Plant Biotechnology Journal

Plant Biotechnology Journal (2016) **14**, pp. 1086–1094 doi: 10.1111/pbi.12456

Review article

Understanding and utilizing crop genome diversity via high-resolution genotyping

Kai Voss-Fels and Rod J. Snowdon*

Department of Plant Breeding, IFZ Research Centre for Biosystems, Land Use and Nutrition, Justus Liebig University, Giessen, Germany

Received 2 June 2015;
revised 17 July 2015;
accepted 20 July 2015.
**Correspondence* (Tel +49 641 9937420;
fax +49 641 9937429;
email rod.snowdon@agrar.uni-giessen.de)

Summary

High-resolution genome analysis technologies provide an unprecedented level of insight into structural diversity across crop genomes. Low-cost discovery of sequence variation has become accessible for all crops since the development of next-generation DNA sequencing technologies, using diverse methods ranging from genome-scale resequencing or skim sequencing, reduced-representation genotyping-by-sequencing, transcriptome sequencing or sequence capture approaches. High-density, high-throughput genotyping arrays generated using the resulting sequence data are today available for the assessment of genomewide single nucleotide polymorphisms in all major crop species. Besides their application in genetic mapping or genomewide association studies for dissection of complex agronomic traits, high-density genotyping arrays are highly suitable for genomic selection strategies. They also enable description of crop diversity at an unprecedented chromosome-scale resolution. Application of population genetics parameters to genomewide diversity data sets enables dissection of linkage disequilibrium to characterize loci underlying selective sweeps. High-throughput genotyping platforms simultaneously open the way for targeted diversity enrichment, allowing rejuvenation of low-diversity chromosome regions in strongly selected breeding pools to potentially reverse the influence of linkage drag. Numerous recent examples are presented which demonstrate the power of next-generation genomics for high-resolution analysis of crop diversity on a subgenomic and chromosomal scale. Such studies give deep insight into the history of crop evolution and selection, while simultaneously identifying novel diversity to improve yield and heterosis.

Keywords: genome diversity, high-throughput genomics, selection, breeding.

Genotyping-by-sequencing in crop plants for discovery of DNA sequence diversity

Novel genomic technologies have achieved exceeding importance for modern crop improvement and are undergoing continual further development in terms of efficiency and costs (Poland and Rife, 2012). Fifteen years after the complete genome sequence of the model plant *Arabidopsis thaliana* was decoded (genome size: 125 mega base pairs) (The Arabidopsis Genome Initiative, 2000), followed shortly afterwards by rice (430 Mbp) (Goff *et al.*, 2002), the rapid advancement of next-generation sequencing (NGS) platforms has today provided reference sequences for the large, complex genomes of many important crop species, such as maize (2500 Mbp) (Schnable *et al.*, 2009), sorghum (730 Mbp) (Paterson *et al.*, 2009), soya bean (1115 Mbp) (Schmutz *et al.*, 2010), potato (850 Mbp) (Xu *et al.*, 2011), barley (5100 Mbp) (Mayer *et al.*, 2012) and rapeseed (1200 Mbp) (Chalhoub *et al.*, 2014). Even in the huge hexaploid genome of bread wheat (17 000 Mbp), a combination of flow cytometry and syntenic mapping with next-generation sequencing technologies has enabled generation of a chromosome-based draft genome sequence (International Wheat Genome Sequencing Consortium, 2014). Pan-genome diversity analysis based on assemblies of crop relatives provides unprecedented insight into the gene diversity available in secondary crop gene pools. Recent pan-genome sequencing studies in maize (Lu *et al.*, 2015) and soya bean (Li *et al.*, 2014) describe the valuable contribution to pan-genomic variation to phenotypic variation for important adaptive traits. Identification and implementation of such adaptive potential using high-resolution genotyping may be a key to targeted rejuvenation of depleted phenotypic diversity in response to climate change, for example.

High-resolution genome information is being increasingly used by plant breeders to characterize germplasm, to identify genes that underlie important agronomic traits or to estimate the breeding values of individuals in breeding programmes in order to accelerate the selection of improved varieties (Varshney *et al.*, 2014). Either with or without a completed reference sequence, the development of reduced-representation genotyping-by-sequencing (GBS) methods has opened the way to use NGS technologies for high-throughput genomic resequencing, even in large plant populations, at a constantly shrinking price (Deschamps *et al.*, 2012). In maize, for example, Romay *et al.* (2013) used the GBS approach described by Elshire *et al.* (2011) to generate almost 700 000 genomewide SNPs in a panel of 2815 diverse inbred lines from globally distributed breeding programmes. Comprehensive sequencing data sets of this kind enable extremely high-resolution evaluation of genetic diversity and population structure, providing insight into the history of recombination and allelic diversity throughout different breeding pools (Qian *et al.*, 2014; Voss-Fels *et al.*, 2015). The broad applicability of GBS techniques for genetic analysis has been

successfully demonstrated in numerous important crops, for example rice (Spindel *et al.*, 2013), barley (Elshire *et al.*, 2011), potato (Uitdewilligen *et al.*, 2013), wheat (Poland *et al.*, 2015) and soya bean (Jarquín *et al.*, 2014). In combination with quantitative phenotype analysis in segregating populations, NGS methods also provide a powerful basis for rapid mapping and identification of genes underlying quantitative traits (e.g. Abe *et al.*, 2012; Gao *et al.*, 2013; Schneeberger *et al.*, 2009).

Targeting of genic variants associated with agronomic traits

Reduced-representation sequencing approaches involving exome capture or transcriptome sequencing enable targeted identification of molecular variants in protein-coding genome regions (Ku *et al.*, 2012). Reference-based assembly of target-captured or transcriptome sequence data in a test panel can allow rapid discovery of hundreds of thousands of molecular variants in gene-coding regions. High-density genic polymorphism data generated via such techniques can be applied to quantitative trait dissection, marker-assisted breeding, genomic selection or for high-resolution exploration of genetic resources (Bolger *et al.*, 2014). Custom design of capture probes or tailed sequencing primers provides a flexibility to target specific chromosome regions harbouring quantitative trait loci, known pathways from related model plants, or candidate genes for traits of interest across a given species of interest. For example, Gholami *et al.* (2012) and Rife *et al.* (2015) describe how tailed PCR can be used to target specific genes for next-generation resequencing in large numbers of individuals, while Schiessl *et al.* (2014) demonstrate the use of bead-based capture technology to resequence a panel of over 30 genes involved in regulating the flowering time pathway. Clarke *et al.* (2013) present an example for resequencing of genetic diversity spanning important meta-QTL in a major crop using a microarray-based capture platform. Harper *et al.* (2012) introduced the concept of associative transcriptomics, in which polymorphic SNP data from transcriptome sequencing in a diversity panel are associated with phenotype variation for QTL identification, and Mascher *et al.* (2013) describe how exome capture sequencing can help in the cost-effective identification of coding sequence variants even in very large genomes like that of barley. Collectively, these methods proved powerful options for mapping and discovery of genes underlying quantitative traits, and development of tightly linked, sequence-based markers for breeding. They also demonstrate the broad diversity of available technological platforms for sequence capture, which enable extremely flexible scaling of resequencing experiments to deal with few to many genes at low cost in large test populations.

