

# NGS technologies for analyzing germplasm diversity in genebanks<sup>★</sup>

Benjamin Kilian and Andreas Graner

Advance Access publication date 17 January 2012

## Abstract

More than 70 years after the first *ex situ* genebanks have been established, major efforts in this field are still concerned with issues related to further completion of individual collections and securing of their storage. Attempts regarding valorization of *ex situ* collections for plant breeders have been hampered by the limited availability of phenotypic and genotypic information. With the advent of molecular marker technologies first efforts were made to fingerprint genebank accessions, albeit on a very small scale and mostly based on inadequate DNA marker systems. Advances in DNA sequencing technology and the development of high-throughput systems for multiparallel interrogation of thousands of single nucleotide polymorphisms (SNPs) now provide a suite of technological platforms facilitating the analysis of several hundred of Gigabases per day using state-of-the-art sequencing technology or, at the same time, of thousands of SNPs. The present review summarizes recent developments regarding the deployment of these technologies for the analysis of plant genetic resources, in order to identify patterns of genetic diversity, map quantitative traits and mine novel alleles from the vast amount of genetic resources maintained in genebanks around the world. It also refers to the various shortcomings and bottlenecks that need to be overcome to leverage the full potential of high-throughput DNA analysis for the targeted utilization of plant genetic resources.

**Keywords:** genetic resources; next-generation sequencing; SNP; allele mining; genetic diversity; association analysis

## INTRODUCTION

Plant breeding needs to focus on traits with the greatest potential to increase yield under changing climate conditions [1]. Agricultural practices have gradually displaced local traditional varieties and crop wild relatives, leading to a dramatic loss of indigenous biodiversity. Tapping into the rich genetic diversity inherent in a crop species and their wild relatives is a prerequisite for germplasm improvement in the future [2–7; <http://www.fao.org>]. Hence, new technologies must be developed to accelerate breeding through improving genotyping and phenotyping methods and by accessing the available genetic diversity stored in genebanks around the world.

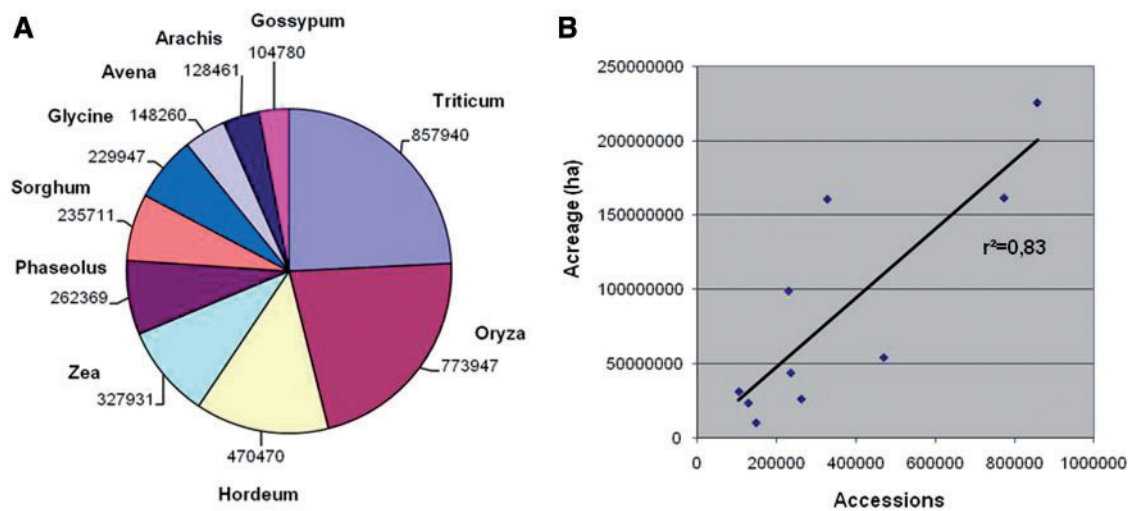
Prior to the advent of molecular characterization, accessions in germplasm collections were mainly examined based on morphological characters and phenotypic traits [8]. The development of molecular techniques now allows a more accurate analysis of large collections. High-throughput (HT) technologies including DNA isolation, genotyping, phenotyping and next-generation sequencing (NGS) provide new tools to add substantial value to genebank collections. The integration of genomic data into genebank documentation systems and its combination with taxonomic, phenotypic and ecological data will usher in a new era for the valorization of plant genetic resources (PGR). From the

Corresponding author. Benjamin Kilian, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Genebank/Genome Diversity, Corrensstrasse 3, 06466 Gatersleben, Germany. Tel.: +49 (0)39482 5-571; Fax: +49 (0)39482 5-500; E-mail: kilian@ipk-gatersleben.de

<sup>★</sup>This article is dedicated to Heiko Parzies, plant geneticist and plant breeder who passed away far too early.

**Benjamin Kilian** is in the research group Genome Diversity at the IPK. His main interests are in genetic diversity, evolution and domestication of Triticeae. He is in charge of projects aiming at exploiting natural genetic diversity by whole-genome association mapping, high-throughput phenotyping and resequencing approaches.

**Andreas Graner** is managing director of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) and the head of the German Federal *ex situ* genebank for agricultural and horticultural plants. His research aims at developing genomics based approaches for the valorization of plant genetics resources of barley (*Hordeum vulgare*).



**Figure 1:** *Ex situ* collections are dominated by major crop species. **(A)** Of the more than 3000 crop species that are maintained *ex situ*, 10 species totaling 3 540 000 accessions represent about half of the global inventory of *ex situ* resources amounting to 7.4 million. **(B)** Correlation of the aggregated size of the *ex situ* collections the acreage fetched by the individual crop species.

determination of phenotypic traits to the application of NGS to whole genomes, every aspect of genomics will have a great impact not only on PGR conservation, but also on their utilization in plant breeding [9].

Identification and tracking of genetic variation has become so efficient and precise that thousands of candidate genes can be tracked within large genebank collections [10]. Using NGS technologies, it is possible to resequence candidate genes, entire transcriptomes or entire plant genomes more efficiently and economically than ever before. Advances in sequencing technology will allow for whole-genome resequencing of hundreds of individuals. In this way, information on thousands of candidate genes and candidate regions can be harnessed for thousands of individuals to sample genetic diversity within and between germplasm pools, to map Quantitative Trait Loci (QTLs), to identify individual genes and to determine their functional diversity. In this review, we outline some important developments in this field, where NGS technologies are expected to enhance the value and thus the usefulness of genebank collections.

