

Chapter 2

Molecular Marker Applications in Plants

Alice C. Hayward, Reece Tollenaere, Jessica Dalton-Morgan,
and Jacqueline Batley

Abstract

Individuals within a population of a sexually reproducing species will have some degree of heritable genomic variation caused by mutations, insertion/deletions (INDELs), inversions, duplications, and translocations. Such variation can be detected and screened using molecular, or genetic, markers. By definition, molecular markers are genetic loci that can be easily tracked and quantified in a population and may be associated with a particular gene or trait of interest. This chapter will review the current major applications of molecular markers in plants.

Key words Molecular markers, SNPs, Association mapping, Genetic diversity, Genetic mapping, Marker-assisted selection

1 Introduction

Genetic markers can be used to study patterns of heredity, genomic variation, evolutionary and selection phenomena, allele–allele linkages, and allele–phenotype associations. The application to which a genetic marker is best suited depends on its physical properties and genomic location, the cost involved, ease of use, and degree of throughput required. Molecular markers have been successfully applied in plant science toward the genetic and physical mapping of genomes, the identification of genes controlling various processes and phenotypes (trait association), genetic diversity and evolutionary analyses, and in marker-assisted breeding for crop improvement.

In the past the physical location of a genetic marker was commonly unknown and not necessary for purposes of diversity and evolutionary analyses and breeding applications. With advances in genome sequencing technologies, genetic markers with a known genomic location and environment are becoming more popular and applicable to an increasingly diverse and high-throughput range of objectives. DNA sequence-based markers are inherently

able to capture vast amounts of variation at single-base resolution, making them particularly useful for the detection of perfect markers (DNA polymorphisms causally linked to traits of interest) and discovery and analysis of the alleles involved.

2 Genetic and Association Mapping

2.1 Genetic Linkage Map Construction and QTL Identification

One of the most important applications of genetic markers has been the construction of genetic linkage maps [1, 2]. These maps are created by genotyping a large mapping population of segregating individuals and studying the resulting recombination frequencies between genetic markers. This enables establishment of linkage groups of associated markers with an approximate relative position along a chromosome based on their likelihood of being co-inherited. A linkage group will inherently often represent a large proportion of an individual chromosome with imputed recombination points. The abundance of SNPs and their ability to be discovered and genotyped rapidly in a high-throughput manner makes them particularly valuable markers for genetic mapping [3–5].

Importantly, when the same mapping population used to derive a linkage map is phenotyped for segregating traits of interest, such as seed color or flowering time, the association between marker patterns and the phenotypic variation can be quantified. This then enables identification of the genomic regions controlling traits of interest. Where these traits are quantitative, the associated genomic region(s) are known as quantitative trait loci (QTL). The identification of markers closely linked to genetic loci of interest, including QTL, enables discovery of the underlying, causative gene(s). Prior to the availability of whole genome sequencing technologies, this involved map-based cloning, which used the known sequence of markers directly flanking a locus to amplify and sequence the intervening region for gene candidate identification. Depending on the resolution of the genetic map as defined by marker density and thus distance between flanking markers, this process was often extremely time and resource intensive. Nonetheless, it enabled the first identification of developmentally and agriculturally important genes in many crop and model plant species. In the crop canola, QTL of importance include those for oil yield, oil quality, disease resistance, and pod shatter tolerance, amongst many others [6–9].

2.2 Genome Assembly, Physical Mapping, and Synteny Mapping

Genetic linkage maps are highly valuable in helping to assemble contigs of next-generation genome sequencing data into chromosomes. This is achieved by physically mapping genetic marker sequences on these contigs and comparing this to their known relative location on the genetic map. The success of this process depends on the accuracy and robustness of the genetic linkage

map, as well as the quality of the original contig sequence assembly. Where markers flanking QTL are physically located on a genome sequence, this enables direct and rapid analysis of the intervening region. With the aid of the plethora of *in silico* sequence analysis, gene prediction, and annotation tools currently available, candidate genes underlying these loci can be rapidly identified [10]. Polymorphisms in the candidate gene regions between individuals segregating for the trait can further narrow down the causal gene. Moreover, identification and genotyping of additional SNPs in the original mapping population enables fine-mapping, or extremely high density mapping, of the QTL [11]. SNPs found to be causally associated with a trait variation are known as “perfect markers”, and these, along with the candidate gene, can be then verified *in vitro* and applied to molecular-assisted breeding programs (*see below*).