Accessing crop diversity with high-density genotyping arrays

Beyond their direct applications for genetic mapping, QTL dissection and characterization of diversity, NGS technologies have also created the basis for the development of high-density SNP genotyping platforms as a high-throughput tool for genetic analysis of large experimental and breeding populations (Ganal *et al.*, 2012). Today, high-capacity SNP arrays are available for a broad range of plant species and are becoming widely used in breeding of major crops like maize (50–600 k SNPs) (Ganal *et al.*, 2011; Unterseer *et al.*, 2014), rice (51.5 k SNPs) (Chen *et al.*, 2014; Zhao *et al.*, 2011), wheat (9 k, 35 k, 90 k and 800 k SNPs) (Cavanagh *et al.*, 2013; Wang *et al.*, 2014; M. Winfield, A. Allen, A. Burridge, G. Barker, H. Benbow, P. Wilkinson, J. Coghill, C. Waterfall, A. Devassi, G. Scopes, T. Webster, F. Brew, C. Bloor, J. King, S. Griffiths, I. King, A. Bentley and K. Edwards *et al.*, unpublished), potato (8.3 k SNPs) (Hamilton *et al.*, 2011), barley (9 k SNPs) (Comadran *et al.*, 2012), soya bean (50 k SNPs) (Song *et al.*, 2013), rapeseed (60 k SNPs), (Edwards *et al.*, 2013) or sorghum (3 k and 90 k SNPs) (Bekele *et al.*, 2013; Wieckhorst *et al.*, 2015). At present, the Infinium® platform from Illumina Inc. (San Diego, CA) and the Axiom® technology from Affymetrix Inc. (Santa Clara, CA) are the most widely used platforms for large-scale SNP genotyping in crop plants (Thomson, 2014). Fixed genotyping chips are often preferred to GBS technologies for scenarios aiming to generate structured data sets of common sequence variants at low cost, with minimal bioinformatic input, for example within an ongoing breeding programme (Varshney *et al.*, 2014). On the other hand, to be effective a fixed SNP genotyping platform must be applicable to a wide range of different genotypes; hence, the alleles of the chosen SNPs must be representative even for diverse germplasm. GBS-based genotyping methods can be more suitable for identifying true, causal genetic variants for phenotypes with a complex genetic architecture, as these are typically influenced in crop species by rare alleles that may not be adequately represented on a SNP array (Huang and Han, 2014). Nevertheless, given the relatively high extent of linkage disequilibrium (LD) throughout the genomes of most crops, SNP markers on a fixed, high-density array are still likely to exhibit genetic associations with phenotypic variation through LD to the causal genes (Wray *et al.*, 2013).

Genome-scale characterization of crop diversity

High-throughput genotyping techniques are an important enabling technology for complex trait dissection by genomewide association studies (GWAS). Detailed molecular characterization of breeding germplasm, providing comprehensive knowledge of population genetic parameters and their relationships to natural and artificial selection for important traits, is a crucial prerequisite for the production of new, improved cultivars. Due to intensive human-mediated selection during plant breeding, modern crop varieties typically exhibit lower levels of genetic variation and biased allele frequency spectra compared to their wild types, especially in chromosome regions that harbour agronomically important loci (Mace *et al.*, 2013). This causes higher levels of overall, chromosome-wise and/or region-specific LD in the respective genomic regions. Agricultural selection furthermore has led to enlarged haplotypes with extended homozygosity, sometimes covering large chromosome segments (Mace *et al.*, 2013; Qian *et al.*, 2014; Voss-Fels *et al.*, 2015).

The enrichment of particular allele variants in gene pools due to directional selection, and the consequential depletion of genetic variation, caused genetic bottlenecks during crop domestication that have resulted in prominent selective sweeps in all major crops (Shi and Lai, 2015). This is particularly troublesome because of linkage drag, the unintentional co-selection of undesirable gene variants that are closely linked to selected loci of interest (Langridge and Fleury, 2011). Because genetic diversity represents the fundamental key to breeding success and a broad variation provides the basis for breeders to select varieties with constantly improving yield performance, these footprints of directional

selection seriously challenge crop improvement, as they lead to deterioration of genetically determined phenotypic variation. High-throughput population genomic analyses can address this dilemma by providing a detailed molecular basis for identification and genomic introgression of novel variation into chromosome segments surrounding directionally selected loci. Using high-resolution tools, we are now able to identify and characterize genome regions in most need of rejuvenation with novel diversity, and utilize genomic selection approaches (Jannink *et al.*, 2010) for the enrichment of depleted gene pools.

Identifying and overcoming signatures of selection

Jiao *et al.* (2012) used genome sequences of 278 maize inbred lines from China and the United States to describe the structural development of the maize genome during domestication and utilization by humans, identifying highly dynamic genetic changes caused by modern breeding. Based on around 4.8 million SNPs present in at least 50% of the population, they identified around 400 different loci that have experienced a selective sweep in different germplasm groups, representing domestication and modern breeding progress. They showed that modern breeding caused an accumulation of rare alleles in elite cultivars, suggesting that the proportion of rare alleles could be used as a selection index in future maize-breeding approaches.

A high-resolution investigation of genome-scale diversity and directional selection in sorghum by Mace *et al.* (2013) used the genome sequences of 44 highly diverse accessions, representative of the diversity spanning the primary gene pool. Strong signatures of selection were identified in different gene pools around major genes related to domestication, eco-geographical adaptation or agricultural traits such as plant height, seed colour and maturity. Investigation of genomewide diversity patterns revealed decreased diversity in landraces and improved germplasm, but discovered untapped genetic variation in related allopatric gene pools (Figure 1; Mace *et al.*, 2013). Similarly, in rice, Huang *et al.* (2012) resequenced 446 diverse accessions of the rice wild relative *Oryza rufipogon*, along with 1083 cultivated *O. indica* and *O. japonica* varieties. This study revealed 55 selective sweeps originating from domestication. Around 8 million SNP markers disclosed extremely high allelic variation in wild rice populations compared to domesticated rice, highlighting genomic target regions for the restoration of genetic diversity in future rice-breeding efforts.