## STATE OF *EX SITU* GERMPLASM RESOURCES

PGR include cultivars, landraces, crop wild relatives and mutants. The loss of genetic diversity in many

crop plants has resulted in efforts to collect PGR which were initiated by Vavilov early in the 20th century aiming at supporting plant breeders with genetic material to extend genetic variability, as a basis to create new crop varieties [11]. A wealth of germplasm collections is available worldwide, with more than 7 million accessions held in over 1.700 genebanks (<http://www.fao.org/docrep/013/i1500e/i1500e00.htm>). These do not evenly cover all crop species but are highly biased regarding their agricultural importance. About 50% of the global *ex situ* germplasm is made up by only 10 crop species with the three largest collections (wheat, rice and barley) representing 28% of the global germplasm (Figure 1). Passport and genotypic data suggest that collections include different degrees of duplications resulting in ~1.9–2.2 million distinct accessions with the remaining being duplicates (<http://www.fao.org/docrep/013/i1500e/i1500e00.htm>). Proper conservation of PGR along with the development of best genebank practices and promoting the effective use is vital for food security in the future [12]. However, *ex situ* conservation is rather fragmented, largely because it is mainly based on national programs and scattered institutional efforts. For instance, barley (*Hordeum vulgare* L.), is maintained in more than 200 collections worldwide amounting to approximately 470 000 accessions [13]. Other crop species follow similar patterns [14]. Despite manifold efforts to coordinate genebank activities conservation

is still inefficient in many places and suffers from variable or even lacking standards, unreliable access and poor characterization and documentation of the material [15]. *Ex situ* germplasm collections for crop wild relatives are rather limited in size due to the difficulties in maintaining non-domesticated plants [16]. Introgression from wild to cultivated germplasm and vice versa both during seed multiplication in genebanks as well as in the wild pose a problem for proper maintenance and correct classification of the material, which usually is based on few morphological characters only. Another problem is that genebank accessions, even if they represent inbreeding crop species, often are genetically heterogeneous and may show residual heterozygosity. While this may reflect the original genetic state, e.g. of a landrace accession, it seriously can impair its molecular characterization and its subsequent use for research and breeding. Thus, most core collections are made up of accessions which underwent purification by single seed descent (SSD).

Systematic phenotypic analysis of genebank collections is a time and resource intense effort which has been mainly restricted to agronomic traits that show a high heritability and can be assessed based on the *per se* performance of an accession. Therefore, most evaluation efforts were focused to combine i.e. disease resistance and important morphological characters (yield components) [8, 17]. Deep genetic and phenotypic characterization of genetic resources by HT techniques, including resequencing of enriched candidate genes and low-coverage full-genome resequencing will increasingly become available. Concomitantly large amounts of data need to be integrated within the current documentation systems. Genebanks have to prepare for entering the genomics era by developing new strategies and novel information tools to assess the genetic diversity represented in their collections. Although there have been some successful examples of extracting useful genes from genebanks, the vast potential of this resource still remains largely untapped [18, 19].

## CHARACTERIZATION OF GERMPLASM BY MOLECULAR MARKERS: THE CURRENT STATE

A large series of studies have been undertaken to study diversity, domestication, evolution and phylogeny of PGR, largely selected from genebank collections. Early studies considered morphological and

cytogenetic characters. Various other techniques and molecular markers have been applied subsequently [20–23]. Until recently, amplified fragment length polymorphism (AFLP) or simple sequence repeats (SSR) were the molecular markers of choice for DNA fingerprinting of crop genomes [24–26]. Owing to their amenability to systematic development and HT detection, SNP markers increasingly applied to study genetic diversity in germplasm collections of up to several hundreds of accessions. Many of these collections have been established as association panels for linkage disequilibrium (LD) mapping, thus providing a first link between phenotypic and genotypic data sets. The corresponding accessions have been selected from various germplasm sources or breeding programs to represent a rough cross section of the overall genetic diversity available for a given species or for an ecogeographical region [27, 28]. This is exemplified by a population comprising 224 spring barley accessions, which were selected from the Barley Core Collection, BCC [29] and complemented by additional accessions to cover the entire distribution range of this crop [30]. More recently, about 1500 spring barley landraces adapted to temperate climate conditions were selected among 22 093 *Hordeum* accessions of the Federal *ex situ* genebank (IPK Gatersleben, Germany), based on their origin and morphology. The whole set has been genotyped by 43 SSR markers and analyzed for its genetic structure. While this is intended to usher in large-scale fingerprinting analysis of barley genebank accessions, the approach still falls short of providing informed molecular access to the entire collection. Different marker systems for genetic diversity studies and population parameters can be compared over a collection as recently shown by [31] who compared the performance of 42 SSR markers and 1536 SNP markers. The marker type of choice and the number of markers to be studied have to be adjusted for each species and project.

## Allele mining of individual loci

Plant accessions from wild or locally adapted landrace genebanks conserved in genebanks contain a rich repertoire of alleles that have been left behind by the selective processes of domestication, selection and cross-breeding that paved the way to today's elite cultivars. These resources stored in genebanks remain underexplored owing to a lack of efficient strategies to screen, isolate and transfer important alleles. The most effective strategy for

determining allelic richness at a given locus is currently to determine its DNA sequence in a representative collection of individuals. Large-scale allele mining projects for germplasm collections at the molecular level are needed as the one described for *Pm3* in wheat. Bhullar *et al.* [18] first selected a set of 1320 bread wheat landraces from a virtual collection of 16 089 accessions, using the focused identification of germplasm strategy (FIGS) and isolated seven new resistance alleles of the powdery mildew resistance gene *Pm3*. Similarly, a series of novel alleles have been detected for a recessive gene conferring virus resistance in barley [32, 33]. Further resequencing studies of candidate genes for agriculturally important traits have been published, however, from smaller collections and mostly without functional characterization [34–40].

Resequencing of candidate genes using Sanger sequencing has been applied to study phylogenetic relationships of crop plants, their domestication, evolution, speciation and ecological adaptation. Early studies resequenced a single locus or few loci in only few individuals per species [41, 42]. Reduced costs for Sanger sequencing using capillary instruments and 96-well formats facilitated multilocus studies in larger collections [43–51].

### NGS technologies to screen germplasm collections

Large-scale NGS is now possible using platforms such as Illumina/GA, Roche/GS FLX, Applied Biosystems/SOLiD and cPAL sequencing [52, 53]. The declining cost of generating such data is transforming all fields of genetics [54]. Many crop plant genomes are characterized by the vast abundance of repetitive DNA. For example, the genome of barley comprises >5 Gb of DNA sequence of which <2% can be accounted for by genes [55]. Therefore, to avoid excessive sequencing of putatively non-informative, repetitive DNAs, reduced-representation sequencing techniques have been developed to home in on subset of the genome for sequencing [56, 57]. When combined with techniques for labeling reads (barcoding), DNA from many individuals can be analyzed in the same pooled sequencing reaction, and NGS provides an increasingly affordable means. These technologies are therefore becoming a standard choice for generating genetic data in fields such as population genetics, conservation genetics and molecular ecology. On the other hand, the deluge of sequence data they will entail the necessity

to develop an appropriate IT infrastructure and new computational solutions [58–64].

Sequencing many individuals at low depth is another attractive strategy e.g. for complex trait association studies as shown by [65]. While detailed analysis of a single individual typically requires deep sequencing, resequencing of many individuals allows drastic reduction of sequencing depth when combined with efficient genotype imputation to match for missing data. Genotype imputation has been used widely in the analysis of genome-wide association studies (GWAS) to boost power and to facilitate the combination of results across different studies using meta-analyses [66, 67].

