Application of genomic tools in plant breeding

A.M. Pérez de Castro, S. Vilanova, J. Cañizares, L. Pascual, J.M. Blanca, M.J. Díez, J. Prohens* and B. Picó

Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain

Running title: Genomic tools in plant breeding

*Address correspondence to this author at the Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain; Tel: +34963879424; Fax: +34963879422; E-mail: jprohens@btc.upv.es

Abstract: Plant breeding has been very successful in developing improved varieties using conventional tools and methodologies. Nowadays, the availability of genomic tools and resources is leading to a new revolution of plant breeding, as they facilitate the study of the genotype and its relationship with the phenotype, in particular for complex traits. Next Generation Sequencing (NGS) technologies are allowing the mass sequencing of genomes and transcriptomes, which is producing a vast array of genomic information. The analysis of NGS data by means of bioinformatics developments allows discovering new genes and regulatory sequences and their positions, and makes available large collections of molecular markers. Genome-wide expression studies provide breeders with an understanding of the molecular basis of complex traits. Genomic approaches include TILLING and EcoTILLING, which make possible to screen mutant and germplasm collections for allelic variants in target genes. Re-sequencing of genomes is very useful for the genome-wide discovery of markers amenable for high-throughput genotyping platforms, like SSRs and SNPs, or the construction of high density genetic maps. All these tools and resources facilitate studying the genetic diversity, which is important for germplasm management, enhancement and use. Also, they allow the identification of markers linked to genes and QTLs, using a diversity of techniques like bulked segregant analysis (BSA), fine genetic mapping, or association mapping. These new markers are used for marker assisted selection, including marker assisted backcross selection, 'breeding by design', or new strategies, like genomic selection. In conclusion, advances in genomics are providing breeders with new tools and methodologies that allow a great leap forward in plant breeding, including the 'superdomestication' of crops and the genetic dissection and breeding for complex traits.

Key Words: bioinformatics, complex traits, genetic maps, marker assisted selection, molecular markers, next-generation-sequencing, quantitative trait loci

INTRODUCTION

Ever since the beginnings of the domestication of plants, some 10,000 years ago, plant breeding has been extremely successful in developing crops and varieties that have contributed to the development of modern societies, and have successively beaten (neo-)Malthusian predictions [1]. Application of

conventional pre-genomics scientific breeding methodologies has led to the development of modern cultivars, which have contributed to the dramatic improvement of yield of most major crops since the middle of the 20th century. The success of plant breeding in the last century has relied in the utilization of natural and mutant induced genetic variation and in the efficient selection, by using suitable breeding methods, of the favorable genetic combinations. In this respect, the evaluation and identification of genetic variants of interest as well as the selection methodologies used have largely been based in the phenotypic evaluation.

Nowadays, genomics provides breeders with a new set of tools and techniques that allow the study of the whole genome, and which represents a paradigm shift, by facilitating the direct study of the genotype and its relationship with the phenotype [2]. While classical genetics revolutionized plant breeding at the beginning of the 20th century, genomics is leading to a new revolution in plant breeding at the beginning of the 21th century.

The field of genomics and its application to plant breeding are developing very quickly. The combination of conventional breeding techniques with genomic tools and approaches is leading to a new genomics-based plant breeding. In this new plant breeding context, genomics will be essential to develop more efficient plant cultivars, which are necessary, according to FAO, for the new 'greener revolution' needed to feed the world’s growing population while preserving natural resources.

One of the main pillars of genomic breeding is the development of high-throughput DNA sequencing technologies, collectively known as next generation sequencing (NGS) methods. These and other technical revolutions provide genome-wide molecular tools for breeders (large collections of markers, high-throughput genotyping strategies, high density genetic maps, new experimental populations, etc.) that can be incorporated into existing breeding methods [2, 3, 4, 5]. Recent advances in genomics are producing new plant breeding methodologies, improving and accelerating the breeding process in many ways (e.g., association mapping, marker assisted selection, ‘breeding by design’, gene pyramiding, genomic selection, etc.) [5, 6, 7].

Genomics approaches are particularly useful when dealing with complex traits, as these traits usually have a multi-genic nature and an important environmental influence. Thanks to these technological improvements it is now feasible for a small laboratory to generate in a short time span (e.g., several months) enough molecular data to obtain a set of mapped quantitative trait loci (QTLs), even in a species lacking any previous genomic information [8]. Genomic tools are thus facilitating the detection of

QTLs and the identification of existing favorable alleles of small effect, which have frequently remained unnoticed and have not been included in the gene pool used for breeding [9, 10].

In this review, we present and discuss the most relevant advances in the development and application of genomic tools for plant breeding, in particular for complex traits. Firstly, we introduce the most relevant genomic tools and secondly we provide examples of the application of these tools to plant breeding. The objective is to provide modern breeders with an updated synthetic view of how genomics can effectively improve the efficiency of breeding programs.

GENOMIC TOOLS AND RESOURCES FOR PLANT BREEDING

Genome and transcriptome sequencing

The availability of the whole genome sequence of a crop is of great utility for plant breeding, but until recently it has been an unachievable goal for most crops. This privilege was restricted to a reduced number of model species with small genomes and to projects with an important investment in time and resources, but now has extended to an increasing number of crops. However, it is also true that for important cultivated species with large and complex genomes such as wheat, sugarcane, or coffee, the sequencing of the whole genome is very challenging and may take several years before a draft is completed. Given the high cost of whole genome sequencing, transcriptome sequencing has been a cheaper alternative. The cDNA sequences (expressed sequence tags, ESTs) provide relevant information about the genes expressed in a certain tissue or organ, at a given stage of development and under particular environmental conditions. ESTs sequencing projects do not provide information about non-coding sequences and, even using diverse libraries, it is difficult to identify all genes and transcripts variants. Despite these limitations, ESTs collections have been very useful for breeders.

The Sanger technology has been the predominant sequencing method for the past thirty years. It has been used to sequence several genomes as well as many transcriptomes. The first international collaborative project resulted in the whole genome sequence of the model plant *Arabidopsis thaliana* [11]. Since then, reference genomes of selected genotypes were completed in a limited number of crops such as rice [12], maize [13], sorghum [14], populous [15], grapevine [16], papaya [17], or soybean [18]. The transcriptomes of most major crops, to a greater or lesser extent, were also sequenced. A global view

105 of the genomes and transcriptomes sequenced can be obtained from the Gene Index Project
106 (http://compbio.dfci.harvard.edu/tgi/plant.html) or in the NCBI Unigene database
107 (http://www.ncbi.nlm.nih.gov/unigene).

108 The field of genomics has changed with the arrival of NGS technologies [19]. These new
109 technologies have reduced the cost of sequencing by more than one thousand times compared to Sanger
110 technology, by avoiding time-consuming and tedious traditional cloning steps and making possible to
111 perform millions of sequencing reactions in parallel (Table 1). Among the "second generation"
112 technologies, the 454 (Roche, http://www.454.com) and Illumina (Illumina, http://www.illumina.com)
113 platforms are already profusely used to sequence crop species. Others, like Solid (Applied Biosystems,
114 http://www.appliedbiosystems. com/technologies), have been less exploited in plants. By using these
115 NGS technologies, the sequencing capacity of laboratories is continuously increasing. For instance, one
116 High-Seq 2000 Illumina Sequencer is able to generate 55 Gb per day, which is roughly eighteen times the
117 size of the human genome. Moreover, new, "third generation" platforms are being developed and
118 incorporated to sequencing projects, such as PacBio RS (Pacific Biosciences,
119 http://www.pacificbiosciences.com), Helicos (Helicos, http://www.helicosbio.com), or Ion Torrent (Life
120 Technologies, http://www.iontorrent.com). The sequence obtained by NGS are generally deposited in the
121 NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/unigene).

122 Nowadays, it is feasible to sequence most crop genomes (excluding those with a very large and
123 complex genome) with a relatively modest budget, by combining Sanger with NGS technologies.
124 Sequencing projects for 135 plant genomes, most of them corresponding to cultivated species or wild
125 relatives, have been completely sequenced (3), are being assembled (27) or are in progress (105), as
126 reported at the NCBI database (http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi). Other databases
127 including plant genome sequences are Plant GDB (http://www.plantgdb.org) and Phytozome
128 (http://www.phytozome.net). A fully sequenced and well annotated genome provides useful tools for the
129 breeders, as it allows the discovery of genes, determining their position and function, as well as the
130 development of large marker collections and high resolution maps. In the cases where only a draft
131 sequence is available, its usefulness depends on the quality of the assembly. For instance on many
132 occasions thousands of scaffolds are obtained but they are not anchored to the genetic map, which
133 difficults the use of the information. Many transcriptomes have also been sequenced, a number of them in
134 species for which no previous sequence information was available. Sweet potato [20], squash [21],

pigeonpea [22], or buckwheat [23] represent some examples published in the last months. These assays are showing the great complexity of plant transcriptomes, allowing the identification of rare transcript variants that are being used to improve gene annotation and our knowledge of gene function and regulation.

Bioinformatics

NGS technologies are facilitating sequencing projects, but have brought new challenges, as millions of short DNA reads have to be analysed and assembled [19]. Also, genetic maps, genotypes, or expression information at a genomic scale have to be processed in order to obtain the relevant biological information. Therefore, it is necessary to develop new bioinformatics tools (algorithms and software), which allow the analyses of huge amounts of genome-wide data, but it is also necessary to change the approaches used to understand complex biological traits [25, 26].

The field of sequence analysis has a long tradition and has enabled the assembly of many genome sequences obtained by Sanger sequencing. Nowadays, the huge amount of sequence reads generated by NGS and the low quality of individual reads requires new software tools and algorithms that allow dealing with these data in an efficient way [27]. We consider this to be a limiting factor for this kind of analyses right now. Although in the last years great advancements in the sequence processing algorithms have been achieved, the software currently available requires improvements in many cases.

Two of the most common analyses carried out on these NGS reads are genome assembly and annotation and mapping. Genome assembly is a complex task requiring powerful computers and skilled bioinformaticians [25]. In particular, the RAM memory requirements of the computers used to assemble eukariotic genomes could hinder the application of this technique by small laboratories. Some of the most commonly used assemblers are Roche's 454 Gsassembler, Celera Assembler, and Mira. Once a reference genome is available in the species it is common to study its variation [19]. To do this, a mapper software is commonly used instead of an assembler software. A mapper tries to align every read against the reference genome. This process is much simpler and faster than the assembly. In this case the computer requirements are usually less demanding and the limiting factor could be the storage capacity. Some commonly used mappers are Bowtie, BWA, and TopHat. Once the reads are aligned, single nucleotide polymorphism (SNPs) can be detected by using the SAMtools or the GigaBayes SNP callers [28].

The open source software mainly used by the bioinformaticians is cumbersome for users not well versed in the Unix command line operating system. Some commercial proprietary solutions easier to use have been developed (e.g., LaserGene or CLC Genomics Workbench), but they have not been widely embraced by the breeders. Galaxy is an open source project devoted to create an easy to use web interface to the open source CLI based applications used in this area.

An important amount of work has been devoted in this field to the creation of standard and open file formats capable of storing information regarding sequence alignment and modelling (SAM) [24], SNP calls using variant call format (VCF; http://1000genomes.org/wiki/doku.php?id=1000_genomes: analysis:vcf4.0), genomic regions with browser extensible data (BED, http://genome.ucsc.edu/FAQ/ FAQformat# format1) and genomic annotations using the general feature format (GFF; http://www.sequenceontology.org/resources /gff3.html]). These open and standard formats allow the interoperability of the different software tools that are being actively developed and used. In addition, the computer requirements might be strong as some analyses require a large amount of RAM memory or storage capability.

The algorithms and methods used to store and process raw genomic data generated by the different technological platforms will depend on the type of data being processed and on the result expected. In any case, once the relevant information is obtained by the bioinformaticians, results have to be made available to the breeders by using an interface as easy and friendly as possible [25]. To provide access to this information, the generation of an easily browseable web site is a common and usually successful approach. Several general purpose web databases exist to make the relevant biological information available to the researchers and breeders (Table 2), like GenBank (http://www.ncbi.nlm.nih.gov/genbank/), EBML (http://www.ebi.ac.uk/embl/), DDBJ (http://www.ddbj.nig.ac.jp/) and Swiss-prot (http://expasy.org/sprot/). These latter databases are devoted to store information about any species, but several other more specific databases focused on species of interest to the breeders also exist, like the Sgn (http://solgenomics.net/), Phytozome (http://www.phytozome.net/), Gramene (http://www.gramene.org/) or CropNet (http://ukcrop.net/), which hold information that could have more specific use for breeding programs.