In many crops, the co-selection of undesired loci due to linkage drag has hampered the efficiency of introgression approaches using exotic plant resources, and the resolution with which introgressed DNA segments could be tracked on a molecular level was extremely poor when concepts for marker-assisted backcrossing were first introduced to breeding. Massively parallel genotyping techniques overcome this dilemma and facilitate marker-assisted selection in very early stages of plant development and the breeding cycle. In a simulation study using maize data, Herzog *et al.* (2014) demonstrated the inherent suitability of high-throughput genotyping arrays for targeted introgression of donor chromosome segments into recipient genotypes. Detection of introgressed DNA fragments using detailed molecular marker information facilitates the utilization of exotic germplasm for the targeted restoration of genetic diversity in crop breeding populations, improving our capacity to introduce novel loci into elite cultivars with minimal linkage drag.

Even in complex polyploid genomes, the use of large-scale molecular information from genomewide SNP markers can be used to reveal the comparative influence of artificial selection and breeding for important agronomic traits on LD and haplotype structure. Divergent bread wheat gene pools show extreme differences in local LD surrounding loci involved in important adaptation and grain quality traits, for example (Figure 2; Voss-Fels *et al.*, 2015). In the highly duplicated A and C subgenomes of rapeseed, Qian *et al.* (2014) demonstrated strong subgenomic bias for selection signatures during breeding for important seed quality traits. Serious erosion of genetic variability in C-subgenome QTL was found to reflect a considerably lower diversity and a reduced recombination rate, which in turn hamper breeding progress and heterotic potential. Such information provides breeders with important information to develop strategies to precisely reinstate diversity in such regions, for example by inducing elevated recombination via de novo polyploidization (Snowdon *et al.*, 2015).

Enriching subgenomic diversity for improvement of heterotic potential

Translating high-resolution genome data to structured breeding populations derived from sequenced founder lines provides a basis for knowledge-based enrichment of low-diversity chromosome segments, to overcome linkage drag associated with large selection signatures (Voss-Fels *et al.*, 2015). This can also enhance heterosis by expanding diversity between hybrid breeding pools into chromosome segments containing strong adaptation signatures. Snowdon *et al.* (2015) present the concept of heterotic haplotype capture (HHC), which uses whole-genome profiling to identify and enrich diversity-poor genome regions and introduce these into hybrid breeding programmes for targeted improvement of heterosis. In HHC, fully sequenced, diverse founder lines are used to generate large structured prebreeding populations, like nested association mapping panels (Gore *et al.*, 2009; Yu *et al.*, 2008). By genotyping an entire population with a high-density SNP array, it is possible to detect crossover breakpoints in every individual at a previously unavailable resolution and combine these with parental sequence data to impute sequenced haplotypes across the whole genome in huge populations. Sequencing of male-sterile maternal lines, used to create and phenotype test hybrids from such a population, can subsequently provide a basis to identify haplotype structures associated with heterotic trait performance (Snowdon *et al.*, 2015). The HHC concept thus enables introduction and characterization of novel diversity on a high-resolution, subchromosomal level, while simultaneously facilitating the replenishment of eroded diversity in strongly selected genome regions (Figure 3).

Unravelling the genetic basis of heterosis

Besides the high-definition molecular characterization and utilization of crop breeding germplasm, novel genotyping technologies can also shed new light on the genetic background of heterosis. Hybrid crops, exploiting heterosis for yield gain and stability, have become one of the major drivers of increased agricultural production during the past few decades. Despite the global importance of heterosis for food security, and the growing tendency towards utilization of hybrid vigour even in inbreeding crops like bread wheat, the molecular and genetic mechanisms underlying this phenomenon are still not completely understood.

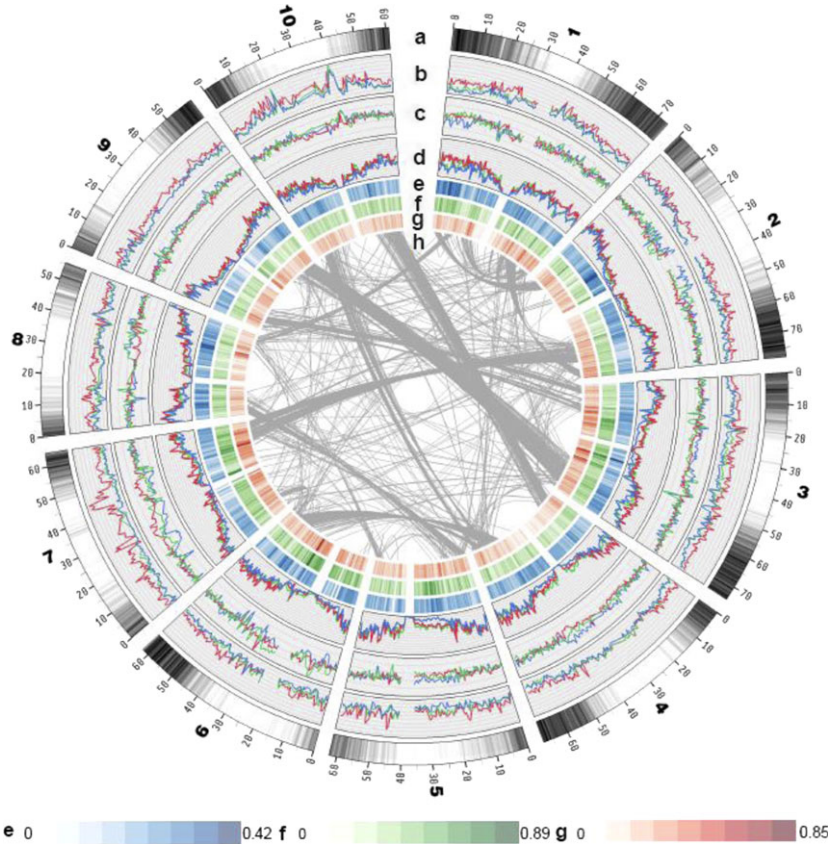

Figure 1 Genomewide patterns of sequence diversity in *Sorghum bicolor*. The 10 chromosomes are portrayed along the perimeter of each circle. Concentric circles display (a) gene content density distribution; (b) genomic diversity of wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (c) selection patterns (Tajima's D statistic) in wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (d) number of SNPs in wild and weedy genotypes (red), landraces (green) and improved inbreds (blue); (e) fixation indices (F_{ST} values) of improved inbreds versus landraces; (f) F_{ST} values of improved inbreds versus wild and weedy genotypes; and (g) F_{ST} values of landraces versus wild and weedy genotypes. (h) A graphical view of duplicated annotated genes is indicated by connections between segments. Figure from Mace *et al.* (2013), reprinted with permission from Macmillan Publishers Ltd, *Nature Communications* © 2013.