We have not yet reached the point at which routine whole-genome resequencing of large numbers of crop plant genomes becomes feasible. Therefore, it is necessary to select genomic regions of interest and to enrich these regions before sequencing. Sequencing targeted regions of DNA (e.g. the exome or parts thereof) rather than complete genomes will be likely the preferred approach for most genomics applications including evolutionary biology, association mapping and biodiversity conservation [68]. Sequencing targeted regions on massively parallel-sequencing instruments requires methods for concomitant enrichment of the templates to be sequenced. There are several enrichment approaches available, each with advantages and disadvantages [69–72]. Resequencing allows fingerprinting of many individuals without ascertainment bias which is inherent to some SNP marker systems [73–75].

As outlined above, targeted resequencing of hundreds of loci in genebank collections is already feasible. Yet, the costs for DNA extraction, complexity reduction and barcoding need to be brought down for systematic resequencing of genebank collections. In this context, large efforts have recently been made to automate protocols for massively parallel (re)sequencing and data analysis in order to match the increasing instrument throughput. These protocols that include e.g. large-scale automatic library preparation and size selection on robots [76] or fully automated construction of bar-coded libraries [77]—might be useful paving the way for automated NGS technologies to screen genebank collections [78].

### Multiparallel resequencing studies

Triggered by advancements in sequencing technologies, several crop genome sequences have been

produced or are underway [79–82]. Once good quality levels have been achieved, these sequences will enable researchers to address all kinds of biological questions or to link sequence diversity accurately to phenotypes.

Rapid developments in NGS will soon make whole-genome resequencing in several individuals or targeted resequencing of large germplasm collections reality. This will help to eliminate an important difficulty in the estimation of LD and genetic relationships between accessions obtained in bi-allelic genotyping studies caused by ascertainment bias i.e. the presence of rare alleles [73, 83–85].

Based on the available *Arabidopsis thaliana* (L.) Heynh. genome sequence, Weigel and Mott [86] advocated a 1001 Genomes project for *Arabidopsis*. Several *Arabidopsis* lines have been sequenced since [87, 88]. First studies on whole-genome resequencing in crop species have been published for rice and maize [66, 89, 90].

Combined genetic approaches for species, where a complete genome sequence and millions of SNPs are available, have been performed. Such approaches that include e.g. large-scale genotyping, targeted genomic enrichment, whole-genome resequencing and GWAS have been addressed to identify allelic diversity, rare genetic variation, QTL and their functional characterization [91–96] or to identify selective sweeps of favorable alleles and candidate mutations that have had a prominent role in domestication [97].

## TRAIT MAPPING IN PLANTS

### Genome-wide marker discovery using NGS

SNPs are the most abundant form of genetic variation in eukaryotic genomes and are not a limiting factor anymore, also not for crop species with large genome sizes like barley [98]. SNP markers are rapidly replacing SSRs or Diversity Arrays Technology (DArT) [99] markers because they are more abundant, reproducible, amenable to automation and increasingly cost-effective [100, 101]. SNP-based resources are presently being developed and made publicly available for broad application in crop research [102].

A high-quality genomic sequence as it is available for *Arabidopsis* and rice represents the ideal blueprint for resequencing and the identification of SNPs. But even for species with less complete genomic

sequences such as barley and wheat [103, 104] or other species [105–109] NGS methods are valuable for genome-wide marker development, genotyping and targeted sequencing across the genomes of populations [110–112]. These new methods—which include e.g. reduced-representation libraries (RRLs) [113–115], complexity reduction of polymorphic sequences (CRoPS) [116, 117], restriction-site-associated DNA sequencing (RAD-seq) [118] and low-coverage sequencing for genotyping [119–121] are applicable for genetic analysis to non-model species, to species with high levels of repetitive DNA or to breeding germplasm with low levels of polymorphism—without the need for prior sequence information. These methods can be applied to compare SNP diversity within and between closely related plant species or within wild natural populations [122, 123].

### Genome-wide association studies in crop plants

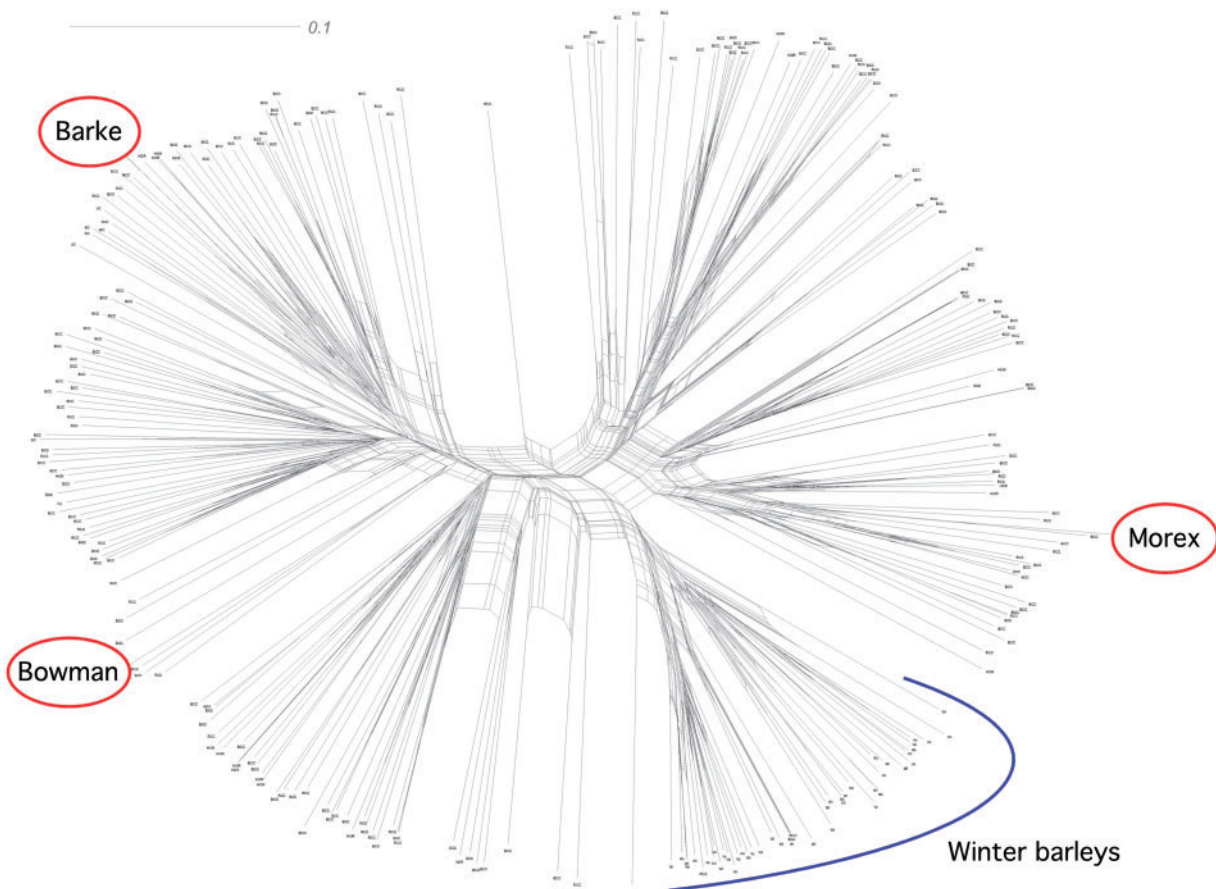
The systematic characterization and utilization of naturally occurring genetic variation has become an important approach in plant genome research and plant breeding. So far, linkage mapping based on bi-parental progenies has proven useful in detecting major genes and QTLs [124, 125]. Although this approach has been successful in many analyses, it suffers from several drawbacks. LD or association mapping is an attractive alternative to traditional linkage mapping and has several advantages over classical linkage mapping i.e. using unstructured populations that have been subjected to many recombination events [126–128]. GWAS in diverse germplasm collections offer new perspectives towards gene and allele discovery for traits of agricultural importance and dissecting the genetic basis of complex quantitative traits in plants [129, 130]. However, GWAS require a genome-wide assessment of genetic diversity (preferably based on a reference genome sequence and resequenced parts thereof), patterns of population structure, and the decay of LD. For this, effective genotyping techniques for plants, high-density marker maps, phenotyping resources, and if possible, a high-quality reference genome sequence is required [131]. The results of GWAS need in many cases confirmation by linkage analysis.