Expression studies, from microarrays to RNA-seq

New genomic tools are also of interest to expand and accelerate gene expression studies. The analysis of gene expression has provided a rich source of biological information, which allows breeders to understand the molecular basis of complex plant processes, leading to the identification of new targets for manipulating these processes.

Gene expression studies were at first based on the classical Northern blot method that only allowed the quantification of tens of genes simultaneously. The QRT-PCR is a more affordable and quantitative technique; but the number of genes analyzed by experiment is also limited [29]. Other approaches allowing the study of thousands of genes were differential display and cDNA amplified fragment length polymorphisms (cDNA-AFLPs) [30]. However, these methods are not really quantitative and are limited by the ability of the developed libraries to capture low-abundance transcripts. Other methods that overcome part of these problems are the serial analysis of gene expression (SAGE) [31] and massively parallel signature sequencing (MPSS) [32]. Nevertheless, the most employed methods at present to analyze transcript profiling are the hybridization-based platforms or microarrays [33]. Expression arrays have several advantages when compared with other methods. They can measure tens of thousands of different transcripts in the same reaction, they are semi-quantitative and sensitive to low-abundance transcripts if those are represented in a given array.

There are several web resources that facilitate microarray data analysis (e.g., http://babelomics.bioinfo.cipf.es/) [34] or even web pages where the breeder can download experiments already performed and analyzed. There are also software packages specializated in microarrays analysis as the Bioconductor (http://www.bioconductor.org/help/workflows/oligo-arrays/) or MeV (ttp://www.tm4.org/mev/) [35]. Probably one of the most useful database is Genevestigator (https://www.genevestigator.com/gv/ doc/plant_biotech.jsp) [36], which contains microarray data from different species. The most extensive data are from the model species *A. thaliana* [37], but an increasing number of studies in crops like maize, wheat, rice, barley, or soybean are already available. All published expression data are public and disposables in databases as GEO (http://www.ncbi.nlm.nih.gov/geo/) [38], ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) [39] or species specific databases. These data can be really useful when analyzing gene expression in these species or other crops [40].

Microarrays make use of the existing EST collections and genome sequence data. The vast increase provided by NGS in the number of sequences opens the possibilities of expression studies in a large number of species lacking previous sequence information. Also, deep NGS sequencing of RNA

transcripts (RNA-seq) is emerging as an alternative to microarray studies to quantify gene expression [41, 42]. RNA-seq does not depend on genome annotation or on the probes contained in the array platform. This technology is also very useful to improve genome annotation, improving the detection of rare transcripts and splicing variants and the mapping of exon/intron boundaries. Moreover, RNA-seq avoids bias introduced during hybridization of microarrays and saturation level problems, haa a greater sensibility, and shows high reproducibility [41, 43] . This approach has been already used in different crops with different breeding objectives, leading to the identification of genes involved in several metabolic pathways, disease response, fruit development, etc. [44, 45, 46, 47]. All these studies show the potential of RNA-seq for complex traits breeding. However, RNA-seq is an emerging technology and the methods used to analyze this kind of data are still being developed.

Mutant and germplasm collections in the genomics era: TILLING and EcoTILLING

Plant breeding requires genetic variability to be selected in order to increase the frequencies of favourable alleles and genetic combinations. Sources of natural genetic variability can be found within the crop, mostly in the form of landraces, and also in the wild relatives. Although many landraces have been substituted by modern and uniform cultivars and genetic erosion has taken place in wild materials, gene banks preserve many of these materials, which constitute an important reservoir of genetic variation useful for breeding [48].

Another important source of variability corresponds to the artificial mutant collections developed in several crop species. These artificial collections are created by radiation, chemical mutagenesis, or transgenic and insertion methods. Artificial mutations, mostly obtained by radiation and chemical methods, were used in plant breeding in the pre-genomics era, but new technologies are allowing the development of other types of collections [49]. For instance, the transferred DNA tagged lines and transposon tagged lines have been used to develop mutant collections in several species such as the model plant *Arabidopsis* (The Arabidopsis Information Resource; http://www.arabidopsis.org) or rice (International Rice Functional Consortium; http://irfgc.irri.org). Gene silencing technologies, using RNA interference, have also been used to create gene specific mutant collections in *Arabidopsis*, like the AGRIKOLA project (http://www.agrikola.org). The artificial mutant collections frequently contain

variability not present in the natural collections, and so are also very useful for the study and development of new traits.

In order to facilitate the identification of the accessions of interest in these collections, a genetic reverse approach has been used. Targeting Induced Local Lesions in Genomes (TILLING) [50] is able to identify all allelic variants of a DNA region present in an artificial mutant collection. A similar procedure called ecotype TILLING (EcoTILLING) [51] can be used to identify allelic variants for targeting genes in natural collections. These two methods are based on the use of endonucleases, such as CEL I or Endo I, that recognize and cut mismatches in the double helix of DNA [52, 53]. Since the TILLING and EcoTILLING techniques identify all allelic variants for a certain genomic region, the phenotypic characterization effort can be concentrated in a reduced number of accessions with different variants. Obviously, the success of the identification of variation useful for breeding programmes will depend on the right selection of target genes. The availability of sequences coming from NGS sequencing projects and the information provided by gene expression studies is significantly increasing the number and quality of candidates for TILLING and EcoTILLING studies

These procedures have been successfully used in many crops [54]. For example, TILLING has been applied to *Arabidopsis* [55], *Lotus* [56], barley [57], maize [58], pea [59], and melon [60]. Rice was the first crop for which EcoTILLING was applied [61]. Subsequently, EcoTILLING has been used in other crops and wild relatives, like barley [62], wheat [63], or the wild peanut *Arachis duranensis* [64], using both genebank collections and natural populations [65]. These assays used gene targets involved in different processes. Many studies have been focused on detecting allelic variants in genes most related to organoleptic quality [66, 67] or disease resistance [68, 69].

Re-sequencing for SNPs discovery and use in genotyping platforms

One of the most interesting applications of NGS for plant breeders is the discovery of genetic variation. Now it is possible to sequence rapidly multiple individuals within a species with limited technical expertise and at minimal cost. The parallel development of computational pipeline tools is greatly accelerating the accurate mining of these sequences for genetic variants that can be converted into genetic markers, mainly microsatellites or simple sequence repeats (SSRs) and SNPs [70]. SSRs and SNPs are now the predominant markers in plant genetic analysis. SNPs are more abundant, stable,

amenable to automation, and increasingly cost-effective, thus are fast becoming the marker system of choice in modern genomics research [71].

The genome-wide SNPs discovery by massive re-sequencing has been performed in model species with small genomes, such as *Arabidopsis thaliana*, where the 1001 Genomes project (http://www.1001genomes.org) [72] is attempting to unveil the whole-genome sequence variation in this reference plant. Similar re-sequencing efforts are becoming possible in rice, maize, grape, soybean, poplar etc. by sequencing sets of related genotypes, individually or pooled, within each species (elite cultivars, breeding lines, ecotypes, landraces, and/or weedy and wild relatives of a crop) [73, 74, 75, 76]. The higher complexity in architecture and repeat content of these genomes has made necessary the use of approaches for genomic complexity reduction that also reduce sequencing cost [77]. Identification of SNPs is also very challenging in species with high levels of heterozygosity and/or with complex ploidy levels, as pseudo-SNPs are identified by misassembly of paralogs [78, 79, 80].

Both Roche 454 and Illumina GA have been mostly used for genome re-sequencing. The aligment difficulties often associtated to the use of short Illumina GA reads (Table 1) are less problematic in species for which available reference genomes facilitates SNPs calling and genome positioning of genetic variation [81]. For most of these species, limited collections of SSRs and SNPs were available from early re-sequencing efforts, previous to the advent of NGS, but new genome-wide re-sequencing is enlarging the SNP pools and making them more representative of the range of natural variation.

For an increasing number of species with high societal or economic value NGS genome re-sequencing is providing the first SSRs and SNPs resources. Examples are the grain amaranths (*Amaranthus* sp.), important pseudocereals, appreciated for their nutritional quality and tolerance to abiotic stress [82], for which no prior genome information existed. In these cases the combination of several sequencing techniques, and the use of paired-end sequencing facilitates SNP discovery. Roche 454 and Illumina GA were combined for high-throughput SNP discovery in common bean [83] and also Solid was used to sequence diploid wheat species, which are donors of the subgenomes of modern hexaploid bread wheat [84].

Most of the effort in species lacking genomic resources is being made through transcriptome re-sequencing, as an alternative way of genome complexity reduction. Targeted amplicon re-sequencing is another strategy for discovering SNPs in specific genes [78].

One of the first examples of deep transcriptome sequencing was a study with two maize inbred lines [85]. This first study was followed by a large and rapidly increasing number of projects using non-model crops, some of them with large, complex, polyploid, and uncharacterized genomes, including forest trees, like *Eucalyptus* [86], oak [87], several polyploid crops, like oilseed rape [79], oats [80], coffee [88], and sweet potato [89], non-model grain legumes as chickpea and chickling pea [90], tomato [91], or several cucurbits, including *Cucurbita* spp. [21], cucumber [92], and melon [93].

These studies employ normalized/non-normalized cDNA libraries generated from single or multi-tissues samples, and derived from single or pooled genotypes, combined or not with multiplex identifier barcodes that allow allele origin identification. Sequencing is often focused on selected genotypes subjected to specific conditions, to detect SNPs in candidate genes involved in complex biological processes of interest to plant breeders. For example, the transcriptomes of two resistant and one susceptible lines of water yam, a major staple crop in Africa, under anthracnose infection, were successfully sequenced detecting SNPs in genes putatively involved in pathogen response [94]. Also, two alfalfa genotypes contrasting for cellulose and lignin content were sequenced, which allowed selecting SNPs useful to improve alfalfa as a forage crop and cellulosic feedstock [95].

The use of genome and transcriptome sequencing for SNP discovery has resulted in large SNPs collections in most of the crops. These large collections are being validated and applied for different purposes such as map construction, map saturation, genome-wide diversity studies, association mapping etc (Table 3). Some of the most important achievements will be described in later sections.

There are many SNPs genotyping techniques, which are more or less appropriate for different scales of individuals/SNPs to be genotyped [107]. The implementation of marker-assisted breeding strategies often requires the generation of thousands of genotypes per population. Thus, one practical way of optimizing the use of these large SNPs collections is using them with cost-effective platforms for medium to high density genotyping. A large number of commercial platforms are available for semiautomated or fully automated SNP genotyping [108, 109]. Genotyping assays usually require a previous process of selection of a set of SNPs, among the hundreds/tens of thousands detected, that are appropriate for the assay objectives.

The Illumina GoldenGate assays have been the most widely used for mid-throughput applications. SNPs platforms with 384, 768, or 1536 SNPs are available for a number of species (Table 3). Popularity is also increasing for the Sequenom Mass array and the KASPar genotyping chemistry [82,

110]. Expanded arrays with tens of thousands SNPs for high-throughput applications have been also developed with the Infinium technology in maize, grape, tomato, pine, and poplar and are under development in soybean and several *Rosaceae* crops (apple, peach, and cherry) [74, 111].

Construction of high density genetic maps

One of the main applications of genomic advances is the development of high density genetic maps. The high-density map construction involves the location of hundreds or even thousand markers in the different linkage groups. In these maps the coverage should be very high and no large gaps must be present. NGS technologies and high-throughput genotyping platforms have allowed the improvement of genetic maps by increasing markers density. Several works include the integration of new markers, basically SNPs derived from re-sequencing studies, into previously developed genetic maps, both in diploid and polyploidy species [80, 112]. Golden Gate has been the most widely used platform. It has been estimated that this genotyping platform is 100-fold faster than gel-based methods for increasing 2-3 times maize map density [101]. Also Sequenom-based SNP-typing assays are starting to be applied. In a recent study, a total of 1.016 SNPs, identified *via* comparative next-generation transcriptomic sequencing, were successfully mapped by genotyping 297 maize recombinant inbred lines (RILs) [113]. Other genotyping strategies based on arrays hybridization, such as the single-feature polymorphisms (SFP), variants detected by a single probe in oligonucleotide arrays, are speeding up genetic map construction. This technique has been used for the construction of a high-density linkage map in species poorly characterized, like *Eucalyptus* [114]. The newly developed maps, enriched in sequence-based markers are facilitating comparative mapping. Recent examples are high density SNPs maps of barley compared with wheat and rice [98, 115].