In species descended from a common ancestor, the preserved order of at least two homologous genes along chromosomes is known as synteny. Synteny mapping uses the locations of conserved genetic markers on the genetic maps of different species to compare interspecies genome organization. This is useful for analyses of gene and genome evolution and in reconstructing ancestral genomes. During evolution, genome rearrangements, expansion, gene loss, and mutation occur at increasing frequency with genetic distance, reducing synteny between distantly related species. When a region of high synteny between species is identified, this suggests a high level of selection for preserving genome sequence and organization in these regions. Such shared synteny is a basic criterion for establishing the functional orthology of genomic regions in different species and can facilitate rapid identification of conserved, agriculturally important gene regions in related crop species. Furthermore, markers associated with different gene paralogues enable localization and comparison of the specific members of multigene family members [12, 13]. Synteny mapping studies were pioneered in grass species [14] but have been conducted in numerous plant species [15–18].

2.3 Association Mapping and Linkage Disequilibrium

Genetic markers that are linked to traits under selection are highly valuable for identifying genetic loci that contribute to phenotypic variation based on linkage disequilibrium (LD). LD refers to the coinheritance of specific genetic markers in ancestrally related individuals at higher frequencies than expected based on recombination distances. Regions that are in high LD may be under high selection pressure for particular allelic combinations, implying a positive relationship between otherwise physically distinct alleles and quantifiable traits. LD mapping, or association mapping, refers to the analysis of statistical associations between genetic markers, usually individual SNPs or SNP haplotypes, and traits (phenotypes) in a collection of individuals [5, 19–21]. SNP haplotypes, which comprise SNP alleles always found in particular allelic

combinations, are found in species with moderate or high levels of LD and may encompass genes or gene clusters [12]. As such, a minimal set of the SNPs normally existing as haplotypes can be used to impute the remainder of the haplotype alleles. This provides the ability to fast-track screening of regions of agronomic interest in breeding programs using a minimal genotyping set [4, 5, 22]. In *Arabidopsis*, identification of linkage disequilibrium based on high-density SNP maps has significantly advanced evolutionary and association genetics studies [23].

Association studies can either be candidate gene-based or whole-genome based [21]. In the candidate gene approach, the aim is to determine correlations between traits of interest and DNA polymorphisms (e.g., SNPs) within candidate genes thought to be involved in those traits. This approach requires prior foresight into the likely biochemistry and genetics of the trait in order to narrow down gene candidates. On the other hand, whole-genome association mapping analyses association of densely mapped genetic markers across all chromosomes with variation in phenotype to identify potential causal or LD associated loci. Association mapping has become popular for identifying trait-marker relationships within many species, particularly for mining new alleles in natural populations or germplasm collections, and/or where the creation of large biparental mapping populations may be less feasible. In this approach, genetic markers are screened across natural populations or a diverse collection of individuals in order to associate alleles with phenotypic traits of interest [24–26]. Since allelic variation in these populations depends on historical recombination and linkage disequilibrium, association studies may produce very high map resolution in species with low levels of LD [27]. LD-based association mapping has been applied in many crop and forage species including maize, barley, wheat, rice, sorghum, sugarcane, sugar beet, soybean, poplar, and grape (reviewed in [19, 28, 29]). In some crop species, for example *Brassica napus*, “diversity fixed foundation sets” have been created, comprising a small number of homozygous lines thought to capture a large proportion of the genetic diversity available for the species. Single nucleotide polymorphisms (SNPs) are currently one of the most popular markers for the fine mapping of heritable traits [30]. The availability of large-scale sequencing and SNP genotyping technologies will support genome-wide association studies in important crop species by enabling screening of large sets of polymorphic markers, even in complex polyploid species [31]. In maize for example, the Illumina Goldengate SNP genotyping assay was used to determine the extent of LD in a diverse global maize collection [32]. Similarly, in black poplar (*Populus nigra*) this same SNP genotyping platform was used to analyze linkage disequilibrium between SNP markers and determine their association to cellulose and lignin biosynthesis properties [33]. This narrowed down candidate genes associated with these traits with the aim of developing a genomics-based breeding platform for bioethanol production.

3 Genomic-Based Breeding

3.1 Marker-Assisted

Selection: Single

Marker–Trait

Associations

Markers provide the potential to fine map important genetic loci with high resolution through the use of mapping populations. Where these populations are phenotyped for traits of agronomic importance, such as disease resistance, the inheritance of particular marker loci or haplotypes in the population can be linked to such phenotypes. Genotyping of markers tightly linked to traits can then rapidly predict the phenotypes of a large selection of segregating individuals at an early stage of development, often well before phenotypic screening would be possible, and at reduced cost. The application of single marker–trait associations to crop breeding is known as marker-assisted selection (MAS). MAS enables efficient selection of breeding lines for the introgression of desirable traits into commercial crop accessions as well as the high-throughput screening of the resulting progeny [34, 35].