GWAS have identified a large number of SNPs associated with disease phenotypes in humans, also in diverse worldwide populations [132]. Early

association mapping studies in crop plants were hampered by the availability of a limited amount of mapped markers and thus were mainly based on resequencing candidate genes [39, 40]. The development of comprehensive sets of SNP markers that can be interrogated in highly multiparallel HT SNP genotyping ushered in the era of germplasm diversity studies and GWAS in crop plants. [87, 98, 119, 133–138].

For barley, few germplasm collections including wild and landrace barley have been genotyped using custom-made OPAs (oligo-pool assays) by Illumina GoldenGate technology [139, 140]. SNP markers significantly associated with traits are being used to identify genomic regions that harbor candidate genes for these traits in various collaborative barley projects. It is relatively easy to detect marker-trait associations in barley cultivar populations that have extensive LD (5–10 cM). Conversely, populations with low LD are supposed to provide

high-resolution associations (landraces, <5 cM; wild barley, <1 cM) but the number of markers needed to find significant associations is relatively high. This rapid decay in LD in populations of wild germplasm is a key generic problem with genotyping for bi-allelic SNPs. Furthermore, ascertainment bias of bi-allelic SNP discovery i.e. caused by rare alleles and alleles not present in the elite cultivars complicates the situation in landraces and wild germplasm [73, 141]. Thus rare alleles are usually excluded from analysis. Higher marker coverage is required in order to identify candidate genes more efficiently in diverse collections. In case of barley, a high density SNP Chip has been developed, which contains 7864 bi-allelic SNPs coming from NGS of a broad range of barley cultivars (R. Waugh *et al.*, unpublished data). Such customized arrays for HT SNP genotyping can accelerate genetic gain in breeding programs. First barley association panels have been genotyped using this resource (Figure 2). Similar SNP chips are



**Figure 2:** NeighborNet [166] of Hamming distances for 6885 polymorphic SNPs among 271 barley cultivars using the 9K Infinium iSELECT HD custom genotyping Bead Chip. Barley cultivars *Barke*, *Bowman* and *Morex* are highlighted as reference genotypes. Winter barleys form a cluster, which separates them clearly from the remaining spring barley accessions.

becoming available for an increasing number of crop plants [142, 143]. Combined studies using GWA mapping, comparative analysis, linkage mapping, resequencing and functional characterization of candidate genes already enabled the identification of candidate genes for selected traits [66, 91, 128].

While genotyping arrays are useful for assessing population structure and the decay of LD across large numbers of samples, low-coverage whole-genome sequencing will become the genotyping method of choice for GWAS in plant species [66]. As for humans, GWAS for plants will become the primary approach for identifying haplotypes and genes with common alleles influencing complex traits. However, common variations identified by GWAS account for only a small fraction of trait heritability and are unlikely to explain the majority of phenotypic variations of common traits. A potential source of the missing heritability is the contribution of rare alleles, insertion–deletion polymorphisms, copy number variants and epigenetic differences—that can be detected by NGS technologies. However, testing the association of rare variants with phenotypes of interest is challenging. Novel powerful association methods designed for large-scale resequencing data have to be developed [144–149].

In the future, it can be expected that mapping by sequencing will become the method of choice to discover the genes underlying quantitative trait variation in large purified germplasm collections [150–152] or epigenetic variation [84, 88, 153–155].

## OUTLOOK

PGR of crop wild relatives or locally adapted crop landraces contain a rich repertoire of alleles that have been lost by selective processes that generated our today's elite cultivars. Such alleles represent an invaluable asset to cope with future challenges for sustainable agricultural development and food production [156, 157]. In the medium run, draft genome sequences will be available for all major and many neglected crops species and resequencing of these genomes in germplasm collections will yield a wealth of information. Transforming this deluge of data to information and knowledge will increase our understanding in all fields of genetics including evolution, ecology, domestication and breeding. Now is a crucial time to explore the potential implications of this information revolution for genebanks and to recognize opportunities and limitations in applying

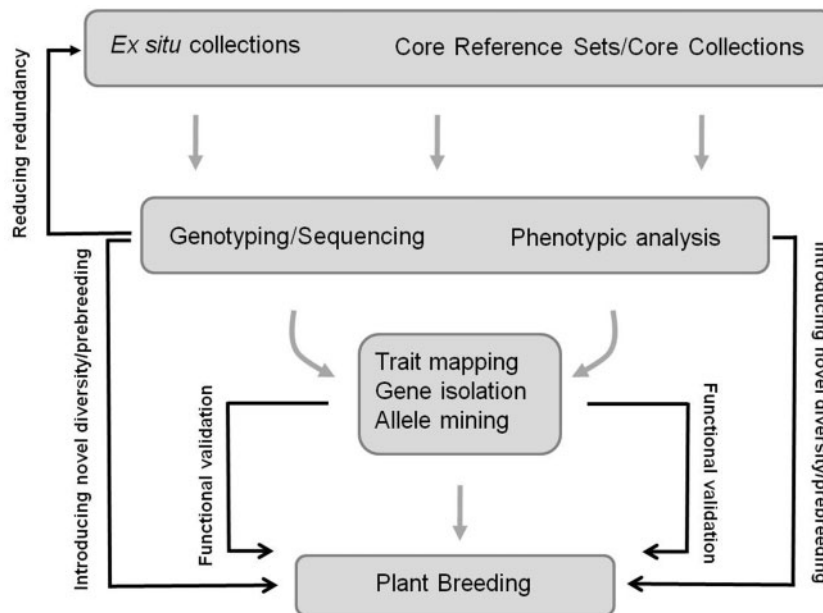
NGS tools and HT technologies to genebank collections [56, 158].