The decrease of sequencing costs is also allowing the detection of new types of genetic markers useful for increasing the density of genetic maps.In this respect, restriction-site associated DNA (RAD) is a kind of marker which detects genetic variation adjacent to restriction enzyme cleavage sites across a target genome. These markers are produced after NGS sequencing of genomic libraries obtained after digestion with different restrictases. As an example of the utility of this technique, a total of 445 RAD markers distributed across all seven barley chromosomes were located, which was very useful for linkage map construction in this crop [116].

The markers derived from NGS can also be useful to position sequence scaffolds onto physical maps. In this respect, a new method combining deep sequencing and the bin mapping strategy has been described [117]. The SNPs identified by re-sequencing genomic libraries from selected progeny individuals, derived from a cross between two closely related diploid strawberry species, were used to anchor 92.8% of the *Fragaria* genome to the genetic map. Results highlighted the potential of this methodology to obtain a robust framework for the anchoring of the genome sequence without the requirement of a high density physical mapping or a well-saturated genetic map.

Whole-genome re-sequencing at different coverage levels is being increasingly applied for map construction using different strategies. As an example, a genetic map for rice has been constructed using whole genome re-sequencing of 150 RILs [118]. These authors concluded that the sequencing-based method was approximately 35 times more precise in recombination breakpoint determination than PCR-based markers maps. Also, the whole genome of 128 chromosome segment substitution lines (CSSLs) of rice was re-sequenced and used it for the construction of an ultra-high quality physical map in this crop [119]. Based on low coverage re-sequencing, a new mapping strategy that allows inferring the parental genotypes of the assayed RILs population has been proposed [120]. An ultra-high density linkage map was obtained with this method and the quality of the map was evaluated by using it to identify a QTL controlling grain width. Further applications of new sequence-based denser genetic maps to QTL discovery and marker assisted selection (MAS) will be discussed later.

TOWARDS A GENOMICS-BASED PLANT BREEDING

Genome-wide genetic diversity studies

One of the main challenges in agricultural genetics is to access and use the tremendous genetic variation present in germplasm collections and in the wild relatives. A significant part of this variation remains untapped because of the difficulties in effectively identifying genetic differences in large collections. Some traits, with high heritability and of simple characterization, are easy to select for. However, desirable allelic variants and genetic combinations for complex traits are difficult to identify. Recent advances in genotyping are enabling genome-wide diversity studies capable of better capturing the spectrum of variability in natural and breeding populations.

Most of the mid- to high-throughput genotyping platforms described above are being used for studies on diversity and population structure in the corresponding crops (Table 3). By using representative diversity panels, polymorphism rates for individual SNP markers, minor allele frequencies (MAFs), etc. are estimated, facilitating the selection of those SNPs with biological interest and highly polymorphic in the different groups. For example, the Infinium arrays developed in some of these crops are being used to create haplotype maps for vast germplasm collections, such as the 18,000 accessions of the USDA soybean germplasm collection [121].

Haplotype maps (hapmap) of entire collections are useful to identify rare, potentially valuable, alleles. Hapmap projects are undergoing in a number of species such as the "*rice diversity project*" (http://www.ricehapmap.org/index.aspx) aimed to develop a 10,000 SNP chip for rice and create a haplotype map to document the differences in allelic variation within and between the different subpopulations of *O. sativa* and its progenitor *O. rufipogon*. Large-scale genetic diversity studies have also been accomplished in maize. Gore et al. [122] identified and genotyped several million sequence polymorphisms among 27 diverse maize inbred lines. This study allowed the discovery of regions with highly suppressed recombination that appear to have influenced the effectiveness of selection during maize inbred development and may be a major component of heterosis. Also, highly differentiated regions were found that probably contained *loci* that are key to geographic adaptation. Also in legumes, the Medicago HapMap Project, that consist in the sequencing the whole-genomes of 30 inbred lines, will explore the genetic basis of symbiosis, creating a robust platform for genome-scale association mapping [123].

The diversity panels can include representatives of close or more distantly related species to check if these sequence-based SNP assays also work in related species [74, 82]. Sometimes sets of SNPs specifically developed for detecting genetic diversity among closely related cultivars are used in genotyping platforms. For example, despite the large amounts of SNPs available in rice obtained from the comparison of the two reference genomic sequences (one *japonica* and one *indica* variety) [124], extremely low levels of DNA polymorphism were detected within *japonica* cultivars. A whole-genome sequencing of an elite Japanese rice cultivar, closely related to the reference *japonica* variety, has been conducted and the SNP information obtained by comparison of the two *japonica* sequences was applied to develop a high-throughput genotyping array used for genotyping a set of representative Japanese cultivars [125]. These experiments are useful for understanding the role of selection and breeding in the

distribution of genetic variation across the crop genome. In fact, this assay led to the identification of several haplotype blocks which are inherited from traditional landraces to current improved varieties. Moreover, it was found that, as predicted, modern breeding practices have generally decreased genetic diversity. On the practical level, the distribution of genetic diversity in modern cultivars plays an important role in the choice of specific mapping and crop improvement strategies.

Genome-wide survey of genetic diversity is useful to elucidate the causative genetic differences that give rise to observed phenotypic variation providing a foundation for dissecting complex traits through genome-wide association studies. However, its utility is limited if phenotypic data are not available. Not just genomics and transcriptomics, but the other 'omics' disciplines, like proteomics and metabolomics, are useful to understand how the changes in the genotype lead to differences in the final phenotype. Phenomics, which uses high-throughput technologies to characterize germplasm, is being developed and will help to deal with this issue [126].

Identification of molecular markers linked to single genes and QTLs

NGS and high-resolution maps have led to a significant improvement in the identification of molecular markers linked to specific genes and to QTLs. The most important advantage comes from the dense genome coverage, which allows the identification of markers closely linked to any target genomic region, with the advantages that this tight linkage provides.

Methods already used in the pre-genomics era to facilitate the identification of markers linked to single loci, such as bulked segregant analysis (BSA), are now optimized. For example, a GoldenGate assay has been combined with BSA to significantly accelerate mapping of the dominant resistance *locus* to soybean rust *Rpp3* [127]. In this respect, there is an increasing number of reports on exploitation of NGS technologies to identify molecular markers tightly linked to major genes. For example, a fine genetic mapping of the single dominant scab resistance gene (*Ccu*) in RILs of cucumber (*Cucumis sativus*) has been conducted [128]. The resistant cucumber genome was sequenced with Solexa/Illumina NGS and compared with the susceptible cucumber genome. As a result, three insertion/deletion (indel) markers closely linked to the *Ccu locus* where obtained. A detailed study of the region delimited by markers revealed four resistance gene analogs as possible candidates for *Ccu*.

QTL detection has traditionally been conducted by linkage mapping. NGS technologies are significantly contributing to increase accuracy in detection of QTLs. They allow increases in many orders of magnitude of the number of markers mapped, ensuring high mapping resolution, and also aid in the development of mapping populations, such as RILs, near isogenic lines (NILs), and CSSLs, more appropriated for QTLs detection. These populations have conventionally been constructed and genotyped using a limited number of markers.

There are increasing reports describing accurate QTLs mapping with different NGS or high-throughput genotyping strategies. For example, a high density rice map constructed by whole-genome re-sequencing of a RILs population, was used to identify four QTLs controlling plant height [90]. On a different study [129] an ultra-high density genetic map based on SNPs, obtained with Illumina GA, was compared with a linkage map based on RFLPs/SSRs in rice. The positions of several cloned genes, two major QTLs for grain length and grain width, and a QTL for pigmentation were evaluated in a RIL population, arising the expected result that the SNPs map detected more QTLs and more accurately than a RFLPs/SSRs based linkage map.

QTL detection based on the linkage analysis method has the disadvantage that the number of recombination events is limited to the generations needed to develop the mapping population. Association mapping or linkage disequilibrium (LD) mapping is a new powerful approach to map complex traits. This method identifies genetic *loci* associated with phenotypic trait variation in a collection of individuals. Association mapping uses the natural diversity, which represents many more recombination events occurred in the history of the population, providing better resolution. Nowadays, two association mapping methodologies are in use: candidate gene association, where a good understanding of the biochemistry and genetics of the trait is needed, and whole genome scan, also called genome-wide association (GWA) studies. New genomic advances are providing the higher density of genetic markers required to ensure enough coverage to detect linkage between markers and a causal *locus*. Also the decrease of sequencing costs (Table 1) has allowed the use of whole genome sequencing in current studies [130].

Nevertheless, association mapping is just rising in model species and major crops. Maize is the most widely studied crop regarding association analysis. Many candidate genes have been successfully associated to morphological or quality traits. As an example, candidate genes *Dwarf8*, *Vgt1* and *ZmRap2.7* were successfully associated to flowering time [131]. Other candidate genes have been associated, among others, to forage quality, carotenoid content, oil content and kernel quality [132, 133,

134, 135]. GWA studies have been more limited, probably due to the large genome of maize (2300 Mbp) and the great number of markers needed to cover it. The first study identified a fatty acid desaturase gene (*fad2*) associated with increased oleic acid levels [99]. More recently, other authors found 32 QTLs associated with southern leaf blight disease resistance [100].

Examples of association mapping approaches in other crops are more limited. Studies based on the candidate gene approach have been reported in some crops, like grape, or conifers [102, 106]. However, GWA studies have only been developed either in the model species *A. thaliana* [136] or in major crops such as rice [96], barley [97], or wheat [104]. Some articles also describe successful mapping processes combining classical linkage mapping and association mapping [137]. Although genetic association mapping is in its early steps, it is a promising tool for the dissection of complex traits in crop plants.

Marker assisted selection

Marker assisted backcross selection

Marker assisted selection (MAS) is an indirect process where selection is carried out on the basis of a marker instead of the trait itself. The successful application of MAS relies on the tight association between the marker and the major gene or QTL responsible for the trait. As we have described before, the new genomic tools accelerate the identification of markers tightly linked to target genomic regions. On the other hand, the new dense genotyping platforms available today accelerate the genotyping of large amounts of samples during the MAS process in a rapid and economically feasible manner. MAS can take benefit from these technologies, speeding up the release of new varieties.

In spite of the close linkage between the marker and the gene, the possibility of recombination limits the use of MAS. The use of intragenic markers, also called functional markers, can help to overcome this limitation [138]. NGS sequencing projects produce large collections of functional markers. These markers enhance real gene assisted breeding, reducing the possibility of losing the desirable trait due to recombination. This is today feasible in many crop species in which NGS cDNA sequencing is being conducted. Some of these studies perform expression profiling, identifying candidates and associated gene targeted markers.

MAS is also frequently applied to perform background selection in the context of backcrossing programmes. Background selection consists in the identification of plants with lower contents in donor genome to continue the breeding scheme, in order to achieve the recovery of the recipient genome. The use of background markers facilitates the quick recovery of the recurrent parent genome [139]. Background selection is being used extensively in rice breeding. High-density genome maps are being effectively used in background analysis. For example, background selection integrated with foreground selection of bacterial blight resistance (*xa13* and *Xa21* genes), amylose content (*waxy* gene) and fertility restorer gene has been performed in order to identify superior lines with maximum recovery of Basmati rice genome along with the quality traits and minimum non-targeted genomic introgressions of the donor chromosomes [140].

In some cases, the problem of recovering the genetic background of the recurrent parent arises because of the linkage drag, that is, the introgression of chromosome regions with deleterious effects which are tightly linked to the gene or QTL of interest. The detection of QTLs responsible of the negative effects and the localization of molecular markers tightly associated to them can be an efficient way to break the genetic drag. A well known example concerns canola (rapeseed) breeding, which began with the discovery of germplasm with low erucic acid content in seeds of a spring forage cultivar in the 1950's. The problem arose because a high association between low erucic acid content and low seed oil content exists. The recent availability of high-density molecular maps has allowed the detection of several QTLs associated to both traits. Moreover, the identification of molecular markers very tightly linked to the QTLs made possible to break the linkage drag between the low oil content and erucic acid concentration in seeds in the process for breeding new high seed oil content canola cultivars [141].