Genome profiling in large hybrid populations today offers an unprecedented resolution for dissection of loci and genes involved in heterotic expression. Recently, Huang *et al.* (2015) described a study in which an extensive population of 1495 elite hybrid rice varieties, along with their inbred parental lines, was subject to detailed genomewide sequence analysis in order to investigate genomic effects on hybrid vigour for 38 agronomic traits. The resequenced genomes of all parental lines harboured around 1.3 million polymorphic SNP markers, which were subsequently used to study population genetic parameters and perform GWAS at an unprecedented resolution. In particular, heterozygous chromosome regions were revealed that contribute to trait expression in the F_1 hybrids. Elucidation of the corresponding genomic effects on phenotypic traits demonstrated that pyramiding of multiple loci facilitates the accumulation of numerous rare superior alleles with positive effects. In other words, dominance complementation contributes most to the heterosis effect in hybrid rice production. A combination of forward and background selection using high-throughput genome screening tools (Herzog and Frisch, 2011; Herzog *et al.*, 2014) can thus be expected to significantly improve potential for increasing breeding gain through efficient exploitation of hybrid vigour.

The idea of genomic hybrid breeding, in which a genome-based prediction strategy based on genome sequence data is applied to estimate the performance of the F_1 progeny in hybrid breeding, was introduced in rice by Xu *et al.* (2014). Using over 250 000 SNP markers, generated by resequencing 210 parental inbred lines from a training set of 278 randomly selected hybrids, this study demonstrated the power of marker-directed estimation of F_1 hybrid yields in rice. The top 100 predicted hybrids, from a total of 21 945 possible combinations between the parental accessions, were estimated to exceed the overall yield average by 16%. This represents a significant improvement on average selection gains from conventional breeding and accelerated hybrid rice production.

Further applications for breeding and crop improvement

High-throughput, high-density genome-profiling tools enable the rapid and low-cost portrayal of crop genome characteristics in a precise, high-resolution manner. Identification of molecular variants on the DNA sequence level opens versatile options for practical application. As described above, breeders can use this detailed information for more targeted germplasm interchange between gene pools for improvement of diversity and heterosis. Generally, crop wild relatives represent an excellent genetic resource to reinstate variation caused by genetic bottlenecks during crop domestication and breeding (Huang and Han, 2014; Lopes *et al.*, 2015). On the other hand, breeders are often reluctant to implement completely novel diversity from distant gene pools due to their lack of adaptation and the consequent performance penalty. High-density genome data may help to improve this problem by enabling more efficient genomic selection strategies that efficiently predict performance based on genomewide marker combinations (Heffner *et al.*, 2009).

One key to improving crop performance and breeding processes is the enhancement of recombination in diversity-poor chromosome regions. Two recent studies have demonstrated how this might be achieved with support from high-resolution genotype data. By high-coverage sequencing of 41 rice offspring

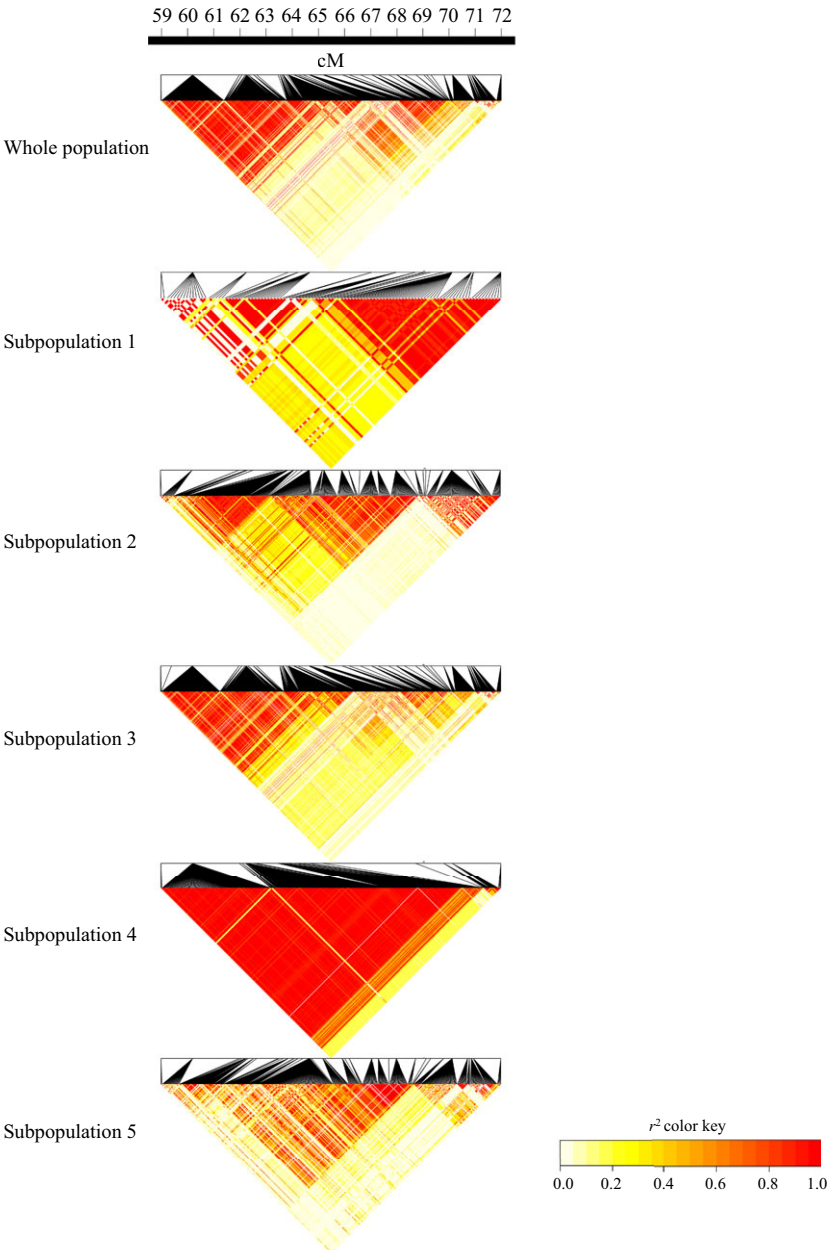

Figure 2 Detailed comparison of local linkage disequilibrium (LD) decay on a 13-cM segment of *Triticum aestivum* chromosome 1B in a population of 460 international wheat accessions, comparing the local genetic diversity within five subpopulations representing distinct breeding pools. This example demonstrates how strong directional selection in distinct breeding pools can lead to highly distinct patterns of LD. Densely spaced SNP markers can assist backcrossing programmes to enrich diversity-poor regions. Figure adapted from Voss-Fels *et al.* (2015), reprinted with permission from the Crop Science Society of America, *The Plant Genome* © 2015.