## Sequence informed conservation and utilization of PGR

The availability of sequence information can make a significant contribution to the conservation of PGR. The high degree of redundancy found between different *ex situ* collections wastes a prohibitive amount of resources (see above). Across the board, two-third of the seed multiplication that is the most resource intense step of all conservation efforts, could be made redundant, if there were ways to unambiguously identify duplicates. Most attempts to identify duplicated samples suffered from the difficulty to agree on a common set of markers for a given species, manifold problems to reproduce DNA marker data between different labs. DNA sequences do not suffer from such shortcomings and therefore represent an ideal information platform to tackle the issue of redundancy. Arguably, sequencing of *ex situ* collections just for the sake of eliminating redundancy would be too expensive an undertaking. Combination of this effort with one of the issues mentioned below could provide an added value.

Clearly large crop collections cannot be sequenced in one draft. Against the backdrop of the evolving technology, a stepwise approach should be envisaged. Glaszmann *et al.* [19] suggested the development of 'core reference sets' for our crops. A core reference set (CRS) is to be understood as 'a set of genetic stocks that are representative of the genetic resources of the crop and are used by the scientific community as a reference for an integrated characterization of its biological diversity'. Every CRS will serve as a public, standardized and well characterized resource for the scientific community. Well characterized, multiplied, isolated CRS have to be maintained for reference purposes, comparative studies, future reanalysis and integrative genomic analysis [59].

For this, already existing core collections must be transformed into genetic stocks, purified (homogeneous/stabilized) and taxonomically classified to facilitate practical choices for comparative association studies. One other approach is to select diverse accessions directly from genebank collections based on all available pre-existing characterization and evaluation data (C&E), pedigree, origin and collection site information. Survey genotyping to test the purity of accessions can be done with various molecular marker types such as inter-simple



**Figure 3:** DNA genotyping and sequencing as integral components for conservation and valorization of plant genetic resources.

sequence repeats (ISSRs) or AFLPs. Mixed accessions including more than one genotype have to be advanced by SSD before entering into systematic molecular and phenotypic characterization (Figure 3).

The scope of a genebank may be extended to that of a DNA bank, similar to biobanks devoted to target medical research [159]. The various implications of DNA banks for PGR have been discussed elsewhere. Common standards and Biobank Information Management Systems (BIMs) have to be developed to deal with highly complex and diverse sets of metadata. Advanced technologies for high-quality bio-sample storage and management systems are available and have to be implemented [160, 161].

Precise phenotyping is one of the major bottlenecks in characterizing large collections. New, non-invasive, automated image analysis technologies are currently under development for systematic phenotyping under greenhouse and field conditions using novel sensing and imaging technologies. Phenomics is an emerging field, in which large and complex data sets are being produced. These require long-term storage for future reanalysis when software tools and algorithms have improved or for comparative analysis [162, 163]. Pre-selection of contrasting accessions by different strategies including allele mining approaches, genotyping using custom-made Bead Chips and morphological characterization are effective strategies to reduce the number of

accessions prior to thorough phenotyping, the latter being the most time consuming step.

The ultimate goal regarding the valorization of PGR will be the deployment of novel alleles that will improve the trait under consideration. While resequencing of candidate genes is a straightforward approach to identify allelic variation, deployment of novel alleles in a breeding program is contingent on prior phenotypic validation. So far, this has been restricted to major genes, e.g. for disease resistance and seed quality. Validation of alleles of candidate genes for quantitative traits still remains a major challenge (i.e. Targeting Induced Local Lesions in Genomes (TILLING)), [164, 165]. In this regard, the ability to replace alleles by site specific recombination could spur the targeted utilization of PGR and thus greatly enhance the value chain of Biodiversity.

### Key Points

- Novel statistical approaches and promising NGS approaches are becoming available to screen major genebank collections. NGS will provide a platform for the large-scale development of SNPs that can be assayed in highly parallel manner for HT genotyping.
- Alternatively to SNP analysis genotyping by sequencing will be employed to obtain information on SNP and haplotype patterns.
- A staggered strategy starting from core collections is proposed to genotype and/or resequences genetic resources.
- Leverage of the full potential of sequence information on PGR depends on the availability of accurate phenotypic information and the potential to validate novel alleles at the phenotypic level.



## FUNDING

This work has been funded by Leibniz Institute of Plant Genetics and Crop Plant Research (IPK).

## References

- Long SP, Ort DR. More than taking the heat: crops and global change. *Curr Opin Pl Biol* 2010;**13**:240–7.
- Tanksley SD, McCouch SR. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 1997;**277**:1063–6.
- Zamir D. Improving plant breeding with exotic genetic libraries. *Nat Rev Genet* 2001;**2**:983–9.
- Hoisington D, Khairallah M, Reeves T, et al. Plant genetic resources: what can they contribute toward increased crop productivity? *Proc Natl Acad Sci USA* 1999;**96**:5937–43.
- Fernie AR, Tadmor Y, Zamir D. Natural genetic variation for improving crop quality. *Curr Opin Pl Biol* 2006;**9**:196–202.
- Takeda S, Matsuoka M. Genetic approaches to crop improvement: responding to environmental and population changes. *Nat Rev Genet* 2008;**9**:444–57.
- Tester M, Langridge P. Breeding technologies to increase crop production in a changing world. *Science* 2010;**327**:818–22.
- Boerner A, Freytag U, Sperling U. Analysis of wheat disease resistance data originating from screenings of Gatersleben genebank accessions during 1933 and 1992. *Genet Resour Crop Evol* 2006;**53**:453–65.
- Van K, Kim DH, Shin JH, et al. Genomics of plant genetic resources: past, present and future. *Pl Genet Res* 2011;**9**:155–8.
- Varshney RK, Nayak SN, May GD, et al. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trend Biotechnol* 2009;**27**:522–30.
- Loskutov IG. Vavilov and his Institute. *A history of the world collection of plant genetic resources in Russia*. Rome, Italy: International Plant Genetic Resources Institute, 1999.
- Van Heerwaarden J, van Eeuwijk FA, Ross-Ibarra J. Genetic diversity in a crop metapopulation. *Heredity* 2009;**104**:28–39.
- Knüpfper H. Triticeae genetic resources in ex situ genebank collections. In: Muehlbauer GJ, Feuillet C (eds). *Genetics and Genomics of the Triticeae*. Springer, 2009:31–79.
- Leung H, Hettel GP, Cantrell RP. International Rice Research Institute: roles and challenges as we enter the genomics era. *Trend Pl Sci* 2002;**7**:139–42.
- Khoury C, Laliberté B, Guarino L. Trends in ex situ conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet Res Crop Evol* 2010;**57**:625–39.
- Kilian B, Ozkan H, Shaaf S, et al. Comparing genetic diversity within a crop and its wild progenitor: a case study for barley. In: Maxted N, Dulloo ME, Ford-Lloyd BV, et al (eds). *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*. CABI, 2011.
- Perovic D, Przulj N, Milovanovic M, et al. Characterisation of spring barley genetic resources in Yugoslavia. In: Knüpfper H, Ochsmann J (eds). *Rudolf Mansfeld and Plant Genetic Resources. Proceedings of the symposium dedicated to the 100th birthday of Rudolf Mansfeld, Gatersleben, Germany, 8–9 October*, Vol. 22. Schriften zu Genetischen Ressourcen, 2001:301–6.
- Bhullar NK, Street K, Mackay M, et al. Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the Pm3 resistance locus. *Proc Natl Acad Sci USA* 2009;**106**:9519–24.
- Glazmann JC, Kilian B, Upadhyaya HD, et al. Accessing genetic diversity for crop improvement. *Curr Opin Pl Biol* 2010;**13**:167–73.
- Kovach MJ, McCouch SR. Leveraging natural diversity: back through the bottleneck. *Curr Opin Pl Biol* 2008;**11**:193–200.
- Feuillet C, Muehlbauer GJ. Genetics and genomics of the triticeae. In: Feuillet C, Muehlbauer GJ (eds). *Plant Genetics and Genomics: Crops and Models*, Vol. 7. Springer, 2009.
- Sang T. Genes and mutations underlying domestication transitions in grasses. *Pl Physiol* 2009;**149**:63–70.
- Kilian B, Mammen K, Millet E, et al. Aegilops L. In: Kole C (ed). *Wild Crop Relatives: Genetic and Breeding Resources*. Springer, 2011.
- Heun M, Schaefer-Pregl R, Klawan D, et al. Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* 1997;**278**:1312–4.
- Castillo A, Dorado G, Feuillet C, et al. Genetic structure and ecogeographical adaptation in wild barley (*Hordeum chilense* Roemer et Schultes) as revealed by microsatellite markers. *BMC Pl Biol* 2010;**10**:266.
- Allender C, King G. Origins of the amphiploid species *Brassica napus* L. investigated by chloroplast and nuclear molecular markers. *BMC Pl Biol* 2010;**10**:54.
- McMullen MD, Kresovich S, Villeda HS, et al. Genetic properties of the maize nested association mapping population. *Science* 2009;**325**:737–40.
- Chao S, Dubcovsky J, Dvorak J, et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genome* 2010;**11**:727.
- Knüpfper H, van Hintum Th. Summarised diversity - the Barley Core Collection. In: Bothmer R, Knüpfper H, van Hintum T, Sato K (eds). *Diversity in barley*. *Hordeum vulgare*: Elsevier Science, 2003:259–367.
- Haseneyer G, Stracke S, Paul C, et al. Population structure and phenotypic variation of a spring barley world collection set up for association studies. *Pl Breed* 2010;**129**:271–9.
- Hübner S, Günther T, Flavell A, et al. Islands and streams: clusters and gene flow in wild barley populations from the Levant. *Mol Ecol*. (in press).
- Stein N, Perovic D, Kumlehn J, et al. The eukaryotic translation initiation factor 4E confers multiallelic recessive *Bymovirus* resistance in *Hordeum vulgare* (L.). *Plant J* 2005;**42**:912–22.
- Hofinger BJ, Russell JR, Bass CG, et al. An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the *eIF4E* resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. *Mol Ecol* 2011;**20**:3653–68.
- Saitoh K, Onishi K, Mikami I, et al. Allelic diversification at the *C (O<sub>5</sub>Cl)* locus of wild and cultivated rice. *Genetics* 2004;**168**:997–1007.