Frequently, current breeding programmes involve the introgression of more than one gene or QTL controlling traits of interest into one genetic background, in a process that is called pyramiding. The most useful application of MAS in the process of pyramiding is related to the introgression of different genes or QTLs whose effect on the phenotype is undistinguishable. The accumulation of genes from different sources which confer resistance against the same disease is an example, and is indeed one of the most widespread applications of gene pyramiding [142]. The main advantages of recent advances in plant genomics incorporated into gene pyramiding will be related to two different aspects. On one hand, the number of plants to be analyzed in a gene pyramiding programme must be increased as the number of *loci* of interest is higher, to ensure with a reasonable likelihood that the genotype combining favorable alleles

552 is present in the population [143]. In this sense, the availability of genotyping platforms will provide the
553 possibility to screen larger generations. On the other hand, the efficiency of the process strongly depends
554 on the tightness of the linkage between markers used and the target genes or QTLs. Again, identification
555 of functional markers will circumvent this limitation.
556
557 *'Breeding by design'*
558
559 The possibility to predict the outcome of a set of crosses on the basis of molecular markers
560 information is known as 'breeding by design' [6]. The process includes three steps: mapping *loci*
561 involved in all agronomically relevant traits, assessment of the allelic variation at those *loci*, and, finally,
562 breeding by design. In the method as initially described by Peleman and van der Voort [6], the first step
563 was proposed to be completed by either using mapping populations segregating for the trait of interest or
564 based on a candidate gene approach (mainly exploiting information from model plant species and
565 increasing understanding of gene function). Also linkage disequilibrium (LD) mapping was suggested,
566 focused on the region previously identified as related to the trait ('targeted LD mapping'). Currently, as
567 previously discussed, other possibilities such as GWA studies allow a more efficient way to accomplish
568 this first step, avoiding limitations of biparental populations. The second step of the process consists in
569 the identification of allelic variation for the *locus* of interest and the assignation of the phenotypic value to
570 each of them. This step cannot be based on biparental populations, given that only two alleles per *locus*
571 are segregating in this case. The analysis should then include plant materials representing the variability
572 of the species. Genotypic and phenotypic data for each plant are required.
573 As previously stated, high level of saturation with markers is not the limiting factor in most
574 cases, and so currently the restrictions mostly come from the phenotyping step. Strictly speaking,
575 'breeding by design' exploits information obtained in the previous steps: once the *loci* of interest have
576 been mapped, and the contribution of each allelic variant has been determined, crosses can be established
577 to generate superior genotypes which combine all favourable alleles. Application of this breeding strategy
578 has been used for different crops and with different objectives, such as breeding for heading date in rice
579 [144] or seed length in soybean [145]. This procedure has also been used in patent applications; as an
580 example, 'breeding by design' has been reported as part of the development of higher quality maize
581 varieties. However, the most effective application of the 'breeding by design' approach will come from

582 the incorporation of the most advanced genomic tools into the process, which will allow the improvement
583 of the predictions.
584
585 *Genomic selection*
586
587 MAS strategies described so far require the identification of markers associated to the traits of
588 interest. This represents one of the weaknesses of traditional MAS approaches [146]. Nevertheless, MAS
589 can also be applied eluding this step, using an approach known as genomic (or genome-wide) selection
590 (Figure 1). The method was first described in 2001 [147], as an attempt to exploit information generated
591 from emerging genotyping technologies. Genomic selection is based on simultaneous estimation of
592 effects on phenotype of all *loci*, haplotypes, and markers available. The difference with other MAS
593 methods relies on the fact that no previous selection of markers with effects on phenotype is developed
594 [148]. Genomic selection requires the availability of phenotypic and genotypic data for the reference
595 population. This data set will allow estimating the parameters for the model, so that the differences at the
596 phenotype level are explained by the markers analysed. Once the model is established, application to
597 breeding populations makes possible to determine the genomic value of each individual, i.e., the expected
598 phenotype based on the genotypic data. The requirement is the availability of enough molecular markers
599 to provide good genome coverage [5, 146].
600 Simulation studies carried out using maize proved the usefulness of genomic selection applied to
601 an initial cross between an adapted line and exotic germplasm. With 512 markers and a reference
602 population of 288 F_2 plants evaluated in six different environments, it was possible to obtain good
603 selection response after 7-8 generations. [149]. Also with maize, simulations showed that response to
604 selection was 18 to 43% larger for genomic selection than for marker assisted recurrent selection [150].
605 Response obtained when using genomic selection can be lower than response by phenotypic selection.
606 However, the reduction in cycle length due to early MAS results in an increase of gain per time unit. This
607 reduction is even more accused for species with a long generation interval, such as tree species [148].
608 The availability of phenotypic databases for different crops has allowed the comparison of
609 predictions about the genotypic value obtained using genomic selection with the true genotypic value as
610 shown by the phenotypic manifestation of the trait. In a study developed with phenotypic and genotypic
611 data from *Arabidopsis*, maize and barley, results obtained were more accurate when genome-wide

612 selection was carried out, if compared with results derived of previous selection of markers with effects
613 on the phenotype [151].
614 Although when applying genomic selection there is no need to previously identify QTLs
615 controlling a certain trait, the utilization of this approach allows detecting the chromosome regions
616 involved in a given trait, as markers with greater effect on the phenotype will indicate the presence of a
617 QTL for this trait [152]. Some studies go one step farther and propose the application of MAS prior to
618 phenotyping. This approach involves the use of *prior indices*, i.e., marker selection indices which have
619 been constructed from a given phenotyped and genotyped population and are applied to different
620 populations which have not been phenotyped [129]. The decrease in the costs of genotyping provides the
621 appropriate scenario for this strategy to become cost-effective.
622 In any case, even the identification of the QTLs responsible for a certain trait does not imply the
623 identification of the gene or genes controlling the trait itself or the understanding of the mode of action.
624 Models applied in genomic selection are useful to predict breeding values and, in some cases, detect
625 regions associated to a trait, but further work is necessary from this point to identify the gene or genes
626 responsible for the phenotypic variability observed. From plant breeders' perspective, the availability of
627 molecular markers which allow MAS to be applied is generally sufficient. However, development of the
628 new high throughput -omics technologies has provided breeders with new strategies to search for
629 candidate genes, mainly based on microarray for differential gene expression, being the possibility to
630 explore more genes the most important advantage. Future exploitation of these strategies could facilitate
631 the identification of candidate genes underlying the traits of interest and make MAS more efficient.
632
633 **CONCLUSIONS**
634
635 For some major crops the pace experimented for genetic gains in yield and other complex traits
636 in the 20th century will be difficult to be maintained if only existing pre-genomics technologies are used
637 [153]. However, plant breeding is a dynamic science and, fortunately, genomics resources and tools are
638 already available and are helping to give another quantitative leap in plant breeding. In this respect many
639 advances are already taking place, and the superdomestication, i.e., "the processes that lead to a
640 domesticate with dramatically increased yield that could not be selected in natural environments from
641 naturally occurring variation without recourse to new technologies" [10], will require the combination of

conventional breeding with crop genomics. Also, genomic tools and approaches will help conventional breeding in achieving important advances in the breeding of crops that from the point of view of genetic improvement have remained either orphaned or neglected [8]. Therefore, while conventional pre-genomics plant breeding has been, is, and will be successful at improving our crops, the application of genomic tools and resources to practical plant breeding will push forward the genetic gains obtained by breeding programmes. New genomic advances, many of which are already being developed, will make easier for breeders to obtain new cultivars with improved characteristics, either by facilitating selection or by improving the variation available for breeders by using precision breeding approaches. In particular, the present and new genomics tools are of great value for the genetic dissection and breeding of complex traits.

ACKNOWLEDGEMENTS

This contribution has been partially funded by the Ministerio de Ciencia y Tecnología (AGL2008-05114, AGL2009-07257, and PIM2010PKB-00691), and Instituto Nacional de Investigación y Tecnología Agraria y Agroalimentaria (RTA2008-00035-C02-02).

ABBREVIATIONS

AFLP = Amplified Fragment Length Polymorphism

BED = Browser Extensible Data

BSA = Bulked Segregant Analysis

CSSL = Chromosome Segment Substitution Line

DArT = Diversity Arrays Technology

EcoTILLING = Ecotype TILLING

EST = Expressed Sequence Tag

GFF = General Feature Format

GWA = Genome Wide Association

LD = Linkage Disequilibrium

MAF = Minor Allele Frequency

MAS = Marker Assisted Selection

MPSS = Massively Parallel Signature Sequencing

NGS = Next Generation Sequencing

NIL = Near Isogenic Line

QTL = Quantitative Trait Locus

RAD = Restriction-Site Associated DNA

RIL = Recombinant Inbred Line

RNA-seq = Sequencing of RNA Transcripts

SAGE = Serial Analysis of Gene Expression

SAM = Sequence Alignment and Modelling

SPF = Single-Feature Polymorphism

SNP = Single Nucleotide Polymorphism

SSR = Simple Sequence Repeat

TILLING = Targeting Induced Local Lesions on Genomes

VCF = Variant Call Format

REFERENCES

[1] Fedoroff, N.V. The past, present and future of crop genetic modification. *New Biotechnol.* **2010**, *27*, 461-465.

[2] Tester, M.; Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **2010**, *327*, 818-822.

[3] Varshney, R.K.; Tuberosa, R. *Genomics-assisted crop improvement vol. 1: Genomics approaches and platforms.* Springer; New York. **2007a**.

[4] Varshney, R.K.; Tuberosa, R. *Genomics-assisted crop improvement vol. 2: Genomics applications in crops.* Springer, New York. **2007b**.

[5] Lorenz, A.J.; Chao, S.; Asoro, F.G.; Heffner, E.L.; Hayashi, T.; Iwata, H.; Smith, K.P.; Sorrells, M.K.; Jannink, J.L. Genomic selection in plant breeding: knowledge and prospects. Adv. Agron. **2011**, *110*, 77-123.

[6] Peleman J.D.; van der Voort, J.R. Breeding by design. *Trends Plant Sci.* **2003**, *8*, 330-334.

[7] Collard, B.C.Y.; Mackill, D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil. Transact. Royal Soc. B* **2008**, *363*, 557-572.

[8] Varshney, R.K.; Glaszmann, J.C.; Leung, H.; Ribaut, J.M. More genomic resources for less-studied crops. *Trends Biotechnol.* **2010**, *28*, 452-460.

[9] Morgante, M.; Salamini, F. From plant genomics to breeding practice. *Curr. Opinion Biotechnol.* **2003**, *14*, 214-219.

[10] Vaughan, D.A.; Balász, E.; Heslop-Harrison, J.S. From crop domestication to super-domestication. *Ann. Bot.* **2007**, *100*, 893-901.

[11] The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **2010**, *408*, 796–815.

[12] International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature* **2005**, 436, 793-800.

[13] 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, MJ.; 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.; Wilson, R.K. The B73 maize genome: complexity, diversity, and dynamics. *Science* **2009**, *326*, 1112-1115.

[14] 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, K.F.X.; Messing, J.; Rokhsar, D.S. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **2009**, *457*, 551-556.

[15] Tuskan, G.A.; Difazio, S,; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; Schein, J.; Sterck, L.; Aerts, A.; Bhalerao, R.R.; Bhalerao, R.P.; Blaudez, D.; Boerjan, W.; Brun, A.; Brunner, A.; Busov, V.; Campbell, M.; Carlson, J.; Chalot, M.; Chapman, J.; Chen, G.L. Cooper, D.; Coutinho, P.M.; Couturier, J.; Covert, S.; Cronk, Q.; Cunningham, R.; Davis, J.; Degroeve, S.; Déjardin, A.; Depamphilis, C.; Detter, J.; Dirks, B.; Dubchak, I.; Duplessis, S.; Ehlting, J.; Ellis, B.; Gendler, K.; Goodstein, D.; Gribskov, M.; Grimwood, J.; Groover, A.; Gunter, L.; Hamberger, B.; Heinze, B.; Helariutta, Y.; Henrissat, B.; Holligan, D.; Holt, R.; Huang, W.; Islam-Faridi, N.; Jones, S.; Jones-Rhoades, M.; Jorgensen, R.; Joshi, C.; Kangasjärvi, J.; Karlsson, J.; Kelleher, C.; Kirkpatrick, R.; Kirst, M.; Kohler, A.; Kalluri, U.; Larimer, F.; Leebens-Mack, J.; Leplé, J.C.; Locascio, P.; Lou, Y.; Lucas, S.; Martin, F.; Montanini, B.; Napoli, C.; Nelson, D.R.; Nelson, C.; Nieminen, K.; Nilsson, O.; Pereda, V.; Peter, G.; Philippe, R.; Pilate, G.; Poliakov, A.; Razumovskaya, J.; Richardson, P.; Rinaldi, C.; Ritland, K.; Rouzé, P.; Ryaboy, D.; Schmutz, J.; Schrader, J.; Segerman, B.; Shin, H.; Siddiqui, A.; Sterky, F.; Terry, A.; Tsai, C.J.; Uberbacher, E.; Unneberg, P.; Vahala, J.; Wall, K.; Wessler, S.; Yang, G.; Yin, T.; Douglas, C.; Marra, M.; Sandberg, G.; Van de Peer, Y.; Rokhsar, D. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **2006**, *313*, 1596-604.