An effective marker for MAS must generally be located within 1 cM of a desired trait and able to be genotyped at high throughput and reproducibility [36]. Low polymorphism, poor genomic distribution, and/or poor reproducibility of marker types including RFLPs (Restriction Fragment Length Polymorphisms) and RAPDs (Randomly Amplified Polymorphic DNA) limit their application to MAS. Microsatellites (also known as Simple Sequence Repeats, SSRs) are highly polymorphic, reproducible alternatives, but are often poorly linked to genes [36, 37]. Nonetheless, SSRs and RAPDs have been applied in *B. napus* (canola) MAS programs for selection for major gene disease resistance [38], yellow seed coat color [39], male fertility restorer lines [40], and improvement of oil quality [41].

SNPs are currently the best markers for MAS due to their high prevalence and polymorphism in the genome and their potential for strong, or even perfect, linkage to traits of interest [5, 42–44]. Perfect linkage is possible where the polymorphism is directly responsible for variation in the desired trait. The development of high-throughput sequencing technologies in recent years has greatly assisted association studies that utilize SNP markers [45, 46].

3.2 Genome-Wide

Marker-Assisted

Selection

The availability of phenotypic data along with genotypic data permits the association of loci, or haplotypes, at a genome-wide scale, which may be used to mine an entire genome for genotype–phenotype correlations.

When there are enough markers, spanning the entire genome in a dense manner, it is expected that the gene, or genes, of interest will be in linkage disequilibrium with at least one or some of the markers, leading to marker-assisted selection on a genomic scale [47]. Genome-wide marker-assisted selection studies will be an important way of safeguarding global food supplies into the future. One study performed by Morris and coworkers investigated

agroclimatic traits, such as drought tolerance, within sorghum lines. The study identified ~265,000 SNPs in 971 worldwide accessions, adapted to diverse agroclimatic conditions. Genome-wide association studies (GWAS) based on the markers identified were then carried out to identify novel loci underlying variations in agroclimatic traits [48]. Another study, utilizing restriction-site associated DNA (RAD) sequencing identified 8,207 SNP markers across the Lupin genome, which once filtered, led to the discovery of 38 molecular markers linked to the *Lanr1* disease resistance gene. Sequences involved in the analysis were derived from 20 informative plants resultant of a cross between a disease resistant and a disease susceptible line [49].

Marker-assisted breeding programs implement the introgression of genomic fragments to deliver a desired trait. In some instances, gene pyramiding is utilized, by which one, two, three, or even more genes carrying a particular trait (i.e., pathogen resistance) are introgressed into a hybrid line. Analysis of the resulting degrees of resistance to the pathogen can then be performed. Jiang et al. [50] carried out marker-assisted gene pyramiding on rice cultivars to introgress rice blast resistance genes. Results from this study indicated that the greater the number of resistance genes contained in the improved lines, the higher the resistance to the pathogen and a subsequent growth benefit [50].

Rice, being the major source of caloric intake globally is critical to worlds food supply. GWAS carried out on *Oryza sativa* aim to improve the quality, safety, reliability, and sustainability of this most important crop in a time of population growth, climate change, and the identification of novel agricultural regions. Rice varieties with high stress tolerance, resource-use efficient, and high productivity will be required utilizing a genomics and plant breeding approach. A study carried out by Zhao et al. [51] genotyped 44,100 SNP variants across 413 diverse varieties collected from 82 countries. For these varieties, 34 morphological, developmental, and agronomic traits were systematically phenotyped over two consecutive field seasons [51].

Tomato introgression lines (ILs) derived from the hybridization of wild tomato (*Solanum pennellii*) and cultivated tomato (*Solanum lycopersicum*) resulting in fertile offspring have been extensively used in the identification of interspecific QTL. These publicly available ILs have been comprehensively phenotyped for hundreds of traits thereby allowing the identification of 2,795 QTL [52]. Further analysis of introgression fragments revealed five genomic regions (BINs, 1C, 2B, 4I, 7H, and 11C) that share colinearity, spanning 104 QTL associated with fruit carbon primary metabolism [53, 54], fruit color [55], volatile content [56], and yield traits linked to metabolite variations found in the fruits [57, 58]. Within these syntenic regions 38 distinct genes with conservation of genomic ordering, orientation, and gene structure (intron/

exon) between the two species were observed with variation in intergenic regions disrupting the near perfect colinearity [59]. Sequencing, annotation, and characterization of the genes within these syntenic regions, along with polymorphism and microsyntenic analysis between the genes have unearthed the basis for evolutionary change for the five regions [59], a resource for understanding the possible future value of these introgression fragments and the role that they might play in increasing genetic diversity and availability of desirable traits in crop species. In canola, MAS has enabled selection of intervarietal substitution lines [60] and enrichment of genomic introgression lines [61].