35. Kilian B, Ozkan H, Deusch O, *et al.* Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. *Mol Biol Evol* 2007;**24**:217–27.
36. Zhu Q, Zheng X, Luo J, *et al.* Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol Biol Evol* 2007;**24**:875–88.
37. Jones H, Leigh FJ, Mackay I, *et al.* Population-based resequencing reveals that the flowering time adaptation of cultivated barley originated east of the Fertile Crescent. *Mol Biol Evol* 2008;**25**:2211–9.
38. Kovach MJ, Calingacion MN, Fitzgerald MA, *et al.* The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proc Natl Acad Sci USA* 2009;**106**:14444–9.
39. Stracke S, Haseneyer G, Veyrieras JB, *et al.* Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet* 2009;**118**:259–73.
40. Haseneyer G, Stracke S, Piepho HP, *et al.* DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC Pl Biol* 2010;**10**:5.
41. Kellogg EA, Appels R, Mason-Gamer AJ. When genes tell different stories: the diploid genera of Triticeae (Gramineae). *Syst Bot* 1996;**21**:321–47.
42. Lin JZ, Brown AHD, Clegg MT. Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). *Proc Natl Acad Sci USA* 2001;**98**:531–6.
43. Vaughan DA, Morishima H, Kadowaki K. Diversity in the *Oryza* genus. *Curr Opin Pl Biol* 2003;**6**:139–46.
44. Wright SI, Bi IV, Schroeder SG, *et al.* The effects of artificial selection on the maize genome. *Science* 2005;**308**:1310–4.
45. Hyten DL, Song Q, Zhu Y, *et al.* Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci USA* 2006;**103**:16666–71.
46. Haudry A, Cenci A, Ravel C, *et al.* Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Mol Biol Evol* 2007;**24**:1506–17.
47. Kilian B, Ozkan H, Walther A, *et al.* Molecular diversity at 18 Loci in 321 wild and 92 domesticate lines reveal no reduction of nucleotide diversity during *Triticum monococcum* (einkorn) domestication: implications for the origin of agriculture. *Mol Biol Evol* 2007;**24**:2657–68.
48. Izawa T, Konishi S, Shomura A, *et al.* DNA changes tell us about rice domestication. *Curr Opin Pl Biol* 2009;**12**:185–92.
49. Labate JA, Robertson LD, Baldo AM. Multilocus sequence data reveal extensive departures from equilibrium in domesticated tomato (*Solanum lycopersicum* L.). *Heredity* 2009;**103**:257–67.
50. Tian F, Stevens NM, Buckler ES. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proc Natl Acad Sci USA* 2009;**106**:9979–86.
51. Escobar J, Scornavacca C, Cenci A, *et al.* Multigenic phylogeny and analysis of tree incongruences in Triticeae (Poaceae). *BMC Evol Biol* 2011;**11**:181.
52. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotech* 2008;**26**:1135–45.
53. Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet* 2010;**11**:31–46.
54. Lister R, Gregory BD, Ecker JR. Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. *Curr Opin Pl Biol* 2009;**12**:107–18.
55. Wicker T, Taudien S, Houben A, *et al.* A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. *Plant J* 2009;**59**:712–22.
56. Paterson AH. Leafing through the genomes of our major crop plants: strategies for capturing unique information. *Nat Rev Genet* 2006;**7**:174–84.
57. Baird NA, Etter PD, Atwood TS, *et al.* Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 2008;**3**:e3376.
58. Alexander RP, Fang G, Rozowsky J, *et al.* Annotating non-coding regions of the genome. *Nat Rev Genet* 2010;**11**:559–71.
59. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010;**11**:476–86.
60. McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;**20**:1297–303.
61. Schadt EE, Linderman MD, Sorenson J, *et al.* Computational solutions to large-scale data management and analysis. *Nat Rev Genet* 2010;**11**:647–57.
62. Surget-Groba Y, Montoya-Burgos JI. Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res* 2010;**20**:1432–40.
63. Nielsen R, Paul JS, Albrechtsen A, *et al.* Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 2011;**12**:443–51.
64. Zhang W, Chen J, Yang Y, *et al.* A practical comparison of *De Novo* genome assembly software tools for next-generation sequencing technologies. *PLoS ONE* 2011;**6**:e17915.
65. Li Y, Sidore C, Kang HM, *et al.* Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res* 2011;**21**:940–51.
66. Huang X, Wei X, Sang T, *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 2010;**42**:961–7.
67. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;**11**:499–511.
68. Kirkness EF. Targeted sequencing with microfluidics. *Nat Biotech* 2009;**27**:998–9.
69. Tewhey R, Nakano M, Wang X, *et al.* Enrichment of sequencing targets from the human genome by solution hybridization. *Genome Biol* 2009;**10**:R116.
70. Gnirke A, Melnikov A, Maguire J, *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotech* 2009;**27**:182–9.
71. Mamanova L, Coffey AJ, Scott CE, *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Meth* 2010;**7**:111–8.
72. Teer JK, Bonnycastle LL, Chines PS, *et al.* Systematic comparison of three genomic enrichment methods for