[16] Jaillon, O.; Aury, J.M.; Noel, B.; Policriti A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; Vezzi, A.; Legeai, F.; Hugueney, P.; Dasilva, C.; Horner, D.; Mica, E.;

Jublot, D.; Poulain, J.; Bruyère, C.; Billault, A.; Segurens, B.; Gouyvenoux, M.; Ugarte, E.; Cattonaro, F.; Anthouard, V.; Vico, V.; Del Fabbro, C.; Alaux, M.; Di Gaspero, G.; Dumas, V.; Felice, N.; Paillard, S.; Juman, I.; Moroldo, M.; Scalabrin, S.; Canaguier, A.; Le Clainche, I.; Malacrida, G.; Durand, E.; Pesole, G.; Laucou, V.; Chatelet, P.; Merdinoglu, D.; Delledonne, M.; Pezzotti, M.; Lecharny, A.; Scarpelli, C.; Artiguenave, F.; Pè, M.E.; Valle, G.; Morgante, M.; Caboche, M.; Adam-Blondon, A.F.; Weissenbach J.; Quétier, F.; Wincker, P.; French-Italian Public Consortium for Grapevine Genome Characterization. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463-7.

[17] Ming, R.; Hou, S.; Feng, Y.; Yu, O.; Dionne-Laporte, A.; Saw, J.H.; Senin, P.; Wang, W.; Ly, B.V.; Lewis, K.L.T.; Salzberg, S.L.; Feng, L.; Jones, M.R.; Skelton, R.L.; Murray, J.E.; Chen, C.; Qian, W.; Shen, J.; Du, P.; Moriah Eustice, M.E.; Tong, E.; Tang, H.; Lyons, E.; Paull, R.E.; Michael, T.P.; Wall, K.; Rice, D.W.; Albert, H.; Wang, M.L.; Zhu, Y.J.; Schatz, M.; Nagarajan, N.; Acob, R.A.; Guan, P.; Blas, A.; Wai, C.M.; Ackerman, C.M.; Ren, Y.; Liu, C.; Wang, J.; Wang, J.; Na, J.K.; Shakirov, E.V.; Brian Haas, B.; Thimmapuram, J.; Nelson, D.; Wang, X.; Bowers, J.E.; Gschwend, A.R.; Delcher, A.L.; Singh, R.; Suzuki, J.Y.; Tripathi, S.; Neupane, K.; Wei, H.; Irikura, B.; Paidi, M.; Jiang, N.; Zhang, W.; Presting, G.; Windsor, A.; Navajas-Pérez, R.; Torres, M.J.; Feltus, F.A.; Porter, B.; Li, Y.; Burroughs, A.M.; Luo, M.C.; Liu, L.; Christopher, D.A.; Mount, S.M.; Moore, P.H.; Sugimura, T.; Jiang, J.; Schuler, M.A.; Friedman, V.; Mitchell-Olds, T.; Shippen, D.E.; dePamphilis, C.W.; Palmer, J.D.; Freeling, M.; Paterson, A.H.; Gonsalves, D.; Wang, L.; Alam, M. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* **2008**, *452*, 991-996

[18] 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.; Jackson, S.A. Genome sequence of the palaeopolyploid soybean *Nature* **2010**, *463*, 178-183.

[19] Metzker, M. Sequencing technologies-the next generation. *Nature Rev. Genet.* **2010**, *11*, 31.

[20] Wang, Z.; Fang, B.; Jingyi, C.; Zhang,X.; Luo, X.; Huang,L.; Chen, X.; Li, Y. De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genoics* **2011**, *11*, 726.

[21] Blanca, J.; Cañizares, J.; Roig, C.; Ziarsolo, P.; Nuez, F.; Picó, B. Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* **2011**, *12*, 104.

[22] Dutta, S.; Kumawat, G.; Singh, B.P.; Gupta, D.K.; Singh, S.; Dogra, V.; Gaikwad, K.; Sharma, T.R.; Raje, R.S.; Bandhopadhya, T.K.; Datta, S.; Singh, M.N.; Bashasab, F.; Kulwal, P.; Wanjari, K.B.; Varshney, R.K.; Cook D.R.; Singh, N.K. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* **2011**, *11*, 17.

[23] Logacheva, M.D.; Kasianov, A.S.; Vinogradov, D.V.; Samigullin, T.H.; Gelfand, M.S.; Makeev, V.J.; Penin, A.A. *De novo* sequencing and characterization of floral transcriptome in two species of buckwheat (*Fagopyrum*). *BMC Genomics* **2011**, *12*, 30.

[24] Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. Genome Project Data P: The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078- 2079.

[25] Pop, M.; Salzberg, S.L. Bioinformatics challenges of new sequencing technology. *Trends Genet.* **2008**, *24*, 142-149.

[26] Horner, D.S.; Pavesi, G.; Castrignanò, T.; Meo, P.D.O.D.; Liuni, S.; Sammeth, M.; Picardi, E.; Presole, G. Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Briefings Bioinformatics* **2009**, *2*, 181-197.

[27] Li, H.; Horner, D.S. A survey of sequence alignment algorithms for next-generation sequencing. *Briefings Bioinformatics* **2010**, *2*, 473-483.

[28] Wang, L.; Li, P.; Brutnell, T.P. Exploring plant transcriptomes using ultra high-throughput sequencing. *Briefings Functional Genomics* **2010**, *9*, 118-128.

[29] VanGuilder, H.D.; Vrana, K.E.; Freeman, W.M. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* **2008**, *44*, 619-626.

[30] Bachem, C.W.; van der Hoeven, R.S.; de Bruijn, S.M.; Vreugdenhil, D.; Zabeau, M.; Visser, R.G. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J.* **1996**, *9*, 745-753.

[31] Anisimov SV. Serial Analysis of Gene Expression (SAGE): 13 years of application in research. *Curr. Pharm. Biotechnol*. **2008**, *9*, 338-350.

[32] Reinartz, J.; Bruyns, E.; Lin, J.Z.; Burcham, T.; Brenner, S.; Bowen, B.; Kramer, M.; Woychik, R. Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. *Briefings Functional Genomics Proteomics*, **2002**, *1*:95-104.

[33] Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **1995**, *270*, 467-470.

[34] Al-Shahrour, F.; Minguez, P.; Vaquerizas, J.M.; Conde, L.; Dopazo, J. BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res*. **2005**, *33*, 460-464.

[35] Saeed, A.I.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; Sturn, A.; Snuffin, M.; Rezantsev, A.; Popov, D.; Ryltsov, A.; Kostukovich, E.; Borisovsky, I.; Liu, Z.; Vinsavich, A.; Trush, V.; Quackenbush, J. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **2003**, *34*, 374-378.

[36] Zimmermann, P.; Laule, O.; Schmitz, J.; Hruz, T.; Bleuler, S.; Gruissem, W. Genevestigator transcriptome meta-analysis and biomarker search using rice and barley gene expression databases. *Mol. Plant.*, **2008**, *1*, 851-857.

[37] Schmid, M.; Davison, T.S.; Henz, S.R.; Pape UJ.; Demar M.; Vingron M.; Scholkopf B.; Weigel D.; Lohmann JU. A gene expression map of Arabidopsis thaliana development. *Nature Genetics* **2005**, *37*, 501-506.

[38] Barrett, T.; Troup, D.B.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Muertter, R.N.; Holko, M.; Ayanbule, O.; Yefanov, A.; Soboleva, A. NCBI GEO: archive for functional genomics data sets -10 years on. *Nucleic Acids Res.*, **2011**, *39*, 1005-1010.

[39] Parkinson, H.; Sarkans, U.; Kolesnikov, N.; Abeygunawardena, N.; Burdett, T.; Dylag, M.; Emam, I.; Farne, A.; Hastings, E.; Holloway, E.; Kurbatova, N.; Lukk, M.; Malone, J.; Mani, R.;

Pilicheva, E.; Rustici, G.; Sharma, A.; Williams, E.; Adamusiak, T.; Brandizi, M.; Sklyar, N.; Brazma A. ArrayExpress update - an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res*., **2011**, *39*, 1002-1004.

[40] Pascual, L.; Blanca, JM.; Canizares, J.; Nuez, F. Analysis of gene expression during the fruit set of tomato: A comparative approach. *Plant Sci.* **2007**, *173*, 609-620.

[41] Marioni, J.C.; Mason, C.E.; Mane, S.M.; Stephens, M.; Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*., **2008**, *18*, 1509-1517.

[42] Stiglic, G.; Bajgot, M.; Kokol, P. Gene set enrichment meta-learning analysis: next-generation sequencing versus microarrays. BMC Bioinformatics **2010**, *11*, 176.

[43] Cloonan, N.; Forrest, A.R.; Kolle, G.; Gardiner, B.B.; Faulkner, G.J.; Brown, M.K.; Taylor, D.F.; Steptoe, A.L.; Wani, S.; Bethel, G.; Robertson, A.J.; Perkins, A.C.; Bruce, S.J.; Lee, C.C.; Ranade, S.S.; Peckham, H.E.; Manning, J.M.; McKernan, K.J.; Grimmond, S.M. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods*., **2008**, *5*, 613-619.

[44] Alagna, F.; D'Agostino, N.; Torchia, L.; Servili, M.; Rao, R.; Pietrella, M.; Giuliano, G.; Chiusano, M.L.; Baldoni, L.; Perrotta. G. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics*, **2009**,*10*, 399.

[45] Zenoni, S.; Ferrarini, A.; Giacomelli, E.; Xumerle, E.; Fasoli, M.; Malerba, G.; Bellin, D.; Pezzotti, M.; Delledonne, M. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Phisiol.* **2010**, *152*, 1787–1795.

[46] Wang Y.; Zhang H.; Li H.; Miao X. Second-generation sequencing supply an effective way to screen RNAi targets in large scale for potential application in pest insect control. *PLoS One* **2011**, 6(4), e18644.

[47] Shi, C.Y.; Yang, H.; Wei, C.L.; Yu, O.; Zhang, Z.Z.; Jiang, C.J.; Sun, J..; Li, Y.Y.; Chen, Q.; Xia, T.; Wan, X.C. Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *BMC Genomics* **2011**, *12*, 131.

[48] Gepts, P. Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci.* **2006**, *46*, 2278-2292.

[49] Parry, M.A.J.; Madgwick, P.J.; Bayon, C.; Tearall, K.; Hernandez-Lopez, A.; Baudo, M.; Rakszegi, M.; Hamada, W.; Al-Yassin, A.; Ouabbou, H.; Labhilili, M.; Phillips, A.L. Mutation discovery for crop improvement. *J. Exp. Bot.* **2009**, *60*, 2817-2825.

[50] Till, B.J.; Reynolds, S.H.; Greene, E.A.; Codomo, C.A.; Enns, L.C.; Johnson, J.E.; Burtner, C.; Odden, A.R.; Young, K.; Taylor, N.E.; Henikoff, J.G.; Comai, L.; Henikoff, S. Large-scale discovery of induced point mutations with high- throughput TILLING. *Genome Res.* **2003**, *13*, 524-530.

[51] Comai, L.; Young, K.; Till, B.J.; Reynolds, S.H.; Greene, E.A.; Codomo, C.A.; Enns, L.C.; Johnson, J.E.; Burtner, C.; Odden, A.R.; Henikoff, S. Efficient discovery of DNA polymorphisms in natural populations by EcoTILLING. *Plant J.* **2004**, *37*, 778-786.

[52] Till, B.J.; Burtner, C.; Comai, L.; Henikoff, S. Mismatch cleavage by single-strand specific nucleases. *Nucleic Acids Res.* **2004**, *32*, 2632-2641.

[53] Triques, K.; Piednoir, E.; Dalmais, M.; Schmidt, J.; Le Signor, C.; Sharkey, M.; Caboche, M.; Sturbois, B.; Bendahmane, A. Mutation detection using ENDO1: Application to disease diagnostics in humans and TILLING and EcoTILLING in plants. *BMC Mol. Biol.* **2008**, 9, 9.