form a biparental cross, Si *et al.* (2015) generated a detailed map of recombination hot spots and cold spots based on 900 000 high-quality polymorphic loci. Interestingly, the recombination hot spot regions were enriched with genes involved in response to environmental stimuli, and environmental stress was found to increase the recombination rate in around one-third of the genotypes. If this can be confirmed, then breeders may be able to increase recombination by making crosses on plants grown under stressed conditions. In another approach, Suay *et al.* (2014) demonstrated that interspecific *Brassica* hybrids carrying a specific chromosome have a significantly elevated recombination rate and reduced crossover interference. As described by Mason *et al.* (2014), high-throughput genomewide SNP genotyping provides an ideal basis for molecular cytogenetic analysis of such materials to infer chromosome behaviour at meiosis and identify individuals with elevated recombination in diversity-poor chromosome regions for use in breeding.

Genomic selection models predict the breeding value of an individual based on molecular marker information, using statistical calibrations from representative test populations with robust phenotype data and genomewide SNP profiles. Genomic selection is becoming a widespread technique in breeding of many important crops, for example maize, wheat, rice and barley (Lin *et al.*, 2014; Poland *et al.*, 2012; Riedelsheimer *et al.*, 2012).

Despite the exceptional recent advances in molecular tools for genotyping, efficient, high-throughput phenotyping platforms are still a major bottleneck in the dissection and understanding of high-value, quantitatively inherited traits. Particularly for traits that are unable to be effectively assessed under controlled conditions, there is still a need for further improvements in automated phenotyping in field trials. Recent advances in remote sensing, field-based robotics and geo-referenced aerial image capture with multisensoric imaging systems (Fahlgren *et al.*, 2015), in combination with high-performance computing, are

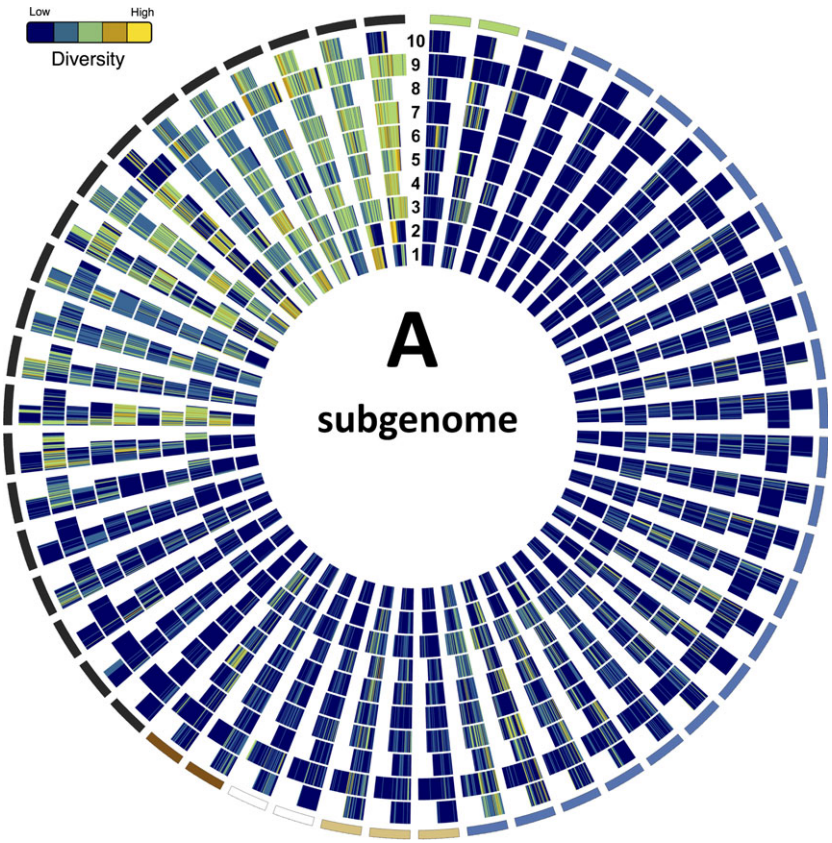

Figure 3 Genomics-assisted rejuvenation of a depleted breeding pool. The founder lines from the *Brassica napus* nested association mapping (NAM) panel (represented in this example by segments describing the 10 chromosomes of the *B. napus* A subgenome) introduce completely novel diversity into the depleted gene pool of cultivated winter oilseed rape. Diversity erosion (dark blue), displayed by many chromosomes of natural *B. napus* (coloured and white bars in external ring), is replaced by considerably enriched sequence diversity (yellow) and strong recombination in the corresponding chromosome regions of many synthetic forms (black bars in external ring). Figure adapted from Snowdon *et al.* (2015) and reprinted with permission from Elsevier, *Trends in Plant Science* © 2015.

improving the phenotyping bottleneck (Araus and Cairns, 2014; White *et al.*, 2012). Standardization of high-throughput phenotype ontologies for automated analysis in association with genotype data remains a challenge, and effective data management and interpretation pipelines are still required to increase the applicability of high-throughput phenotyping platforms in crop improvement.