4 Genetic Diversity Analyses

The development and implementation of molecular marker technology has paved the way for large-scale analyses of genetic diversity in and between species. This is valuable for clarifying evolutionary relationships and taxonomies as well as providing an understanding of genome change rates within and between different species. Importantly, the ability to assess genetic diversity in crops also has implications for crop breeding and sustainability [62, 63].

SSRs and SNPs have been widely applied to crop genetic diversity analyses [64–67]. SNPs, as the most common form of highly heritable genetic variation across the genome, are superior indicators of genetic diversity and phylogeny, particularly in crop species with ancient genome duplications. Moreover, genomic SNPs are most often free of selective pressures, allowing a more complete estimate of diversity levels based on random genetic drift [68]. This makes them highly useful in identifying regions of LD and then in tracking chromosome segments to identify recombination events that break up such regions [42, 69]. In maize, an Illumina SNP genotyping assay using over 1,000 SNPs was used to estimate the genetic diversity, population structure, and familial relatedness across a highly diverse global maize collection from temperate, tropical, and subtropical public breeding programs [32]. A similar study in cassava assessed the diversity of 53 varieties from the Americas and Africa to reveal substructure based on geographical origin [70]. In the genus *Arabidopsis*, a genotyping by sequencing approach of 80 diverse accessions from different habitats throughout Eurasia is being used to assess genetic variation contributing to adaptation to diverse environments [71].

4.1 Crop Breeding

For agriculturally important species, a high level of allelic diversity provides an essential resource for mining beneficial trait variants associated with this diversity. In the context of a changing climate, a diverse germplasm set provides a valuable degree of genetic plasticity and adaptive potential for breeding-based crop

improvements and future food security. Unfortunately, extensive artificial selection and inbreeding has severely limited the genetic diversity in many major crop species [63, 72–74]. Canola (*B. napus*), for instance, is a recent allopolyploid that contains only a fraction of the genetic diversity present in its progenitor species' *B. rapa* and *B. oleracea* [75]. Compounding this, inbreeding depression and the associated large blocks of linkage disequilibrium in rapeseed breeding populations have created linkage drag, whereby desirable alleles are inextricably linked to undesirable alleles [25]. As such it has become a priority for many breeders to identify the degree of genetic diversity in not only commercial germplasm, but also wild relatives, of crop species through the use of molecular markers.

Understanding genetic diversity creates great scope for crop improvement and heterosis via wide hybridization and introgression of genetic diversity [76]. Blackleg disease is a fungal disease that devastates canola crops worldwide. Recently, Yu et al. [77] successfully introgressed two known blackleg resistance genes, *LepRI* and *LepR2* from *B. rapa* subspecies *sylvestris* into the related allotetraploid *B. napus* via interspecific hybridizations [77]. Furthermore, diversity analyses enable best choice of lines within germplasm banks for preservation of genetic diversity and breeding potential. In collections of black mustard (*B. nigra*; [66]) and castor bean (*Ricinus communis*; [63]) SSR and SNP markers were used, respectively, to analyze the diversity within geographically distinct populations. A similar study using SSR markers in feral and cultivated alfalfa germplasm concluded that feral alfalfa populations may provide a source of new germplasm for plant improvement [65].

4.2 Comparative Genomics

The ability to compare genomic properties of various evolutionarily related individuals can provide a wealth of information regarding the mechanisms underlying genome evolution, hybridization, polyploidization, and speciation [4, 5]. Molecular markers facilitate rapid and high-throughput comparative genomics analyses and enable analysis of presence/absence variation (PAV), copy number variation (CNV) and, for physically mapped markers such as SNPs, genomic rearrangements between individuals or species. In addition to elucidating the mechanisms and patterns of genome evolution, this information can then be linked to phenotype to better understand the influence of various selective pressures on genome stability and phenotype expression [4]. For example, many genomic regions associated with disease resistance in plants are rapidly evolving due to constant selective pressure from rapidly evolving pathogens. This information can be highly useful for elucidating the genetic basis for disease resistance and coevolution of the pathogen and host plant [78]. Another example is comparing genomic structural change and any associated effects on agricultural vigor in hybrids expressing heterosis for any trait of interest.

4.3 Taxonomic Classification

Comparing the genetic similarity of related species is the most accurate method of resolving taxonomic classifications. Various molecular marker methods provide a fast, high-throughput, and effective means to determine evolutionary relationships at differing resolution. Within the Brassicaceae family, phylogenies remain somewhat confused as a result of recurrent hybridization and polyploidization events [79]. SNPs for high-throughput evolutionary analysis are being applied to resolve ancestral karyotypes in the Brassicaceae and the origin and timing of whole genome duplication and hybridization events [80–82]. The ability to efficiently classify large numbers of samples into species groups also has applications for germplasm banks by facilitating routine verification of stored lines and control of potential contamination. In the study by Pradhan and coworkers [66], SSR markers with known genomic locations from each of the three *Brassica* “A,” “B,” and “C” genomes were used to confirm species identity in a collection of *B. nigra* accessions found to be contaminated with *B. juncea* and *B. rapa* species. Thus, genetic markers with known genomic origin can be valuable for species classification where identification based solely on morphological characters is difficult [66].