[54] Barkley, N.A.; Wang, M.L. Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. *Curr. Genomics*, **2008**, *9*, 212-226.

[55] Colbert, T.; Till, B.J.; Tompa, R.; Reynolds, S.; Steine, M.N.; Yeung, A.T.; McCallum, C.M.; Comai, L.; Henikoff, S. High-throughput screening for induced point mutations. *Plant Physiol*., **2001**, *126*, 480-484.

[56] Perry, J.A.; Wang, T.L.; Welham, T.J.; Gardner, S.; Pike, J.M.; Yoshida, S.; Parniske, M. A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol*., **2003**, *131*, 866-871.

[57] Caldwell, D.G.; McCallum, N.; Shaw, P.; Muehlbauer, G.J.; Marshall, D.F.; Waugh R. A structured mutant population for forward and reverse genetics in barley (*Hordeum vulgare* L). *Plant J*., **2004**, *40*:143-150.

[58] Weil, C.F.; Monde, R.A. Getting to the point - mutations in maize. *Crop Sci*., **2007**, *47*, S60-S67.

[59] Triques, K.; Sturbois, B.; Gallais, S.; Dalmais, M.; Chauvin, S.; Clepet, C.; Aubourg, S.; Rameau, C.; Caboche, M.; Bendahmane, A. Characterization of *Arabidopsis thaliana* mismatch specific

endonucleases: application to mutation discovery by TILLING in pea. *Plant J.*, **2007**, 51, 1116-1125.

[60] Dahmani-Mardas, F.; Troadec, C.; Boualem, A.; Lévêque, S.; Alsadon, A.A.; Aldoss, A.A.; Dogimont, C.; Bendahmane, A. Engineering melon plants with improved fruit shelf life using the TILLING approach. *PLoS One* **2010**, *5 (12)*, e15776.

[61] Kadaru, S.B.; Yadav, A.S.; Fjellstrom, R.G.; Oard J.H. Alternative EcoTILLING protocol for rapid, cost-effective single-nucleotide polymorphism discovery and genotyping in rice (*Oryza sativa* L.). *Plant Mol. Biol. Reporter* **2006**, *24*, 3-22

[62] Mejlhede, N.; Kyjovska, Z.; Backes, G.; Burhenne, K.; Rasmussen, S.K.; Jahoor, A. EcoTILLING for the identification of allelic variation in the powdery mildew resistance genes *mlo* and *Mla* of barley. *Plant Breed.* **2006**, *125*, 461-467.

[63] Wang, J.; Sun, J.Z.; Liu, D.C.; Yang, W.L.; Wang, D.W.; Tong, Y.P.; Zhang, A.M. Analysis of *Pina* and *Pinb* alleles in the micro-core collections of Chinese wheat germplasm by EcoTILLING and identification of a novel *Pinb* allele. *J. Cereal Sci.* **2008**, *48*, 836-842.

[64] Ramos, M.L.; Huntley, J.J.; Maleki, S.J.; Ozias-Akins, P. Identification and characterization of a hypoallergenic ortholog of Ara h 2.01. *Plant Mol. Biol.* **2009**, *69*, 325-335.

[65] Galeano, C.H.; Gomez, M.; Rodriguez, L.M.; Blair, M.W. CEL I Nuclease Digestion for SNP Discovery and Marker Development in Common Bean (*Phaseolus vulgaris* L.). *Crop Sci.* **2009**, *49*, 381-394.

[66] Wang, J.; Sun, J.Z.; Liu, D.C.; Yang, W.L.; Wang, D.W.; Tong, Y.P.; Zhang, A.M. Analysis of *Pina* and *Pinb* alleles in the micro-core collections of Chinese wheat germplasm by EcoTILLING and identification of a novel *Pinb* allele. *J. Cereal Sci.*, **2008**, *48*, 836-842.

[67] Ramos, M.L.; Huntley, J.J.; Maleki, S.J.; Ozias-Akins, P. Identification and characterization of a hypoallergenic ortholog of Ara h 2.01. *Plant Mol. Biol.*, **2009**, *69*:325-335.

[68] Piron, F.; Nicolaï, M.; Minoïa, S.; Piednoir, E.; Moretti, A.; Salgues, A.; Zamir, D.; Caranta, C.; Bendahmane, A. An induced mutation in tomato eIF4E leads to immunity to two potyviruses. *PLoS One* **2010**, 5 (6), e11313.

[69] Ibiza, V.P.; Cañizares, J.; Nuez, F. EcoTILLING in Capsicum species: searching for new virus resistances. *BMC Genomics* **2010**, *11*, 631.

[70] Deschamps, S.; Campbell, M.A. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Mol. Breed.* **2010**, *25*, 553-570.

[71] Ganal, M.W.; Altmann, T.; Röder, M.S. SNP identification in crop plants. *Curr. Opin. Plant Biol*., **2009**,*12*, 211-217.

[72] Weigel, D.; Mott, R. The 1001 genomes project for *Arabidopsis thaliana*. *Genome Biol*., **2009**, *10*,107.

[73] Gore, M.A.; Wright, M.H.; Ersoz, E.S.; Bouffard, P.; Szekeres, E.S.; Jarvie, T.P.; Hurwitz, B.L.; Narechania, A.; Harkins, T.T.; Grills, G.S.; Ware, D.H.; Buckler, E.S. Large-scale discovery of gene-enriched SNPs. *The Plant Genome* **2009**, *2*, 121-133.

[74] Myles, S.; Chia, J.M.; Hurwitz, B.; Simon, C.; Zhong, G.Y.; Buckler, E.; Ware, D. Rapid genomic characterization of the genus *Vitis*. *PLoS One* **2010**, 5 (1): e8219.

[75] McCouch, S.R.; Zhao, K.; Wright, M.; Tung, C.W.; Ebana, K.; Thomson, M.; Reynolds, A.; Wang, D.; DeClerck, G.; Ali, M.L.; McClung, A.; Eizenga, G.; Bustamante, C. Development of genome-wide SNP assays for rice. *Breed. Sci.* **2010**, *60*, 524-535.

[76] Wu, X.; Ren, C.; Joshi, T.; Vuong, T.; Xu, D.; Nguyen, H.T. SNP discovery by high-throughput sequencing in soybean. *BMC Genomics*. **2010**, 11:469.

[77] Deschamps, S.; La Rota, M.; Ratashak, J.P.; Biddle, P.; Thureen, D.; Farmer, A.; Luck, S.; Beatty, M.; Nagasawa, N.; Michael, L.; Llaca, V.; Sakai, H.; May, G.; Lightner, J.; Campbell, M.A. Rapid genome-wide single nucleotide polymorphism discovery in soybean and rice via deep resequencing of reduced representation libraries with the Illumina genome analyzer. *The Plant Genome*, **2010**, *3*, 1.

[78] Bundock, P.C.; Eliott, F.G.; Ablett, G.; Benson, A.D.; Casu, R.E.; Aitken, K.S.; Henry, R.J. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol. J*. **2009**, *7*, 347-354.

[79] Trick, M.; Long, Y.; Meng, J.;, Bancroft, I. Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol. J.* **2009**, *7*, 334-346.

[80] Oliver, R.E.; Lazo, G.R.; Lutz, J.D.; Rubenfield, M.J.; Tinker, N.A.; Anderson, J.M.; Morehead N.H.W.; Adhikary, D.; Jellen, E.N.; Maughan, P.J.; Brown Guedira, G.L.; Chao, S.; Beattie, A.D.; Carson, M.L.; Rines, H.W.; Obert, D.E.; Bonman, J.M.; Jackson, E.W. Model SNP development

for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. *BMC Genomics* **2011**, *12*, 77.

[81] Shen, Y.; Wan, Z.; Coarfa, C.; Drabek, R.; Chen, L.; Ostrowski, E.A.; Liu, Y.; Weinstock, G.M.; Wheeler, D.A.; Gibbs, R.A.; Yu, F. A SNP discovery method to assess variant allele probability from next-generation resequencing data. *Genome Res*., **2010**, *20*, 273-280.

[82] Maughan, P.J; Smith, S.M.; Fairbanks, D.J.; Jellen, E.N. Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). *The Plant Genome* **2011**, *4*, 92-101.

[83] Hyten, D.L.; Cannon, S.B.; Song, Q.; Weeks, N.; Fickus, E.W.; Shoemaker, R.C.; Specht, J.E.; Farmer, A.D.; May, G.D.; Cregan, P.B. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genomics* **2010**, *11*, 38.

[84] You, F.M.; Huo, N.; Deal, K.R.; Gu, Y.Q.; Luo, M.C.; McGuire, P.E.; Dvorak, J.; Anderson, O.D. Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* **2011**, *12*, 59.

[85] Barbazuk, W.B.; Emrich, S.J.; Hsin, D.; Chen, L.; Li, P.; Schnable, P.S. SNP discovery via 454 transcriptome sequencing. *Plant J.* **2007**, *51*, 910-918.

[86] Novaes, E.; Drost, D.R.; Farmerie, W.G.; Pappas, G.J.Jr.; Grattapaglia, D.; Sederoff, R.R.; Kirst, M. High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* **2008**, *9*, 312.

[87] Ueno, S.; Le Provost, G.; Léger, V.; Klopp, C.; Noirot, C.; Frigerio, JM.; Salin, F.; Salse, J.; Abrouk, M.; Murat, F.; Brendel, O.; Derory, J.; Abadie, P.; Léger, P.; Cabane, C.; Barré, A.; de Daruvar, A.; Couloux, A.; Wincker, P.; Reviron, M.P.; Kremer, A.; Plomion, C. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **2010**, *11*, 650.

[88] Vidal, R.O.; Mondego, J.M.; Pot, D.; Ambrósio, A.B.; Andrade, A.C.; Pereira, L.F.; Colombo, C.A.; Vieira, L.G.; Carazzolle, M.F.; Pereira, G.A. A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential

995 homeologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiol.* **2010**, *154*, 1053-
996 1066.
997 [89] Schafleitner, R.; Tincopa, L.R.; Palomino, O.; Rossel, G.; Robles, R.F.; Alagon, R.; Rivera, C.;
998 Quispe, C.; Rojas, L.; Pacheco, J.A.; Solis, J.; Cerna, D.; Kim, J.Y.; Hou, J.; Simon, R. A sweet
999 potato gene index established by de novo assembly of pyrosequencing and Sanger sequences and
1000 mining for gene-based microsatellite markers. *BMC Genomics* **2010**, *11*, 604.
1001 [90] Garg, R.; Patel, R.K.; Tyagi, A.K.; Jain, M. De novo assembly of chickpea transcriptome using
1002 short reads for gene discovery and marker identification. *DNA Res.* **2011**, *18*, 53-63.
1003 [91] Shirasawa, K.; Isobe, S.; Hirakawa, H.; Asamizu, E.; Fukuoka, H.; Just, D.; Rothan, C.; Sasamoto,
1004 S.; Fujishiro, T.; Kishida, Y.; Kohara, M.; Tsuruoka, H.; Wada, T.; Nakamura, Y.; Sato, S.;
1005 Tabata, S. SNP discovery and linkage map construction in cultivated tomato *DNA Res.* **2010**, *17*,
1006 381-391.
1007 [92] Guo, S.; Zheng, Y.; Joung, J.G.; Liu, S.; Zhang, Z.; Crasta, O.R.; Sobral, B.W.; Xu, Y.; Huang, S.;
1008 Fei, Z. Transcriptome sequencing and comparative analysis of cucumber flowers with different sex
1009 types. *BMC Genomics* **2010**, *11*, 384.
1010 [93] Blanca, J.; Cañizares, J.; Ziarsolo, P.; Esteras, C.; Mir, G.; Nuez, F.; García-Mas, J.; Picó, B.
1011 Melon transcriptome characterization. SSRs and SNPs discovery for high throughput genotyping
1012 across the species. *The Plant Genome*, **2011**, in press.
1013 [94] Narina, S.S.; Buyyarapu, R.; Kottapalli, K.R.; Sartie, A.M.; Ali, M.I.; Robert, A.; Hodeba, M.J.D.;
1014 Sayre, B.L.; Scheffler, B.E. Generation and analysis of expressed sequence tags (ESTs) for marker
1015 development in yam (*Dioscorea alata* L.). *BMC Genomics* **2011**, *12*, 100.
1016 [95] Yang, S.S.; Tu, Z.J.; Cheung, F.; Xu, W.W.; Lamb, J.F.S.; Jung, H.J.G.; Vance, C.P.; Gronwald,
1017 J.W. Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in
1018 two alfalfa genotypes with divergent dell wall composition in stems. *BMC Genomics* **2011**, *12*,
1019 199.
1020 [96] Huang, X.H.; Wei, X.H.; Sang, T.; Zhao, Q.A.; Feng, Q.; Zhao, Y.; Li, C.Y.; Zhu, C.R.; Lu, T.T.;
1021 Zhang, Z.W.; Li, M.; Fan, D.L.; Guo, Y.L.; Wang, A.; Wang, L.; Deng, L.W.; Li, W.J.; Lu, Y.Q.;
1022 Weng, Q.J.; Liu, K.Y.; Huang, T.; Zhou, T.Y.; Jing, Y.F.; Li, W.; Lin, Z.; Buckler, E.S.; Qian,
1023 Q.A.; Zhang, Q.F.; Li, J.Y.; Han, B. Genome-wide association studies of 14 agronomic traits in
1024 rice landraces. *Nature Genet.* **2010**, *42*: 961-976.