- massively parallel DNA sequencing. *Genome Res* 2010;**20**:1420–31.
73. Moragues M, Comadran J, Waugh R, *et al.* Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *Theor Appl Genet* 2010;**120**:1525–34.
  74. Cosart T, Beja-Pereira A, Chen S, *et al.* Exome-wide DNA capture and next generation sequencing in domestic and wild species. *BMC Genome* 2011;**12**:347.
  75. Schuenemann VJ, Bos K, DeWitte S, *et al.* Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proc Natl Acad Sci USA* 2011;**108**:E746–52.
  76. Borgström E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. *PLoS ONE* 2011;**6**:e19119.
  77. Lennon N, Lintner R, Anderson S, *et al.* A scalable, fully automated process for construction of sequence-ready bar-coded libraries for 454. *Genome Biol* 2010;**11**:R15.
  78. Zheng J, Moorhead M, Weng L, *et al.* High-throughput, high-accuracy array-based resequencing. *Proc Natl Acad Sci USA* 2009;**106**:6712–7.
  79. Feuillet C, Leach JE, Rogers J, *et al.* Crop genome sequencing: lessons and rationales. *Trend Pl Sci* 2011;**16**:77–88.
  80. Schmutz J, Cannon SB, Schlueter J, *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* 2010;**463**:178–83.
  81. Young ND, Debelle F, Oldroyd GED, *et al.* The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* 2011;**480**:520–24.
  82. Varshney RK, Chen W, Li Y, *et al.* Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat Biotech* 2011, doi:10.1038/nbt.2022.
  83. Li R, Li Y, Fang X, *et al.* SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009;**19**:1124–32.
  84. Rafalski JA. Genomic tools for the analysis of genetic diversity. *Pl Genet Res* 2011;**9**:159–62.
  85. Wang L, Li P, Brutnell TP. Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief Funct Genome* 2010;**9**:118–28.
  86. Weigel D, Mott R. The 1001 Genomes Project for *Arabidopsis thaliana*. *Genome Biol* 2009;**10**:107.
  87. Cao J, Schneeberger K, Ossowski S, *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 2011;**43**:956–63.
  88. Lister R, Ecker JR. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 2009;**19**:959–66.
  89. Lai J, Li R, Xu X, *et al.* Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* 2010;**42**:1027–30.
  90. He Z, Zhai W, Wen H, *et al.* Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet* 2011;**7**:e1002100.
  91. Ramsay L, Comadran J, Druka A, *et al.* *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. *Nat Genet* 2011;**43**:169–72.
  92. Muir WM, Wong GK-S, Zhang Y, *et al.* Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *Proc Natl Acad Sci USA* 2008;**105**:17312–7.
  93. Huang X, Qian Q, Liu Z, *et al.* Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat Genet* 2009;**41**:494–7.
  94. Todesco M, Balasubramanian S, Hu TT, *et al.* Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 2010;**465**:632–6.
  95. Mokry M, Nijman I, van Dijken A, *et al.* Identification of factors required for meristem function in *Arabidopsis* using a novel next generation sequencing fast forward genetics approach. *BMC Genome* 2011;**12**:256.
  96. Yan J, Kandianis CB, Harjes CE, *et al.* Rare genetic variation at *Zea mays* *ctRB1* increases  $\beta$ -carotene in maize grain. *Nat Genet* 2010;**42**:322–7.
  97. Rubin CJ, Zody MC, Eriksson J, *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* 2010;**464**:587–91.
  98. Close T, Bhat P, Lonardi S, *et al.* Development and implementation of high-throughput SNP genotyping in barley. *BMC Genome* 2009;**10**:582.
  99. Wenzl P, Li H, Carling J, *et al.* A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. *BMC Genome* 2006;**7**:206.
  100. Ganal MW, Altmann T, Roeder MS. SNP identification in crop plants. *Curr Opin Pl Biol* 2009;**12**:211–7.
  101. Alsop B, Farre A, Wenzl P, *et al.* Development of wild barley-derived DArT markers and their integration into a barley consensus map. *Mol Breed* 2011;**27**:77–92.
  102. McCouch SR, Zhao K, Wright M, *et al.* Development of genome-wide SNP assays for rice. *Breed Sci* 2010;**60**:524–35.
  103. Paux E, Sourdille P, Salse J, *et al.* A physical map of the 1-Gigabase bread wheat chromosome 3B. *Science* 2008;**322**:101–4.
  104. Mayer KFX, Martis M, Hedley PE, *et al.* Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell* 2011;**23**:1249–63.
  105. Kuelheim C, Hui Yeoh S, Maintz J, *et al.* Comparative SNP diversity among four *Eucalyptus* species for genes from secondary metabolite biosynthetic pathways. *BMC Genome* 2009;**10**:452.
  106. Hribova E, Neumann P, Matsumoto T, *et al.* Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biol* 2010;**10**:204.
  107. Griffin P, Robin C, Hoffmann A. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to *Poa* grasses. *BMC Biol* 2011;**9**:19.
  108. Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* 2011;**475**:189–95.
  109. Shulaev V, Sargent DJ, Crowhurst RN, *et al.* The genome of woodland strawberry (*Fragaria vesca*). *Nat Genet* 2011;**43**:109–16.
  110. Bansal V, Harismendy O, Tewhey R, *et al.* Accurate detection and genotyping of SNPs utilizing population sequencing data. *Genome Res* 2010;**20**:537–45.