5 Complications Arising from Polyploidy

Due to the majority of agriculturally important crop species containing genomes with complex polyploidy, effective SNP discovery can be hampered by possible misidentification of variation between homoeologous (between genome) or paralogous (within genome) loci as true SNPs. Polyploidization events have also resulted in larger genome sizes, with organisms such as maize, barley, and wheat having genome sizes comparable to, or much larger than, humans [4, 83, 84].

During SNP prediction, calibration of the software parameters is required to enable the best trade-off between detection of false positives and the exclusion of some real polymorphism. Many studies have addressed this issue by adjusting the stringency level required for read depth allowing polymorphism detection. In Trick et al. [82], a direct comparison of SNP detection rates at varying stringency levels was presented, demonstrating a large degree of difference. While it is possible to predict a large number of polymorphic sites, sequencing or read mapping errors can produce synthetic polymorphism. Validation of a subset of the predicted SNPs is required to estimate the true rate of variation. This has traditionally been achieved using Sanger sequencing, however it is also possible to utilize higher throughput SNP assays, such as GoldenGate, for this purpose [85, 86].

Despite these hurdles, SNP discovery has been performed in the crop *B. napus* with a validation rate of 95 % and these have then

used to produce successful SNP assays (J. Batley, pers. comm.). The approach used recently in the allogamous species *Juglans regia* was to detect SNPs within one line and then use this SNP pool to genotype populations generated by crossings to that line [87]. The issue of homologous genes within polyploid genomes interfering in SNP discovery has been managed in sugarcane by Bundock et al. [88], directing the discovery effort toward intergenomic SNPs. Two separate sugarcane lines, parents of a mapping population, were sequenced within regions of interest. When used in conjunction with the analysis of wild or progenitor species related to the organism of interest, these can aid in the analysis of evolutionary relationships and are able to provide information on both the diploid and polyploid organisms. The five allele combinations that arrive from either tetraploidy or the existence of two paralogous loci can be accommodated in Illumina's SNP assay software GenomeStudio, however these two scenarios are indistinguishable without prior knowledge. Higher rates of ploidy produce results that are unable to be discriminated into discrete allele combinations.

Complexity reduction is a common strategy used to deal with the issues deriving from complex polyploids where high coverage is required. A reduction in the complexity of the template to be sequenced can be achieved in a number of different ways, depending on the desired approach. Limiting the sequence to expressed sequence tags (ESTs) may produce an appropriate amount of sequence data and can be a useful alternative in gene discovery and has also been employed in crop species lacking a reference genome for SNP discovery [82, 89]. Other complexity reduction methods are based on enzymatic digestion or AFLP (Amplified Fragment Length Polymorphism) amplification using the CROPs system [90].

Although most SNP identification using next-generation sequencing approaches can be utilized without prior knowledge of the reference genome, the sequence capture approach enriches for regions of interest based on predesigned probes, which can describe one contiguous region or many small regions up to a total size limit. A similar output can be generated using long amplicons, sequenced in multiple individuals, such as for Eucalyptus [91] or rice [92]. The best situation is having access to a reference genome for the species of interest, however, a number of groups have developed methods to get around the lack of information. If the focus is on the transcriptome, ESTs can be assembled into a rough draft, using either newly generated data [80] or, if do not have the resources to sequence, it is possible to use publicly available EST data [82]. Broadening the focus to the whole genome, an "on the fly" reference genome has been generated in the wheat progenitor species *Aegilops tauschii* [93], performing a low level of coverage in one individual using the longer reads generated on the Roche 454 platform. The individuals that were the focus of that study could

then be sequenced to a much greater depth on other platforms and the results compared to the generated genome. This enabled the successful prediction of nearly half a million SNPs. A similar approach is to isolate individual chromosome arms and use these as the template. This is a suitable method for size and complexity reduction and has already been used to sequence a number of the large chromosomes of *T. aestivum* [94–97].

6 Conclusion

Molecular markers offer abundant applications in plant molecular genomics and breeding. Despite increasing accessibility and improvement of genome sequencing technologies, molecular markers remain essential components of all large-scale genomic analyses, not only by facilitating genome assembly but via their demonstrated value in high-throughput genotyping, comparative and evolutionary genomics, trait mapping, and plant breeding. As such molecular markers are likely to continue to be developed and successfully applied toward advancing plant genomics for many years to come.