[97] Massman, J.; Cooper, B.; Horsley, R.; Neate, S.; Dill-Macky, R.; Chao, S.; Dong, Y.; Schwarz, P.; Muehlbauer, G.J.; Smith, K.P. Genome-wide association mapping of Fusarium head blight resistance in contemporary barley breeding germplasm. *Mol. Breed.* **2011**, *27*: 439-454.

[98] Close, T.J.; Bhat, P.R.; Lonardi, S.; Wu, Y.H.; Rostoks, N.; Ramsay, L.; Druka, A.; Stein, N.; Svensson, J.T.; Wanamaker, S.; Bozdag, S.; Roose, M.L.; Moscou, M.J.; Chao, S.A.M.; Varshney, R.K.; Szucs, P.; Sato, K.; Hayes, P.M.; Matthews, D.E.; Kleinhofs, A.; Muehlbauer, G.J.; DeYoung, J.; Marshall, D.F.; Madishetty, K.; Fenton, R.D.; Condamine, P.; Graner, A.; Waugh, R. Development and implementation of high-throughput SNP genotyping in barley. *BMC Genomics* **2009**, *10*, 582.

[99] Belo, A.; Zheng, P.Z.; Luck, S.; Shen, B.; Meyer, D.J.; Li, B.L.; Tingey, S.; Rafalski, A. Whole genome scan detects an allelic variant of fad2 associated with increased oleic acid levels in maize. *Mol. Genet. Genomics*, **2008**, *279*, 1-10.

[100] Kump, K.L.; Bradbury, P.J.; Wisser, R.J.; Buckler, E.S.; Belcher, A.R.; Oropeza-Rosas, M.A.; Zwonitzer, J.C.; Kresovich, S.; McMullen, M.D.; Ware, D.; Balint-Kurti, P.J.; Holland, J.B. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. *Nature Genet.* **2011**, *43*, 163-168.

[101] Yan, J.B.; Yang, X.H.; Shah, T.; Sanchez-Villeda, H.; Li, J.S.; Warburton, M.; Zhou, Y.; Crouch, J.H.; Xu, Y.B. High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol. Breed.* **2009**, *25*, 441-451.

[102] Emanuelli, F.; Battilana, J.; Costantini, L.; Le Cunff, L.; Boursiquot, J.M.; This, P.; Grando, M.S. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* **2010**, *10*, 241.

[103] Deulvot, C.; Charrel, H.; Marty, A.; Jacquin, F.; Donnadieu, C.; Lejeune-Hénaut, I.; Burstin, J.; Aubert, G. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics.* **2011**, *11*, 468.

[104] Neumann, K.; Kobiljski, B.; Dencic, S.; Varshney, R.K.; Borner, A. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* **2011**, 27, 37-58.

[105] Chao, S.; Dubcovsky, J.; Luo, M.C.; Baenziger, S.P.; Matnyazov, R.; Clark, D.R.; Talbert, L.E.; Anderson, J.A.; Dreisigacker, S.; Glover, K.; Chen, J.; Campbell, K.; Bruckner, P.L.; Rudd, J.C.; Haley, S.; Carver, B.F.; Perry, S.; Sorrells, M.E.; Akhunov, E.D. Population- and genome-specific

patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics.* **2011**, *11*, 727.

[106] Beaulieu, J.; Doerksen, T.; Boyle, B.; Clément, S.; Deslauriers, M.; Beauseigle, S.; Poulin, P.; Lenz, P.; Caron, S.; Rigault, P.; Bicho, P.; Bousquet, J.; Mackay, J. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics, **2011**, *188*, 197-214.

[107] Edenberg, H.J.; Yunlong, L. Laboratory methods for high-throughput genotyping. *Cold Spring Harb Protoc.* **2009**, *16*, 183-193.

[108] Appleby, N.; Edwards, D.; Batley, J. New technologies for ultra-high throughput genotyping in plants. *Methods Mol Biol.* **2009**, *513*, 19-39.

[109] Lin, C.H.; Yeakley, J.M.; McDaniel, T.K.; Shen, R. Medium- to high-throughput SNP genotyping using VeraCode microbeads. *Methods Mol. Biol.* **2009**, *496*, 129-142.

[110] Gabriel, S.; Ziaugra, L.; Tabbaa D. 2009. SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Curr. Protocols Human Genet.*, **Suppl. 60**, 2.12.1-2.12.18.

[111] Eckert, A.J.; van Heerwaarden, J.; Wegrzyn, J.L.; Nelson, C.D.; Ross-Ibarra, J.; Gonzalez-Martinez, S.C.; Neale, D.B. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **2010**, *185*, 969-982

[112] Eckert, A.J.; Pande, B.; Ersoz, E.S.; Wright, M.H.; Rashbrook, V.K.; Nicolet, C.M.; Neale, D.B. High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genet. Genomes*, **2009**, *5*, 225-234.

[113] Liu, S.; Chen, H.D.; Makarevitch, I.; Shirmer, R.; Emrich, S.J.; Dietrich, C.R.; Barbazuk, W.B.; Springer, N.M.; Schnable, P.S. High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. *Genetics* **2010**, *184*, 19-26.

[114] Neves, L.; Mamani, E.; Alfenas, A.; Kirst, M.; Grattapaglia, D. A high-density transcript linkage map with 1,845 expressed genes positioned by microarray-based Single Feature Polymorphisms (SFP) in *Eucalyptus*. *BMC Genomics* **2011**, *12*, 189.

[115] Sato, K.; Nankaku, N.; Takeda, K. A high-density transcript linkage map of barley derived from a single population. *Heredity* **2009**, *103*: 110-117.

[116] Chutimanitsakun, Y.; Nipper, R.W.; Cuesta-Marcos, A.; Cistue, L.; Corey, A.; Filichkina, T.; Johnson, E.A.; Hayes, P.M. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics* **2011**, *12*, 4.

[117] Celton, J.M.; Christoffels, A.; Sargent, D.J.; Xu, X.M.; Rees, D.J.G. Genome-wide SNP identification by high-throughput sequencing and selective mapping allows sequence assembly positioning using a framework genetic linkage map. *BMC Biol.* **2010**, *8*, 155.

[118] Huang, X.H, Feng, Q.; Qian, Q.; Zhao, Q.; Wang, L.; Wang, A.H.; Guan, J.P.; Fan, D.L.; Weng, Q.J.; Huang, T.; Dong, G.J.; Sang, T.; Han, B. High-throughput genotyping by whole-genome resequencing. *Genome Res.* **2009**, *19*: 1068-1076.

[119] Xu, J.J.; Zhao, Q.A.; Du, P.N.; Xu, C.W.; Wang, B.H.; Feng, Q.; Liu, Q.Q.; Tang, S.Z.; Gu, M.H.; Han, B.; Liang, G.H. Developing high throughput genotyped chromosome segment substitution lines based on population whole-genome re-sequencing in rice (*Oryza sativa* L.). *BMC Genomics* **2010**, *11*, 656.

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

[121] Boerma, R.; Wilson, R.; Ready, E. Soybean genomics research program strategic plan implementing research to meet 2012–2016 strategic milestones. *The Plant Genome* **2011**, *4*, 1-11.

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

[123] Branca, A.; Paape, T.; Briskine, R.; Zhou, P.; Wang, S.; Denny, R.; Mudge, J.; Bharti, A.K.; Farmer, A.; May, G.D.; Tiffin, P.L.; Young, N.D. The *Medicago truncatula* HapMap project: deep coverage sequencing of 30 inbred lines using Illumina's Solexa technology. *Plant & Animal Genomes XVIII Conference, San Diego, CA*, **2010**, P417.

[124] McNally, K.L.; Childs, K.L.; Bohnert, R.; Davidson, R.M.; Zhao, K.; Ulat, V.J.; Zeller, G.; Clark, R.M.; Hoen, D.R.; Bureau, T.E.; Stokowski, R.; Ballinger, D.G.; Frazer, K.A.; Cox, D.R.; Padhukasahasram, B.; Bustamante, C.D.; Weigel, D.; Mackill, D.J.; Bruskiewich, R.M.; Rätsch, G.; Buell, C.R.; Leung, H.; Leach, J.E. Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12273-12278.

[125] Yamamoto, T.; Nagasaki, H.; Yonemaru, J.; Ebana, K.; Nakajima, M.; Shibaya, T.; Yano, M. Definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. *BMC Genomics* **2010**, *11*, 267.

[126] Finkel E. Imaging. With 'phenomics,' plant scientists hope to shift breeding into overdrive. *Science*, **2009**, *325*, 380-381.

[127] Hyten, D.L.; Smith, J.R.; Frederick, R.D.; Tucker, M.L.;Song, Q.; Cregan, P.B. Bulked segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. *Crop Sci.* **2009**, *49*, 265-271.

[128] Kang, H.X.; Weng, Y.Q.; Yang, Y.H.; Zhang, Z.H.; Zhang, S.P.; Mao, Z.C.; Cheng, G.H.; Gu, X.F.; Huang, S.W.; Xie, B.Y. Fine genetic mapping localizes cucumber scab resistance gene *Ccu* into an R gene cluster. *Theor. Appl. Genet.* **2011**, *122*, 795-803.

[129] Yu, H.H.; Xie, W.B.; Wang, J.; Xing, Y.Z.; Xu, C.G.; Li, X.H.; Xiao, J.H.; Zhang, Q.F. Gains in QTL detection using an ultra-high density SNP map based on population sequencing relative to traditional RFLP/SSR markers. *PLoS One* **2011**, *6 (3)*, e17595.

[130] Rafalski JA. Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Sci.* **2002**, 162, 329-33.

[131] Buckler, E.S.; Holland, J.B.; Bradbury, P.J.; Acharya, C.B.; Brown, P.J.; Browne, C.; Ersoz, E.; Flint-Garcia, S.; Garcia, A.; Glaubitz, J.C.; Goodman, M.M.; Harjes, C.; Guill, K.; Kroon, D.E.; Larsson, S.; Lepak, N.K.; Li, H.H.; Mitchell, S.E.; Pressoir, G.; Peiffer, J.A.; Rosas, M.O.; Rocheford, T.R.; Romay, M.C.; Romero, S.; Salvo, S.; Villeda, H.S.; da Silva, H.S.; Sun, Q.; Tian, F.; Upadyayula, N.; Ware, D.; Yates, H.; Yu, J.M.; Zhang, Z.W.; Kresovich, S.; McMullen, M.D. The genetic architecture of maize flowering time. *Science* **2009**, *325*, 714-718.

[132] Andersen, J.R.; Zein, I.; Wenzel, G.; Darnhofer, B.; Eder, J.; Ouzunova, M.; Lubberstedt, T. Characterization of phenylpropanoid pathway genes within European maize (*Zea mays* L.) inbreds. *BMC Plant Biol.*, **2008**, *8*, 2.

[133] Harjes, C.E.; Rocheford, T.R.; Bai, L.; Brutnell, T.P.; Kandianis, C.B.; Sowinski, S.G.; Stapleton, A.E.; Vallabhaneni, R.; Williams M.; Wurtzel E.T.; Yan, J.B.; Buckler ES. Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **2008**, *319*: 330-333.

[134] Zheng, P.; Allen, W.B.; Roesler, K.; Williams, M.E.; Zhang, S.; Li, J.; Glassman, K.; Ranch, J.; Nubel, D.; Solawetz, W.; Bhattramakki, D.; Llaca, V.; Deschamps, S.; Zhong, G.Y.; Tarczynski, M.C.; Shen, B. A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nature Genet.* **2008**, *40*: 367-372.