Conclusions and outlook

The global demand for major crops is expected to strongly increase as a consequence of the steadily growing world population. On the other hand, the decreasing agricultural area, and increasingly stressed environments for plant production in the face of climate change, challenges plant breeders worldwide to produce constantly improving varieties that produce high and stable yields. The fundamental key to breeding success is genetic diversity, which provides breeders the basis for the selection of cultivars with a superior arable performance. The availability of powerful genomics tools provides an unprecedented basis to accelerate crop improvement and increase genetic gain in breeding programmes, especially for traits which are difficult, time-consuming or expensive to accurately evaluate phenotypically. The genome sequences of the world's most important crop plants open the way to identify and characterize all available diversity not only within crop primary gene pools, but also throughout related species. Ultra-high-throughput genotyping techniques like GBS or array-based SNP genotyping enable the prompt genome profiling of even large plant populations at constantly shrinking costs, providing breeders detailed, high-resolution molecular information. On the one hand, this can greatly improve the genetic resolution for mapping and cloning of useful genes and QTL, particularly as large structured populations like NAM or HHC panels become available for important crop species. High-resolution analysis of breeding germplasm allows breeders to gain deep and highly precise insights into genetic diversity on a subgenomic and chromosomal level. This is particularly helpful for targeted enrichment of depleted gene pools, in which strong, human-mediated selection during domestication and breeding has caused a dramatic loss of genetic diversity in many genome regions of modern varieties. Chromosome segments harbouring loci involved in essential traits such as vernalization requirement, winter hardiness, flowering time, seed quality or resistances against biotic stress are often associated with large blocks of very strong LD which can negatively impact yield via linkage drag. The ability to precisely identify these signatures of directional selection provides a basis to precisely reinstate diversity in affected genomic regions. Detailed genome profiles can also be very helpful to identify germplasm that is most suitable for crossing in order to reverse the erosion of genetic variation in elite material. Genomics-directed strategies to simultaneously increase recombination rates can facilitate this process. The targeted introgression of loci from exotic germplasm into high-performance varieties is likely to play a key role in continual improvement of heterotic performance in major crops. Ultimately, genomic selection and hybrid prediction strategies based on cheap, high-throughput, high-density genetic marker data will be a key factor in the acceleration of breeding progress to provide food, feed, fibre and fuel for future generations. Translating the huge progress in low-cost, high-resolution genotyping to significant increases in crop improvement requires similar advances in the development, cost and field applicability of high-throughput

phenotyping tools, including improved automated analysis and interpretation of geo-referenced, multisensor, field-based crop imaging data.

Acknowledgements

The authors were supported for this work by funding from the German Research Foundation (DFG grant number SN14/16-1), the Federal Ministry of Food and Agriculture (BMEL grant number FNR-22408212) and the Federal Ministry of Food and Agriculture (BMEL grant number 0315964H).

References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S. and Terauchi, R. (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotech.* **30**, 174–178.

Araus, J.L. and Cairns, J.E. (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* **19**, 52–61.

Bekele, W.A., Wieckhorst, S., Friedt, W. and Snowdon, R.J. (2013) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array. *Plant Biotech. J.* **11**, 1112–1125.

Bolger, M.E., Weisshaar, B., Scholz, U., Stein, N., Usadel, B. and Mayer, Klaus.F.X. (2014) Plant genome sequencing - applications for crop improvement. *Curr. Opin. Biotech.* **26**, 31–37.

Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A., See, D., Bai, G., Pumphrey, M., Tomar, L., Wong, D., Kong, S., Reynolds, M., da Silva, M.L., Bockelman, H., Talbert, L., Anderson, J.A., Dreisigacker, S., Baenziger, S., Carter, A., Korzun, V., Morrell, P.L., Dubcovsky, J., Morell, M.K., Sorrells, M.E., Hayden, M.J. and Akhunov, E. (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. *Proc. Nat. Acad. Sci. USA*, **110**, 8057–8062.

Chalhoub, B., Denoeud, F., Liu, S., Parkin Isobel, A.P., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.-C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.-H., Thi, V.H.D., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom Christine, H.D., Wang, X., Canaguier, A., Chauveau, A., Bérard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.-M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C. and Wincker, P. (2014) Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*, **345**, 950–953.

Chen, H., Xie, W., He, H., Yu, H., Chen, W., Li, J., Yu, R., Yao, Y., Zhang, W., He, Y., Tang, X., Zhou, F., Deng, X.W. and Zhang, Q. (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Mol. Plant*, **7**, 541–553.

Clarke, W.E., Federico, M.L., Gajardo, H.A., Gerhardt, D.J., Higgins, E., Sharpe, A.G., Snowdon, R.J., Parkin, I.A.P. and Iniguez-Luy, F.L. (2013) Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. *PLoS ONE*, **8**, e81992.

Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganal, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., Hedley, P., Tondelli, A., Pecchioni, N., Francia, E., Korzun, V., Walther, A. and Waugh, R. (2012) Natural variation in a homolog of Antirrhinum CENTRORADIALIS contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **44**, 1388–1392.

Deschamps, S., Llaca, V. and May, G.D. (2012) Genotyping-by-sequencing in plants. *Biology*, **1**, 460–483.

Edwards, D., Batley, J. and Snowdon, R.J. (2013) Accessing complex crop genomes with next-generation sequencing. *Theor Appl Genet.* **126**, 1–11.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Fahlgren, N., Gehan, M.A. and Baxter, I. (2015) Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* **24**, 93–99.

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.-M., Hansen, M., Joets, J., Le Paslier, M.-C., McMullen, M.D., Montalent, P., Rose, M., Schön, C.-C., Sun, Q., Walter, H., Martin, O.C. and Falque, M. (2011) A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, **6**, e28334.

Ganal, M.W., Polley, A., Graner, E.-M., Plieske, J., Wieseke, R., Luerssen, H. and Durstewitz, G. (2012) Large SNP arrays for genotyping in crop plants. *J. Biosci.* **37**, 821–828.

Gao, Z.Y., Zhao, S.C., He, W.M., Guo, L.B., Peng, Y.L., Wang, J.J., Guo, X.S., Zhang, X.M., Rao, Y.C., Zhang, C., Rao, Y.C., Zhang, C., Dong, G.J., Zheng, F.Y., Lu, C.X., Hu, J., Zhou, Q., Liu, H.J., Wu, H.Y., Xu, J., Ni, P.X., Zeng, D.L., Liu, D.H., Tian, P., Gong, L.H., Ye, C., Zhang, G.H., Wang, J., Tian, F.K., Xue, D.W., Liao, Y., Zhu, L., Chen, M.S., Li, J.Y., Cheng, S.H., Zhang, G.Y., Wang, J. and Qian, Q. (2013) Dissecting yield-associated loci in super hybrid rice by resequencing recombinant inbred lines and improving parental genome sequences. *Proc. Natl Acad. Sci. USA*, **110**, 14492–14497.

Gholami, M., Bekele, W.A., Schondelmaier, J. and Snowdon, R.J. (2012) A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. *Plant Biotech. J.* **10**, 635–645.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchison, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W.-L., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R.M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science*, **296**, 92–100.

Gore, M.Q., Chia, J.M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.L., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., Ware, D.H. and Buckler, E.S. (2009) A first-generation haplotype map of maize. *Science*, **326**, 1115–1117.

Hamilton, J.P., Hansey, C.N., Whitty, B.R., Stoffel, K., Massa, A.N., van Deynze, A., Jong, D., Walter, S., Douches, D.S. and Buell, C.R. (2011) Single nucleotide polymorphism discovery in elite North American potato germplasm. *BMC Genom.* **12**, 302.