References

1. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376
2. Duran C, Edwards D, Batley J (2009) Genetic maps and the use of synteny. In: Somers DJ, Langridge P, Gustafson JP (eds) *Plant genomics: methods and protocols*. Humana Press, New York, NY, pp 41–56
3. Duran C, Edwards D, Batley J (2009) Molecular marker discovery and genetic map visualisation. In: Edwards D (ed) *Applied bioinformatics*. Springer, New York, pp 165–189
4. Edwards D, Wilcox S, Barrero RA, Fleury D, Cavanagh CR, Forrest KL, Hayden MJ, Moolhuijzen P, Keeble-Gagnere G, Bellgard MI, Lorenc MT, Shang CA, Baumann U, Taylor JM, Morell MK, Langridge P, Appels R, Fitzgerald A (2012) Bread matters: a national initiative to profile the genetic diversity of Australian wheat. *Plant Biotechnol J* 10:703–708
5. Hayward A, Dalton-Morgan J, Mason A, Zander M, Edwards D, Batley J (2012) SNP discovery and applications in *Brassica napus*. *J Plant Biotechnol* 39:49–61
6. Kaur S, Cogan NO, Ye G, Baillie RC, Hand ML, Ling AE, McGearey AK, Kaur J, Hopkins CJ, Todorovic M, Mountford H, Edwards D, Batley J, Burton W, Salisbury P, Gororo N, Marcroft S, Kearney G, Smith KF, Forster JW, Spangenberg GC (2009) Genetic map construction and QTL mapping of resistance to blackleg (*Leptosphaeria maculans*) disease in Australian canola (*Brassica napus* L.) cultivars. *Theor Appl Genet* 120:71–83
7. Pilet ML, Delourme R, Foisset N, Renard M (1998) Identification of loci contributing to quantitative field resistance to blackleg disease, causal agent *Leptosphaeria maculans* (Desm.) Ces. et de Not., in Winter rapeseed (*Brassica napus* L.). *Theor Appl Genet* 96:23–30
8. Qiu D, Morgan C, Shi J, Long Y, Liu J, Li R, Zhuang X, Wang Y, Tan X, Dietrich E, Weihmann T, Everett C, Vanstraelen S, Beckett P, Fraser F, Trick M, Barnes S, Wilmer J, Schmidt R, Li J, Li D, Meng J, Bancroft I (2006) A comparative linkage map of oilseed rape and its use for QTL analysis of seed oil and erucic acid content. *Theor Appl Genet* 114: 67–80
9. Smooker AM, Wells R, Morgan C, Beaudoin F, Cho K, Fraser F, Bancroft I (2011) The identification and mapping of candidate genes and QTL involved in the fatty acid desaturation pathway in *Brassica napus*. *Theor Appl Genet* 122:1075–1090
10. Tollenaere R, Hayward A, Dalton-Morgan J, Campbell E, Lee JRM, Lorenc MT, Manoli S, Stiller J, Raman R, Raman H, Edwards D,