111. Davey JW, Hohenlohe PA, Etter PD, *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 2011;**12**:499–510.
112. Luca F, Hudson RR, Witonsky DB, *et al.* A reduced representation approach to population genetic analyses and applications to human evolution. *Genome Res* 2011; doi:10.1101/gr.119792.110.
113. Hyten D, Cannon S, Song Q, *et al.* High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genome* 2010;**11**:38.
114. You F, Huo N, Deal K, *et al.* Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genome* 2011;**12**:59.
115. Gompert Z, Forister ML, Fordyce JA, *et al.* Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. *Mol Ecol* 2010;**19**:2455–73.
116. van Orsouw NJ, Hogers RCJ, Janssen A, *et al.* Complexity Reduction of Polymorphic Sequences (CRoPS™): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2007;**2**:e1172.
117. Mammadov J, Chen W, Ren R, *et al.* Development of highly polymorphic SNP markers from the complexity reduced portion of maize [*Zea mays* L.] genome for use in marker-assisted breeding. *Theor Appl Gen* 2010;**121**:577–88.
118. Baxter SW, Davey JW, Johnston JS, *et al.* Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS ONE* 2011;**6**: e19315.
119. Huang X, Feng Q, Qian Q, *et al.* Weng Q, Huang T, Dong G, Sang T, Han B: High-throughput genotyping by whole-genome resequencing. *Genome Res* 2009;**19**:1068–76.
120. Andolfatto P, Davison D, Erezyilmaz D, *et al.* Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res* 2011;**21**:610–7.
121. Elshire RJ, Glaubitz JC, Sun Q, *et al.* A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS ONE* 2011;**6**:e19379.
122. Ossowski S, Schneeberger K, Lucas-Lledó JI, *et al.* The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 2010;**327**:92–4.
123. Pool JE, Hellmann I, Jensen JD, *et al.* Population genetic inference from genomic sequence variation. *Genome Res* 2010;**20**:291–300.
124. Frary A, Nesbitt TC, Frary A, *et al.* *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. *Science* 2000;**289**:85–8.
125. Komatsuda T, Pourkeirandish M, He C, *et al.* Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci USA* 2007;**104**:1424–9.
126. Oraguzie NC, Rikkerink EHA, Gardiner SE, De Silva HN (eds). *Association mapping in plants*. Springer, 2007.
127. Waugh R, Jannink JL, Muehlbauer GJ, *et al.* The emergence of whole genome association scans in barley. *Curr Opin Pl Biol* 2009;**12**:218–22.
128. Atwell S, Huang YS, Vilhjalmsón BJ, *et al.* Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 2010;**465**:627–31.
129. Mackay TFC, Stone EA, Ayroles JF. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 2009;**10**:565–77.
130. Hall D, Tegström C, Ingvarsson PK. Using association mapping to dissect the genetic basis of complex traits in plants. *Brief Funct Genomic* 2010;**9**:157–65.
131. Rafalski JA. Association genetics in crop improvement. *Curr Opin Pl Biol* 2010;**13**:174–80.
132. Rosenberg NA, Huang L, Jewett EM, *et al.* Genome-wide association studies in diverse populations. *Nat Rev Genet* 2010;**11**:356–66.
133. Yan J, Shah T, Warburton ML, *et al.* Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS ONE* 2009;**4**: e8451.
134. Deulvot C, Charrel H, Marty A, *et al.* Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genome* 2010;**11**:468.
135. Myles S, Chia JM, Hurwitz B, *et al.* Rapid genomic characterization of the genus *Vitis*. *PLoS ONE* 2010;**5**:e8219.
136. Grattapaglia D, Silva-Junior O, Kirst M, *et al.* High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC Pl Biol* 2011;**11**:65.
137. Myles S, Boyko AR, Owens CL, *et al.* Genetic structure and domestication history of the grape. *Proc Natl Acad Sci USA* 2011;**108**:3530–5.
138. Tian F, Bradbury PJ, Brown PJ, *et al.* Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 2011;**43**:159–62.
139. Russell J, Dawson IK, Flavell AJ, *et al.* Analysis of >1000 single nucleotide polymorphisms in geographically matched samples of landrace and wild barley indicates secondary contact and chromosome-level differences in diversity around domestication genes. *New Phytol* 2011;**191**:564–78.
140. Comadran J, Russell JR, Booth A, *et al.* Mixed model association scans of multi-environmental trial data reveal major loci controlling yield and yield related traits in *Hordeum vulgare* in Mediterranean environments. *Theor Appl Genet* 2011;**122**:1363–73.
141. Abdurakhmonov IY, Abdurakarimov A. Application of association mapping to understanding the genetic diversity of Plant Germplasm Resources. *Int J Pl Genomic* 2008; doi:10.1155/2008/574927.
142. Durstewitz G, Polley A, Plieske J, *et al.* SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species *Brassica napus*. *Genome* 2010;**53**: 948–56.
143. Zhao K, Wright M, Kimball J, *et al.* Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PLoS ONE* 2010;**5**: e10780.
144. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747–53.
145. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet* 2010;**11**:191–203.
146. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet* 2011;**12**:363–76.
147. Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* 2011;**12**:628–40.

148. Luo L, Boerwinkle E, Xiong M. Association studies for next-generation sequencing. *Genome Res* 2011;**21**:1099–108.
149. Swanson-Wagner RA, Eichten SR, Kumari S, *et al.* Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res* 2010;**20**:1689–99.
150. Bergelson J, Roux F. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nat Rev Genet* 2010;**11**:867–79.
151. Austin RS, Vidaurre D, Stamatiou G, *et al.* Next-generation mapping of Arabidopsis genes. *Plant J* 2011;**67**:715–25.
152. Schneeberger K, Weigel D. Fast-forward genetics enabled by new sequencing technologies. *Trend Pl Sci* 2011;**16**:282–8.
153. Delker C, Quint M. Expression level polymorphisms: heritable traits shaping natural variation. *Trend Pl Sci* 2011;**16**:481–8.
154. Rakyant VK, Down TA, Balding DJ, *et al.* Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2011;**12**:529–41.
155. Schmitz RJ, Zhang X. High-throughput approaches for plant epigenomic studies. *Curr Opin Pl Biol* 2011;**14**:130–6.
156. Khush GS. Green revolution: the way forward. *Nat Rev Genet* 2001;**2**:815–22.
157. Varshney RK, Bansal KC, Aggarwal PK, *et al.* Agricultural biotechnology for crop improvement in a variable climate: hope or hype? *Trend Pl Sci* 2011;**16**:363–71.
158. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. *Nat Rev Genet* 2010;**11**:697–709.
159. Angelow A, Schmidt M, Weitmann K, *et al.* Methods and implementation of a central biosample and data management in a three-centre clinical study. *Comput Methods Programs Biomed* 2008;**91**:82–90.
160. Wan E, Akana M, Pons JCJ, *et al.* Green technologies for room temperature nucleic acid storage. *Curr Issues Mol Biol* 2010;**12**:135–42.
161. Peplies J, Fraterman A, Scott R, *et al.* Quality management for the collection of biological samples in multicentre studies. *Eur J Epidemiol* 2010;**25**:607–17.
162. Montes JM, Melchinger AE, Reif JC. Novel throughput phenotyping platforms in plant genetic studies [abstract]. *Trend Pl Sci* 2007;**12**:433–6.
163. Zhu J, Ingram PA, Benfey PN, *et al.* From lab to field, new approaches to phenotyping root system architecture. *Curr Opin Pl Biol* 2011;**14**:310–7.
164. Gottwald S, Bauer P, Komatsuda T, *et al.* TILLING in the two-rowed barley cultivar ‘Barke’ reveals preferred sites of functional diversity in the gene *HvHox1*. *BMC Res Notes* 2009;**17**:258.
165. Hein I, Kumlehn J, Waugh R. Functional validation in the Triticeae. In: Feuillet C, Muehlbauer GJ (eds). *Genetics and Genomics of the Triticeae*. *Plant Genetics and Genomics: Crops and Models*, Vol. 7. Springer: Science+Business Media, LLC, New York, 2009:359–85.
166. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;**23**:254–67.