Harper, A.L., Trick, M., Higgins, J., Fraser, F., Clissold, L., Wells, R., Hattori, C., Werner, P. and Bancroft, I. (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. *Nat. Biotech.* **30**, 798–802.

Heffner, E.L., Sorrells, M.E. and Jannink, J.L. (2009) Genomic selection for crop improvement. *Crop Sci.* **49**, 1–12.

Herzog, E. and Frisch, M. (2011) Selection strategies for marker-assisted backcrossing with high-throughput marker systems. *Theor. Appl. Genet.* **123**, 251–260.

Herzog, E., Falke, K.C., Presterl, T., Scheuermann, D., Ouzunova, M. and Frisch, M. (2014) Selection strategies for the development of maize introgression populations. *PLoS ONE*, **9**, e92429.

Huang, X. and Han, B. (2014) Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **65**, 531–551.

Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., Guo, Y., Lu, Y., Zhou, C., Fan, D., Weng, Q., Zhu, C., Huang, T., Zhang, L., Wang, Y., Feng, L., Furuumi, H., Kubo, T., Miyabayashi, T., Yuan, X., Xu, Q., Dong, G., Zhan, Q., Li, C., Fujiyama, A., Toyoda, A., Lu, T., Feng, Q., Qian, Q., Li, J. and Han, B. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, **490**, 497–501.

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., Chen, N., Hao, Z., Liu, K., Zhu, C., Huang, T., Zhao, Q.,

Zhang, L., Fan, D., Zhou, C., Lu, Y., Weng, Q., Wang, Z.-X., Li, J. and Han, B. (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Comm.* **6**, 6258.

International Wheat Genome Sequencing Consortium. (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*, **345**, 1251788.

Jannink, J.L., Lorenz, A.J. and Iwata, H. (2010) Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics*, **9**, 166–177.

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G. and Lorenz, A. (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genom.* **15**, 740.

Jiao, Y., Zhao, H., Ren, L., Song, W., Zeng, B., Guo, J., Wang, B., Liu, Z., Chen, J., Li, W., Zhang, M., Xie, S. and Lai, J. (2012) Genome-wide genetic changes during modern breeding of maize. *Nat. Genet.* **44**, 812–815.

Ku, C.-S., Wu, M., Cooper, D.N., Naidoo, N., Pawitan, Y., Pang, B., Iacopetta, B. and Soong, R. (2012) Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev. Mol. Diagn.* **12**, 241–251.

Langridge, P. and Fleury, D. (2011) Making the most of 'omics' for crop breeding. *Trends Biotechnol.* **29**, 33–40.

Li, Y.H., Zhou, G., Ma J., J., Jin, L., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., Zhang, S., Zuo, Q., Shi, X., Li, Y., Zhang, W., Hu, Y., Kong, G., Hong, H., Tan, B., Song, J., Liu, Z., Wang, Y., Ruan, H., Yeung, C.K.L., Liu, J., Wang, H., Zhang, L., Guan, R., Wang, K., Li, W., Chen, S., Chang, R., Jiang, Z., Jackson, S.A., Li, R. and Qiu, L. (2014) *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nat. Biotech.* **32**, 1045–1052.

Lin, Z., Hayes, B.J. and Daetwyler, H.D. (2014) Genomic selection in crops, trees and forages: a review. *Crop Pasture Sci.* **65**, 1177.

Lopes, M.S., El-Basyoni, I., Baenziger, P.S., Singh, S., Royo, C., Ozbek, K., Aktas, H., Ozer, E., Ozdemir, F., Manickavelu, A., Ban, T. and Vikram, P. (2015) Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *J. Exp. Bot.* **66**, 3477–3486. doi:10.1093/jxb/erv122.

Lu, F., Romay, M.C., Glaubitz, J.C., Bradbury, P.J., Elshire, R.J., Wang, T., Li, Y., Li, Y., Semagn, K., Zhang, X., Hernandez, A.G., Mikel, M.A., Soifer, I., Barad, O. and Buckler, E.S. (2015) High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* **6**, 6914.

Mace, E.S., Tai, S., Gilding, E.K., Li, Y., Prentis, P.J., Bian, L., Campbell, B.C., Hu, W., Innes, D.J., Han, X., Cruickshank, A., Dai, C., Frère, C., Zhang, H., Hunt, C.H., Wang, X., Shatte, T., Wang, M., Su, Z., Li, J., Lin, X., Godwin, I.D., Jordan, D.R. and Wang, J. (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Commun.* **4**, 2320.

Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., Akhunov, E.D., Hedley, P.E., Gonzales, A.M., Morrell, P.L., Kilian, B., Blattner, F.R., Scholz, U., Mayer, K.F., Flavell, A.J., Muehlbauer, G.J., Waugh, R., Jeddeloh, J.A. and Stein, N. (2013) Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505.

Mason, A.S., Batley, J., Bayer, P.E., Hayward, A., Cowling, W.A. and Nelson, M.N. (2014) High-resolution molecular karyotyping uncovers pairing between ancestrally related *Brassica* chromosomes. *New Phytol.* **202**, 964–974.

Mayer, K.F.X., Waugh, R., Brown, J.W.S., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., Close, T.J., Wise, R.P. and Stein, N. (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**, 711–716.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Rahman, M., Ware, D., Westhoff, P., Mayer Klaus, F.X., Messing, J. and Rokhsar, D.S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Poland, J. (2015) Breeding-assisted genomics. *Curr. Opin. Plant Biol.* **24**, 119–124.

Poland, J.A. and Rife, T.W. (2012) Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome*, **5**, 92.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannink, J.L. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, **5**, 103–113.

Qian, L., Qian, W. and Snowdon, R.J. (2014) Sub-genomic selection patterns as a signature of breeding in the allopolyploid *Brassica napus* genome. *BMC Genom.* **15**, 1170.

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

Rife, T.W., Wu, S., Bowden, R. and Poland, J.A. (2015) Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. *BMC Genom.* **16**, 169.

Romay, M.C., Millard, M.J., Glaubitz, J.C., Peiffer, J.A., Swarts, K.L., Casstevens, T.M., Elshire, R.J., Acharya, C.B., Mitchell, S.E., Flint-Garcia, S.A., McMullen, M.D., Holland, J.B., Buckler, E.S. and Gardner, C.A. (2013) Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol.* **14**, R55.

Schiessl, S., Samans, B., Hüttel, B., Reinhard, R. and Snowdon, R.J. (2014) Capturing sequence variation among flowering-time regulatory gene homologues in the allopolyploid crop species *Brassica napus*. *Front. Plant Sci.* **5**, 404.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, K., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.T., Wing, R.A., Cregan, P., Specht, J., Grimwood, J., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla, A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.-T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.-P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.-M., Deragon, J.-M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y., Lee, H., Li, P., Lisch, D.R., Liu, S., Liu, Z., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L., Yu, Y., Zhang, L., Zhou, S., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A. and Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Lehmann Nielsen, K., Jørgensen, J. and Weigel Andersen, S.U. (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Meth.* **6**, 550–551.