- Batley J (2012) Identification and characterization of candidate *Rlm4* blackleg resistance genes in *Brassica napus* using next-generation sequencing. *Plant Biotechnol J* 10:709–715
11. Choi SR, Teakle GR, Plaha P, Kim JH, Allender CJ, Beynon E, Piao ZY, Soengas P, Han TH, King GJ, Barker GC, Hand P, Lydiat DJ, Batley J, Edwards D, Koo DH, Bang JW, Park BS, Lim YP (2007) The reference genetic linkage map for the multinational *Brassica rapa* genome sequencing project. *Theor Appl Genet* 115:777–792
 12. Edwards D, Batley J, Cogan NOI, Forster JW, Chagné D (2007) Single nucleotide polymorphism discovery. In: Oraguzie N, Rikkerink E, Gardiner S, Silva H (eds) *Association mapping in plants*. Springer, New York, pp 53–76
 13. Love C, Logan E, Erwin T, Kaur J, Lim GAC, Hopkins C, Batley J, James N, May S, Spangenberg G, Edwards D (2006) Integrating and interrogating diverse *Brassica* data within an EnSEMBL structured database. *Proceedings of the joint meeting of the fourteenth crucifer genetics workshop and fourth ishs symposium on Brassicas*. *Acta Hort* 706:77–82
 14. Bevan M, Murphy G (1999) The small, the large and the wild: the value of comparison in plant genomics. *Trends Genet* 15: 211–214
 15. Feuillet C, Keller B (2002) Comparative genomics in the grass family: molecular characterization of grass genome structure and evolution. *Ann Bot* 89:3–10
 16. Galvão VC, Nordstrom KJV, Lanz C, Sulz P, Mathieu J, Pose D, Schmid M, Weigel D, Schneeberger K (2012) Synteny-based mapping-by-sequencing enabled by targeted enrichment. *Plant J* 71:517–526
 17. McClean PE, Mamidi S, McConnell M, Chikara S, Lee R (2010) Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics* 11:184
 18. Zhu HY, Kim DJ, Baek JM, Choi HK, Ellis LC, Kuester H, McCombie WR, Peng HM, Cook DR (2003) Syntenic relationships between *Medicago truncatula* and *Arabidopsis* reveal extensive divergence of genome organization. *Plant Physiol* 131:1018–1026
 19. Abdurakhmonov IY, Abdurakhimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. *Int J Plant Genomics* 2008: 574927
 20. Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol Biol* 57:461–485
 21. Rafalski JA (2010) Association genetics in crop improvement. *Curr Opin Plant Biol* 13: 174–180
 22. Cowling WA, Balázs E (2010) Prospects and challenges for genome-wide association and genomic selection in oilseed *Brassica* species. *Genome* 53:1024–1028
 23. Atwell S, Huang YS, Vilhjalmsso BJ et al (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631
 24. Cardon LR, Bell JI (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91–99
 25. Flint-Garcia SA, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Physiol Plant Mol Biol* 54:357–374
 26. Oraguzie N (2007) An overview of association mapping. In: Oraguzie N, Rikkerink E, Gardiner S, Silva H (eds) *Association mapping in plants*. Springer, New York, pp 1–9
 27. Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends Plant Sci* 9:325–330
 28. Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. *Curr Opin Plant Biol* 12:218–222
 29. Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160
 30. Chagné D, Batley J, Edwards D, Forster JW (2007) Single nucleotide polymorphisms genotyping in plants. In: Oraguzie N, Rikkerink E, Gardiner S, Silva H (eds) *Association mapping in plants*. Springer, New York, pp 77–94
 31. Duran C, Eales D, Marshall D, Imelfort M, Stiller J, Berkman PJ, Clark T, McKenzie M, Appleby N, Batley J, Basford K, Edwards D (2010) Future tools for association mapping in crop plants. *Genome* 53:1017–1023
 32. Yan JB, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4:e8451
 33. Guerra FP, Wegrzyn JL, Sykes R, Davis MF, Stanton BJ, Neale DB (2013) Association genetics of chemical wood properties in black poplar (*Populus nigra*). *New Phytol* 197:162–176
 34. Appleby N, Edwards D, Batley J (2009) New technologies for ultra-high throughput genotyping in plants. In: Somers DJ, Langridge P, Gustafson JP (eds) *Plant genomics: methods and protocols*. Humana Press, New York, NY, pp 19–39

35. Semagn K, Bjornstad A, Ndjioudjop MN (2006) An overview of molecular marker methods for plants. *Afr J Biotechnol* 5:2540–2568
36. Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Mol Breed* 3:87–103
37. Hong CP, Piao ZY, Kang TW, Batley J, Yang TJ, Hur YK, Bhak J, Park BS, Edwards D, Lim YP (2007) Genomic distribution of simple sequence repeats in *Brassica rapa*. *Mol Cells* 23:349–356
38. Chèvre AM, Barret P, Eber F, Dupuy P, Brun H, Tanguy X, Renard M (1997) Selection of stable *Brassica napus*-*B.juncea* recombinant lines resistant to blackleg (*Leptosphaeria maculans*): identification of molecular markers, chromosomal and genomic origin of the introgression. *Theor Appl Genet* 95:1104–1111
39. Somers DJ, Rakow G, Prabhu VK, Friesen KRD (2001) Identification of a major gene and RAPD markers for yellow seed coat colour in *Brassica napus*. *Genome* 1077–1082
40. Hansen M, Hallden C, Nilsson NO, Sall T (1997) Marker-assisted selection of restored male-fertile *Brassica napus* plants using a set of dominant RAPD markers. *Mol Breed* 3: 449–456
41. Tanhuanpää PK, Vilkkilä JP, Vilkkilä HJ (1995) Association of a RAPD marker with linolenic acid concentration in the seed oil of rapeseed (*Brassica napus* L.). *Genome* 38:414–416
42. Barker GLA, Edwards KJ (2009) A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops. *Plant Biotechnol J* 7:318–325
43. Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 3:19
44. Snowden RJ, Friedt W (2004) Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. *Plant Breed* 123: 1–8
45. Syvänen AC (2005) Toward genome-wide SNP genotyping. *Nat Genet* 37:S5–S10
46. Varshney RK, Nayak SN, May GD, Jackson SA (2009) Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol* 27:522–530
47. Meuwissen T (2007) Genomic selection: marker assisted selection on a genome wide scale. *J Anim Breed Genet* 124:321–322
48. Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES, Kresovich S (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc Natl Acad Sci U S A* 110:453–458
49. Yang HA, Tao Y, Zheng ZQ, Li CD, Sweetingham MW, Howieson JG (2012) Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius* L. *BMC Genomics* 13:318
50. Jiang HC, Feng YT, Bao L, Li X, Gao GJ, Zhang QL, Xiao JH, Xu CG, He YQ (2012) Improving blast resistance of Jin 23B and its hybrid rice by marker-assisted gene pyramiding. *Mol Breed* 30:1679–1688
51. Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2:467
52. Lippman ZB, Semel Y, Zamir D (2007) An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr Opin Genet Dev* 17:545–552
53. Schauer N, Semel Y, Balbo I, Steinfath M, Reipsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. *Plant Cell* 20:509–523
54. Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, Willmitzer L, Zamir D, Fernie AR (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. *Nat Biotechnol* 24:447–454
55. Liu YS, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003) There is more to tomato fruit colour than candidate carotenoid genes. *Plant Biotechnol J* 1:195–207
56. Tieman DM, Zeigler M, Schmelz EA, Taylor MG, Bliss P, Kirst M, Klee HJ (2006) Identification of loci affecting flavour volatile emissions in tomato fruits. *J Exp Bot* 57: 887–896
57. Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. *Genetics* 141:1147–1162
58. Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z,

