to a single contig finished state in only a few days [28] and the resulting reads can be corrected using, for example Illumina reads [29].

What to do with all these data — is one genome enough?

Although generally considered to be part of the genome sequence, the generation of good gene models as a basis for genome annotation is also a challenge. At present, gene calling that includes alternative splicing and small RNA genes is best achieved by combining intrinsic evidence (*de novo* gene finding) with extensive extrinsic evidence in the form of RNA data obtained from NGS platforms. Also analysing close relatives or resequencing multiple cultivars will further improve the depth by which the genomes can be analysed. As an example, the resequencing of >1000 wild and cultivated rice accessions has been carried out and revealed thousands of genes with lower diversity in cultivated rice which helped localize the origin of rice domestication [30••,31]. When exploiting genomic data, special emphasis can be directed to genes or gene families of interest (e.g. resistance gene analogues) and these often vary among cultivars. Also, domestication genes can be moved in the focus since they often have an impact on the optimization of plant architecture which is another major component of yield. Furthermore, the analysis of copy number variation among and between species, which will become feasible on the basis of a reference genome sequence and resequencing data, will contribute to the understanding of the mechanisms of heterosis [32].

Resequencing leads to more informative and optimized marker applications, enables and improves smart breeding, and allows functional allele mining and SNP value determination (coding SNPs opposed to intergenic SNPs). By analysing more than a million SNPs per genotype using NGS in a single measurement, breeders are able to use approaches like genomic selection or sequence based association studies. These developments have the potential to cause conceptual changes in hybrid breeding, and to perform genotyping by (population) sequencing [33] as standard.

Genomes directly impacting crop research

Whilst there are still many crops to have their genomes sequenced and many improvements to current genome assemblies to be undertaken, it is already clear that crop genomes have made a massive impact in a variety ways (see Figure 2 for some of the highlights shown here). One of the overwhelming uses of genomes is in the availability of high-density molecular markers which can be used to quickly map agronomically desirable traits and to identify candidate genes within a region of interest. This can be seen in the increased use of genomes for QTL mapping of desirable traits even if only some sequence is available [34]. These traits, once characterized, can then be bred into elite varieties. High throughput genotyping by sequencing are options for efficient marker assisted breeding, for genomic selection, and for efficient management of genetic resources at a higher resolution than is otherwise possible [35].

Once the rice genome became available, it was immediately used to help elucidate a major QTL for rice grain production which was found to be a cytokinin oxidase [36]. Interestingly, almost ten years later a transcription factor controlling the expression of the gene was identified as DST [37]. This gene encodes a transcription factor which had previously been shown to also regulate drought and salt tolerance in rice [38]. Similarly, availability of the maize genome sequence [39] made it possible to develop powerful haplotype maps [40] and to find QTLs for biomass and bioenergy using whole genome and metabolic prediction [41]. These examples illustrate how a genome sequence can impact data integration.

Within the Asterid branch, the tomato genome [42•] came in a timely manner to identify an esterase responsible for differences in volatile ester content in different tomato species [43]. A gene underlying the uniformly ripening locus in tomato was also identified which turned out to be a Golden 2-like Transcription Factor which determines chlorophyll distribution in unripe fruits [44••]. The genome together with the draft genome of its wild relative *Solanum pennellii* illuminated the evolution of the terpene biosynthesis [45•]. These new insights will undoubtedly spur breeding towards improved quality crops.

Apart from crop yield (seed, grain, etc.), the control of flowering and maturation time is a recurring interest, as it represents an agronomical important trait for the adaptation to different photoperiod regimes and geographic latitude. Thus, it is not surprising that the soybean genome was used to unravel the maturity locus E1 which has a major impact on flowering time [46]. This represents an important trait for the adaptation to photoperiods. Similarly, the potato genome sequence has helped in the identification of a transcription factor regulating plant maturity and life cycle [47••]. The sugar beet genome was used to determine the biology of its flowering time control [48]. For the latter, the hope is to create a 'winter beet' which avoids bolting induction during the winter, thereby providing a prolonged growing season with increased yields. Finally, in a large genome wide association study in rice, flowering time was one of the many identified traits for which a QTL was found [49•].