Shi, J. and Lai, J. (2015) Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* **24**, 47–53.

Si, W., Yuan, Y., Huang, J., Zhang, X., Zhang, Y., Zhang, Y., Tian, D., Wang, C., Yang, Y. and Yang, S. (2015) Widely distributed hot and cold spots in meiotic recombination as shown by the sequencing of rice F_2 plants. *New Phytol.* **206**, 1491–1502.

Snowdon, R.J., Abbadi, A., Kox, T., Schmutzer, T. and Leckband, G. (2015) Heterotic haplotype capture: precision breeding for hybrid performance. *Trends Plant Sci.* **20**, 410–413.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE*, **8**, e54985.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N. and McCouch, S. (2013) Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* **126**, 2699–2716.

Suay, L., Zhang, D., Eber, F., Jouy, H., Lodé, M., Huteau, V., Coriton, O., Szadkowski, E., Leflon, M., Martin, O.C., Falque, M., Jenczewski, E., Paillard, S. and Chèvre, A.-M. (2014) Crossover rate between homologous chromosomes and interference are regulated by the addition of specific unpaired chromosomes in *Brassica*. *New Phytol.* **201**, 645–656.

The Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Thomson, M.J. (2014) High-throughput SNP genotyping to accelerate crop improvement. *Plant Breed. Biotech.* **2**, 195–212.

Uitdewilligen, J.G.A.M.L., Wolters, A.A., D'hoop, B.B., Borm, T.J.A. and Visser, R.G.F. (2013) A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE*, **8**, e62355.

Unterseer, S., Bauer, E., Haberer, G., Seidel, M., Knaak, C., Ouzunova, M., Meitinger, T., Strom, T.M., Fries, R., Pausch, H., Bertani, C., Davassi, A., Mayer, K.F. and Schön, C.-C. (2014) A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genom.* **15**, 823.

Varshney, R.K., Terauchi, R. and McCouch, S.R. (2014) Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biol.* **12**, e1001883.

Voss-Fels, K., Frisch, M., Qian, L., Kontowski, S., Friedt, W., Gottwald, S. and Snowdon, R.J. (2015) Subgenomic diversity patterns caused by directional selection in bread wheat gene pools. *Plant Genome*, **8**, doi:10.3835/plantgenome2015.03.0013.

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M., Salvi, S., Milner, S.G., Cattivelli, L., Mastrangelo, A.M., Whan, A., Stephen, S., Barker, G., Wieseke, R., Plieske, J., Lillemo, M., Mather, D., Appels, R., Dolferus, R., Brown-Guedira, G., Korol, A., Akhunova, A.R., Feuillet, C., Salse, J., Morgante, M., Pozniak, C., Luo, M.-C., Dvorak, J., Morell, M., Dubcovsky, J., Ganal, M., Tuberosa, R., Lawley, C., Mikoulitch, I., Cavanagh, C., Edwards, K.J., Hayden, M. and Akhunov, E. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotech. J.* **12**, 787–796.

White, J.W., Andrade-Sanchez, P., Gore, M.A., Bronson, K.F., Coffelt, T.A., Conley, M.M., Feldmann, K.A., French, A.N., Heun, J.T., Hunsaker, D.J., Jenks, M.A., Kimball, B.A., Roth, R.L., Strand, R.J., Thorp, K.R., Wall, G.W. and Wang, G. (2012) Field-based phenomics for plant genetics research. *Field. Crop. Res.* **133**, 101–112.

Wieckhorst, S., Bekele, W., Kloiber-Maitz, M., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Davassi, A. and Snowdon, R. (2015) *A high-density SNP genotyping array for genome-based breeding of energy sorghum for central Europe. Plant and Animal Genome XXIII Conference, San Diego, CA, USA, January 10-14*, 2015. https://pag.confex.com/pag/xxiii/webprogram/Paper15935.html

Wray, N.R., Yang, J., Hayes, B.J., Price, A.L., Goddard, M.E. and Visscher, P.M. (2013) Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515.

Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., Ni, P., Zhang, G., Yang, S., Li, R., Wang, J., Orjeda, G., Guzman, F., Torres, M., Lozano, R., Ponce, O., Martinez, D., De la Cruz, G., Chakrabarti, S.K., Patil, V.U., Skryabin, K.G., Kuznetsov, B.B., Ravin, N.V., Kolganova, T.V., Beletsky, A.V., Mardanov, A.V., Di Genova, A., Bolser, D.M., Martin David, M.A., Li, G., Yang, Y., Kuang, H., Hu, Q., Xiong, X., Bishop, G.J., Sagredo, B., Mejía, N., Zagorski, W., Gromadka, R., Gawor, J., Szczesny, P., Huang, S., Zhang, Z., Liang, C., He, J., Li, Y., He, Y., Xu, J., Zhang, Y., Xie, B., Du, Y., Qu, D., Bonierbale, M., Ghislain, M., Herrera, M.D.R., Giuliano, G., Pietrella, M., Perrotta, G., Facella, P., O'Brien, K., Feingold, S.E., Barreiro, L.E., Massa, G.A., Diambra, L., Whitty, B.R., Vaillancourt, B., Lin, H., Massa, A.N., Geoffroy, M., Lundback, S., DellaPenna, D., Buell, C.R., Sharma, S.K., Marshall, D.F., Waugh, R., Bryan, G.J., Destefanis, M., Nagy, I., Milbourne, D., Thomson, S.J., Fiers, M., Jacobs Jeanne, M.E., Nielsen, K.L., Sønderkær, M., Iovene, M., Torres, G.A., Jiang, J., Veilleux, R.E., Bachem Christian, W.B., de Boer, J., Borm, T., Kloosterman, B., van Eck, H., Datema, E., Hekkert, B.T., Goverse, A., van Ham, R.C.H.J. and Visser, R.G.F. (2011) Genome sequence and analysis of the tuber crop potato. *Nature*, **475**, 189–195.

Xu, S., Zhu, D. and Zhang, Q. (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Nat. Acad. Sci. USA*, **111**, 12456–12461.

Yu, J., Holland, J.B., McMullen, M.D. and Buckler, E.S. (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, **178**, 539–551.

Zhao, K., Tung, C.-W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., McClung, A.M., Bustamante, C.D. and McCouch, S.R. (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat. Commun.* **2**, 467.