[135] Manicacci, D.; Camus-Kulandaivelu, L.; Fourmann, M.; Arar, C.; Barrault, S.; Rousselet, A.; Feminias, N.; Consoli, L.; Frances, L.; Mechin, V.; Murigneux, A.; Prioul, J.L.; Charcosset, A.; Damerval, C. Epistatic interactions between Opaque2 transcriptional activator and its target gene CyPPDK1 control kernel trait variation in maize. *Plant Physiol.* **2009**, *150*: 506-520.

[136] Atwell, S.; Huang, Y.S.; Vilhjalmsson, B.J.; Willems, G.; Horton, M.; Li, Y.; Meng, D.Z.; Platt, A.; Tarone, A.M.; Hu, T.T.; Jiang, R.; Muliyati, N.W.; Zhang, X.; Amer, M.A.; Baxter, I.; Brachi, B.; Chory, J.; Dean, C.; Debieu, M.; de Meaux, J.; Ecker, J.R.; Faure, N.; Kniskern, J.M.; Jones, J.D.G.; Michael, T.; Nemri, A.; Roux, F.; Salt, D.E.; Tang, C.L.; Todesco, M.; Traw, M.B.; Weigel, D.; Marjoram, P.; Borevitz, J.O.; Bergelson, J.; Nordborg, M. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **2010**, *465*, 627-631.

[137] Simko, I.; Pechenick, D.A.; McHale, L.K.; Truco, L.J.; Ochoa, O.E.; Michelmore, R.W.; Scheffler, B.E. Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. *BMC Plant Biol.* **2009**, *9*: 135-151.

[138] Andersen, J.R.; Lübberstedt, T. Functional markers in plants. *Trends Plant Sci*, **2003**, *8*, 554-560.

[139] Hospital, F.; Chevalet, C.; Mulsant, P. Using markers in gene introgression breeding programs. *Genetics*, **1992**, *132*, 1199-1210.

[140] Gopalakrishnan, S.; Sharma, R.K.; Anand Rajkumar, K.A.; Joseph, M.; Singh, V.P.; Bhat, K.V.; Singh, N.K.; Mohapatra, T. Integrating marker assisted background analysis with foreground selection for identification of superior bacterial blight resistant recombinants in Basmati rice. *Plant Breed.* **2008**, *127*, 131-139.

[141] Cao, Z.; Tian, F.; Wang, N.; Jiang, C.; Lin, B.; Xia, W.; Shi, J.; Long, Y.; Zhang, C.; Meng, J. Analysis of QTLs for erucic acid and oil content in seeds on A8 chromosome and the linkage drag between the alleles for the two traits in *Brassica napus*. *J. Genet. Genomics* **2010**, *37*, 231-240.

[142] Huang, N.; Angeles, E.R.; Domingo, J.; Magpantay, G.; Singh, S.; Zhang, G.; Kumaravadivel, N.; Bennett, J.; Khush, G.S. Pyramiding of bacterial blight resistance genes in rice: marker-assisted selection using RFLP and PCR. *Theor. Appl. Genet.* **1997**, *95*, 313-320.

[143] Ishii, T.; Yonezawa, K. Optimization of the marker-based procedures for pyramiding genes from multiple donor lines: II. strategies for selecting the objective homozygous plant. *Crop Sci*., **2007**, *47*, 1878-1886.

[144] Wei, X.; Liu, L.; Xu, J.; Jiang, L.; Zhang, W.; Wang, J.; Zhai, H.; Wan, J. Breeding strategies for optimum heading date using genotypic information in rice. *Mol Breeding* **2010**, *25*, 287-298.

[145] Lü, H.Y.; Liu, X.F.; Wei, S.P.; Zhang, Y.M. Epistatic association mapping in homozygous crop cultivars. *PLoS One* **2011**, *6 (3)*, e17773.

[146] Jannink, J.L.; Lorenz, A.J. Iwata, H. Genomic selection in plant breeding: from theory to practice. *Briefings Functional Genomics* **2010**, *9*, 166-177.

[147] Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819-1829.

[148] Heffner, E.L.; Sorrells, M.E.; Jannink, J.L. Genomic selection for crop improvement. *Crop Sci.* **2009**, *49*: 1-12.

[149] Bernardo, R. Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci. **2009**, *49*, 419-425.

[150] Bernardo, R.; Yu, J. Prospects for genomewide selection for quantitative traits in maize. Crop Sci, **2007**, *47*, 1082-1090.

[151] Lorenzana, R.E.; Bernardo, R. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor. Appl. Genet.* **2009**, *120*: 151-161.

[152] Bernardo, R. Molecular markers and selection for complex traits in plants: learning from the last 20 years. Crop Sci. **2008**, *48*, 1649-1664.

[153] Araus, J.L.; Slafer, G.A.; Royo, C.; Serret, M.D. Breeding for yield potential and stress adaptation in cereals. *Critical Rev. Plant Sci.* **2008**, *27*, 377-412.

Table 1. Comparison of the main characteristics of the conventional Sanger and some of the most currently used next generation sequencing (NGS) technologies and approximate sequencing cost (in US $ per Mbp).

Technology	Read length (bp)	Mbp per run	Cost ($/Mbp)
Sanger	1000	0.001	3000.00
454 Roche	450	450	66.00
Illumina Hi-Seq2000	100	270000	0.07
Solid 5500xl	50	270000	0.07

1201 Table 2. Some important databases and repositories of genomic information of interest for breeders.

Database	Description	URL
Genbank	General public sequence repository	http://www.ncbi.nlm.nih.gov/genbank/
EMBL	General public sequence repository	http://www.ebi.ac.uk/embl/
DDBJ	General public sequence repository	http://www.ddbj.nig.ac.jp
UniProt	Protein sequences and functional information	http://www.uniprot.org/
NCBI	Biomedical and genomical information	http://www.ncbi.nlm.nih.gov/
Gene Index Project	Transcriptome repository	http://compbio.dfci.harvard.edu/tgi/
GOLD	Repository of genomes databases	http://genomesonline.org/cgi-bin/GOLD/bin/gold.cgi
Phytozome	Genomic plant database	http://www.phytozome.net/
Plantgdb	Genomic plant database	http://www.plantgdb.org
CropNet	Genomic plant database	http://ukcrop.net/
SGN	Solanaceae information resource	http://solgenomics.net/
Gramene	Grass information resource	http://www.gramene.org/
MaizeGDB	Maize infornation resource	http://www.maizegdb.org/
Tair	Arabidopsis information resource	http://www.arabidopsis.org/
CotthonDB	Cotton information resource	http://cottondb.org/
CPGR	Phytopathogen genomic resource	http://cpgr.plantbiology.msu.edu/

1202

1203

1204

1205 Table 3.Some examples of the utility of molecular markers developed by means of high-throughput
1206 genomics techniques for the breeding of important crops.

Crop	Markers	Plant material	Use for breeding	Reference
Rice (*Oryza sativa*)	~$3.6 \cdot 10^6$ SNPs	517 rice landraces	Association studies for 14 agronomic traits	[96]
Barley (*Hordeum vulgare*)	1,536 SNPs	768 breeding lines	Association studies for *Fusarium* head blight resistance	[97]
	3,072 SNPs	336 DH lines and 213 germplasm selections	High-density genetic map construction and MAF estimation	[98]
Maize (*Zea mays*)	8,590 SNPs	553 elite maize inbred lines	Association studies for oleic acid content	[99]
	1,106 SNPs	5,000 RILs	Association studies for resistance to southern leaf blight	[100]
	1,536 SNPs	154 maize inbred lines	Diversity studies	[101]
Grapevine (*Vitis vinifera*)	94 SNPs and 7 indels	148 grape varieties	Association studies for muscat flavor candidate gene VvDXS	[102]
	9000 SNPs	10 cultivated *Vitis vinifera* and 7 wild *Vitis* spp.	Diversity and population structure studies	[74]
Pea (*Pisum sativum)*	384 SNPs	91 RIL mapping population and 373 *Pisum* accessions	Linkage map construction and diversity studies.	[103]
Wheat (*Triticum aestivum*)	874 DArT markers	winter wheat core collection of 96 accessions	Association studies for 20 agronomic traits	[104]

	1,536 SNPs	478 spring and winter wheat cultivars	Diversity studies	[105]
White spruce (*Picea glauca*)	944 SNPs	492 individuals	Association studies with 549candidate genes and 25 wood quality traits	[106]

1207

1208

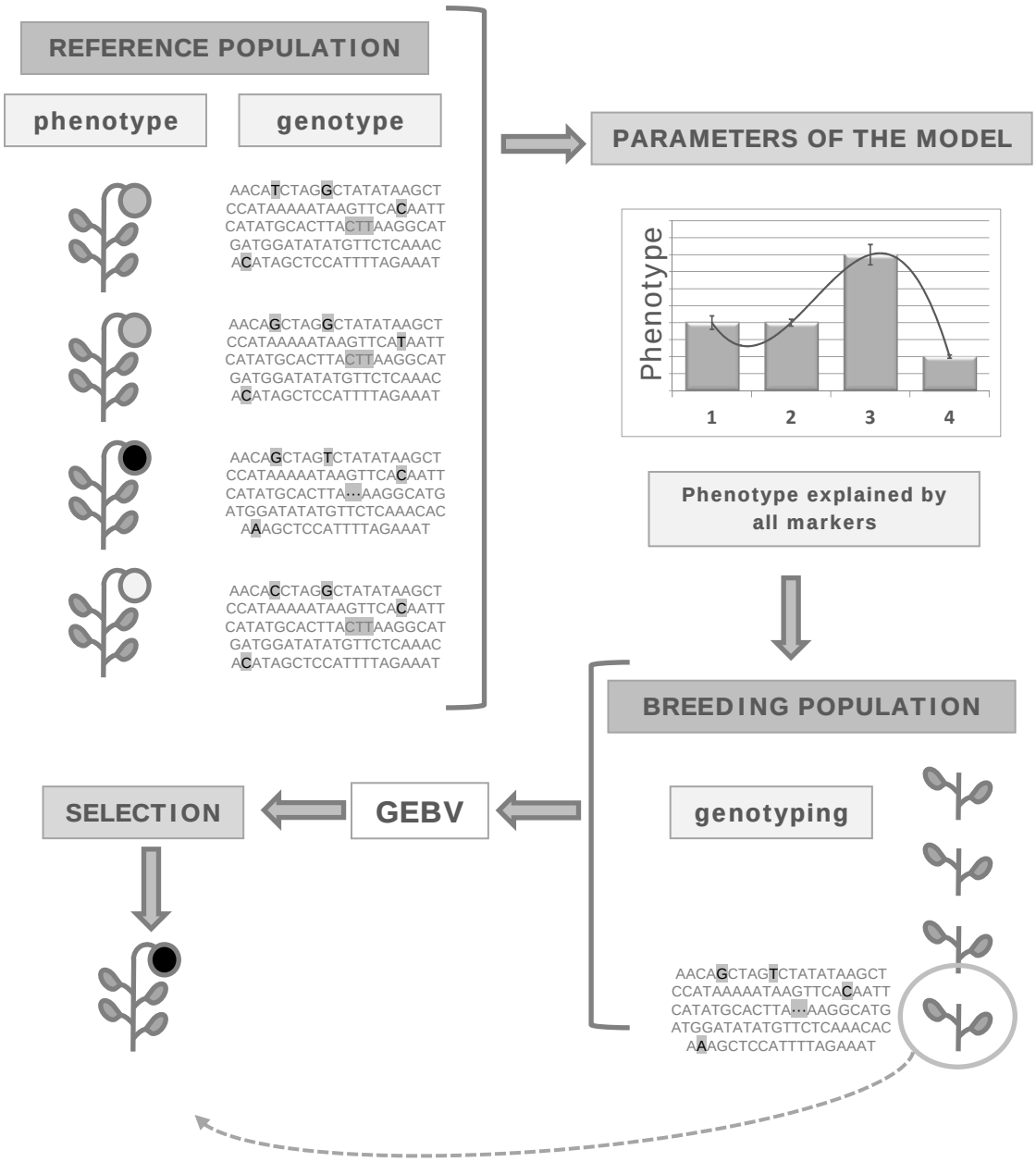

Figure 1. Genomic selection scheme. Information on phenotype and genotype for a reference population allows estimating parameters for the model. This model explains phenotype based on all markers analyzed. The model predicts the phenotype of plants in a breeding population on the basis of the genotyping results: this is the genomic estimated breeding value (GEBV), used to select the desired phenotypes.