Conclusion

Crop genome sequences, even at the current levels of completeness, have had a major impact on crop research/improvement in a relatively short time. The 'success stories' indicate that additional breakthroughs are to be expected when sequencing multiple cultivars or land-races. Since improvements in NGS in terms of library preparation and sequence runs have seen a rapid development in the last years, it will only be a matter of time until sequencing smaller genomes for QTL and genome wide association studies will become commonplace.

Together with deep phenotyping platforms which promise to overcome the phenoytping bottleneck [50,51], we can expect an even faster elucidation of numerous QTL, but will be challenged with the sheer magnitude of data available.

Acknowledgement

We thank Anthony Bolger for critical reading of the manuscript and for help with formatting the references. Also BMBF grant 0315961 for funding of BU.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. AGI Initiative: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana***. *Nature* 2000, **408**:796-815.

2. Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*)**. *Science* 2002, **296**:79-92.

3. Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H *et al.*: **A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*)**. *Science* 2002, **296**:92-100.

4. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**:1596-1604.

5. Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C *et al.*: **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**. *Nature* 2007, **449**:463-467.

6. Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J *et al.*: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety**. *PLoS ONE* 2007, **2**:e1326.

7. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P *et al.*: **The genome of the cucumber, *Cucumis sativus* L.**. *Nat Genet* 2009, **41**:1275-1281.

8. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D *et al.*: **The genome of the domesticated apple (*Malus* × *domestica* Borkh.)**. *Nat Genet* 2010, **42**:833-839.

9. Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN *et al.*: **The genome of theobroma cacao**. *Nat Genet* 2011, **43**:101-108.

10. Garcia-Mas J, Benjak A, Sanseverino W, Bourgeois M, Mir G, González VM, Hénaff E, Câmara F, Cozzuto L, Lowy E *et al.*: **The genome of melon (*Cucumis melo* L.)**. *Proc Natl Acad Sci U S A* 2012, **109**:11872-11877.

11. • Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP *et al.*: **The genome of woodland strawberry (*Fragaria vesca*)**. *Nat Genet* 2011, **43**:109-116.

Even though the genome size is small, this publication demonstrated that complete NGS sequencing of plant genomes is a possibility. It also provides an insightful decision: Sequencing a wild relative rather than the polyploid domesticated crop to learn about the gene space rather than wait and try and sequence the domesticated crop.

12. Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F *et al.*: **The genome of the mesopolyploid crop species *Brassica rapa***. *Nat Genet* 2011, **43**:1035-1039.

13. Potato Genome Consortium, Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R *et al.*: **Genome sequence and analysis of the tuber crop potato**. *Nature* 2011, **475**:189-195.

14. D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M *et al.*: **The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants**. *Nature* 2012, **488**:213-217.

15. Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B *et al.*: **Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement**. *Nat Biotechnol* 2013, **31**:240-246.

16. Xu Q, Chen LL, Ruan X, Chen D, Zhu A, Chen C, Bertrand D, Jiao WB, Hao BH, Lyon MP *et al.*: **The draft genome of sweet orange (*Citrus sinensis*)**. *Nat Genet* 2013, **45**:59-66.

17. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MTA, Azam S, Fan G, Whaley AM *et al.*: **Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers**. *Nat Biotechnol* 2012, **30**:83-89.

18. Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z *et al.*: **The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions**. *Nat Genet* 2013, **45**:51-58.

19. • Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A *et al.*: **The Norway spruce genome sequence and conifer genome evolution**. *Nature* 2013, **497**:579-584.

Assembling a gigantic genome (20 GB) was performed using WGS including some fosmid pool sequences. In addition, low coverage sequencing of other conifers was included to provide some insights into the conifer evolution. Furthermore data were provided on how to proceed with biotechnological biomass improvement. This is particularly meaningful as less than 30k protein coding genes were found providing an angle for biotechnology driven improvement.

20. Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, Eversole K, Appels R: **Integrating cereal genomics to support innovation in the triticeae**. *Funct Integr Genomics* 2012, **12**:573-583.

21. • IBGS Consortium, Mayer KFX, Waugh R, Brown JWS, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K *et al.*: **A physical, genetic and functional sequence assembly of the barley genome**. *Nature* 2012, **491**:711-716.
This manuscript describes the complex dual approach that was necessary to tackle the genome sequence of the barley genome. It demonstrates very well, that approaches for medium sized genomes need to be rethought when dealing with large grass genomes.

22. • Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H *et al.*: **Unlocking the barley genome by chromosomal and comparative genomics**. *Plant Cell* 2011, **23**:1249-1263.
One of several applications of chromosome sorting, a promising technique which can be coupled with other 'classical' approaches and shows promise in helping reducing the complexity of even the largest genomes.

23. International Brachypodium Initiative: **Genome sequencing and analysis of the model grass *Brachypodium distachyon***. *Nature* 2010, **463**:763-768.

24. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A *et al.*: **The sorghum bicolor genome and the diversification of grasses**. *Nature* 2009, **457**:551-556.

25. • Pfeifer M, Martis M, Asp T, Mayer KFX, Lübberstedt T, Byrne S, Frei U, Studer B: **The perennial ryegrass genomezipper: targeted use of genome resources for comparative grass genomics**. *Plant Physiol* 2013, **161**:571-582.
What do you do if you cannot sequence the full genome of a grass yet? Extensive synteny using a 'GenomeZipper' approach is used here to get a useful first look.

26. Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J *et al.*: **Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content**. *Plant J* 2012, **69**:377-386.

27. •• Glenn TC: **Field guide to next-generation DNA sequencers**. *Mol Ecol Resour* 2011, **11**:759-769.
Provides an overview of current next generation sequencing techniques including read length, pricing per base as well as machine prices allowing one to perform back of the envelope calculations on how much a genome would be. The accompanying online resource has been updated twice already.

28. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE *et al.*: **Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data**. *Nat Methods* 2013, **10**:563-569.

29. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED *et al.*: **Hybrid error correction and de novo assembly of single-molecule sequencing reads**. *Nat Biotechnol* 2012, **30**:693-700.

30. •• Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, Dong Y, Gutenkunst RN, Fang L, Huang L *et al.*: **Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes**. *Nat Biotechnol* 2012, **30**:105-111.
Having a good reference genome is often only a start, which is demonstrated in this publication detailing the resequencing of 50 rice cultivars to identify SNPs and their direct application.

31. Huang X, Kurata N, Wei X, Wang ZX, Wang A, Zhao Q, Zhao Y, Liu K, Lu H, Li W *et al.*: **A map of rice genome variation reveals the origin of cultivated rice**. *Nature* 2012, **490**:497-501.

32. Schnable PS, Springer NM: **Progress toward understanding heterosis in crop plants**. *Annu Rev Plant Biol* 2013, **64**:71-88.

33. Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q: **Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing**. *Proc Natl Acad Sci U S A* 2010, **107**:10578-10583.

34. Paux E, Sourdille P, Mackay I, Feuillet C: **Sequence-based marker development in wheat: advances and applications to breeding**. *Biotechnol Adv* 2012, **30**:1071-1088.

35. Cabezas JA, Ibáñez J, Lijavetzky D, Vélez D, Bravo G, Rodríguez V, Carreño I, Jermakow AM, Carreño J, Ruiz-García L *et al.*: **A 48 snp set for grapevine cultivar identification**. *BMC Plant Biol* 2011, **11**:153.

36. Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M: **Cytokinin oxidase regulates rice grain production**. *Science* 2005, **309**:741-745.

37. Li S, Zhao B, Yuan D, Duan M, Qian Q, Tang L, Wang B, Liu X, Zhang J, Wang J *et al.*: **Rice zinc finger protein dst enhances grain production through controlling gn1a/osckx2 expression**. *Proc Natl Acad Sci U S A* 2013, **110**:3167-3172.

38. Huang XY, Chao DY, Gao JP, Zhu MZ, Shi M, Lin HX: **A previously unknown zinc finger protein, dst, regulates drought and salt tolerance in rice via stomatal aperture control**. *Genes Dev* 2009, **23**:1805-1817.

39. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA *et al.*: **The b73 maize genome: complexity, diversity, and dynamics**. *Science* 2009, **326**:1112-1115.

40. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J *et al.*: **A first-generation haplotype map of maize**. *Science* 2009, **326**:1115-1117.

41. Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L, Melchinger AE: **Genomic and metabolic prediction of complex heterotic traits in hybrid maize**. *Nat Genet* 2012, **44**:217-220.

42. • The Tomato Genome Consortium: **The tomato genome sequence provides insights into fleshy fruit evolution**. *Nature* 2012, **485**:635-641.
The tomato was initially intended to be sequenced using a BAC by BAC approach. However, due to the advances in NGS, the consortium decided upon a hybrid approach, thus demonstrating the power of NGS for medium scale genomes.

43. Goulet C, Mageroy MH, Lam NB, Floystad A, Tieman DM, Klee HJ: **Role of an esterase in flavor volatile variation within the tomato clade**. *Proc Natl Acad Sci U S A* 2012, **109**:19009-19014.

44. •• Powell ALT, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons C, Fernández-Muñoz R, Vicente A *et al.*: **Uniform ripening encodes a golden 2-like transcription factor regulating tomato fruit chloroplast development**. *Science* 2012, **336**:1711-1715.
This paper describes how a breeding selection may have resulted in a loss of function gene that negatively impacts fruit quality. They use two interspecific populations to map the location of the causal gene.

45. • Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schäfer P, Kudrna D, Wing RA, Bolger AM *et al.*: **Evolution of a complex locus for terpene biosynthesis in solanum**. *Plant Cell* 2013, **25**:2022-2036.
This paper describes how even different species genomes were used to elucidate the evolution of a complex Locus for a secondary metabolite, highlighting that the multispecies approach can be very interesting in the future.

46. Xia Z, Watanabe S, Yamada T, Tsubokura Y, Nakashima H, Zhai H, Anai T, Sato S, Yamazaki T, Lü S *et al.*: **Positional cloning and characterization reveal the molecular basis for soybean maturity locus e1 that regulates photoperiodic flowering**. *Proc Natl Acad Sci U S A* 2012, **109**:E2155-E2164.

47. •• Kloosterman B, Abelenda JA, Gomez MdMC, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S *et al.*: **Naturally occurring allele diversity allows potato cultivation in northern latitudes**. *Nature* 2013, **495**:246-250.
Potatoes having evolved from the Andean region of South America should be short day dependent to form tubers. Here the authors use two potato populations, one wild and one domesticated, to identify a gene which explain the long day tuberization onset seen in European potato types.

48. Pin PA, Benlloch R, Bonnet D, Wremerth-Weich E, Kraft T, Gielen JJL, Nilsson O: **An antagonistic pair of ft homologs mediates the control of flowering time in sugar beet**. *Science* 2010, **330**:1397-1400.

49. • Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C *et al.*: **Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm**. *Nat Genet* 2012, **44**:32-39.

This interesting article unlocks the power of genome wide association studies for grain related traits as well as flowering time relying on nearly thousand rice varieties.

50. Fiorani F, Rascher U, Jahnke S, Schurr U: **Imaging plants dynamics in heterogenic environments**. *Curr Opin Biotechnol* 2012, **23**:227-235.

51. Fiorani F, Schurr U: **Future scenarios for plant phenotyping**. *Annu Rev Plant Biol* 2013, **64**:267-291.