- Gur A, Zamir D (2006) Overdominant quantitative trait loci for yield and fitness in tomato. *Proc Natl Acad Sci U S A* 103:12981–12986
59. Kamenetzky L, Asis R, Bassi S, de Godoy F, Bermudez L, Fernie AR, Van Sluys MA, Vrebalov J, Giovannoni JJ, Rossi M, Carrari F (2010) Genomic analysis of wild tomato introgressions determining metabolism- and yield-associated traits. *Plant Physiol* 152:1772–1786
60. Howell PM, Marshall DF, Lydiate DJ (1996) Towards developing intervarietal substitution lines in *Brassica napus* using marker-assisted selection. *Genome* 39:348–358
61. Zou J, Zhu JL, Huang SM, Tian ET, Xiao Y, Fu DH, Tu JX, Fu TD, Meng JL (2010) Broadening the avenue of intersubgenomic heterosis in oilseed *Brassica*. *Theor Appl Genet* 120:283–290
62. Cowling WA (2007) Genetic diversity in Australian canola and implications for crop breeding for changing future environments. *Field Crop Res* 104:103–111
63. Foster JT, Allan GJ, Chan AP, Rabinowicz PD, Ravel J, Jackson PJ, Keim P (2010) Single nucleotide polymorphisms for assessing genetic diversity in castor bean (*Ricinus communis*). *BMC Plant Biol* 10:13
64. Allan G, Williams A, Rabinowicz PD, Chan AP, Ravel J, Keim P (2008) Worldwide genotyping of castor bean germplasm (*Ricinus communis* L.) using AFLPs and SSRs. *Genet Resour Crop Evol* 55:365–378
65. Bagavathiannan MV, Julier B, Barre P, Gulden RH, Van Acker RC (2010) Genetic diversity of feral alfalfa (*Medicago sativa* L.) populations occurring in Manitoba, Canada and comparison with alfalfa cultivars: an analysis using SSR markers and phenotypic traits. *Euphytica* 173:419–432
66. Pradhan A, Nelson MN, Plummer JA, Cowling WA, Yan GJ (2011) Characterization of *Brassica nigra* collections using simple sequence repeat markers reveals distinct groups associated with geographical location, and frequent mislabelling of species identity. *Genome* 54:50–63
67. Wang J, Kaur S, Cogan NOI, Dobrowolski MP, Salisbury PA, Burton WA, Baillie R, Hand M, Hopkins C, Forster JW, Smith KF, Spangenberg G (2009) Assessment of genetic diversity in Australian canola (*Brassica napus* L.) cultivars using SSR markers. *Crop Pasture Sci* 60:1193–1201
68. Edwards D, Forster J, Chagné D, Batley J (2007) What are SNPs? In: Oraguzie N, Rikkerink E, Gardiner S, Silva H (eds) *Association mapping in plants*. Springer, New York, pp 41–52
69. Fourmann M, Barret P, Froger N, Baron C, Charlot F, Delourme R, Brunel D (2002) From *Arabidopsis thaliana* to *Brassica napus*: development of amplified consensus genetic markers (ACGM) for construction of a gene map. *Theor Appl Genet* 105:1196–1206
70. Ferguson ME, Hearne SJ, Close TJ, Wanamaker S, Moskal WA, Town CD, de Young J, Marri PR, Rabbi IY, de Villiers EP (2012) Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor Appl Genet* 124:685–695
71. Cao J, Schneeberger K, Ossowski S, Gunther T, Bender S, Fitz J, Koenig D, Lanz C, Stegle O, Lippert C, Wang X, Ott F, Muller J, Alonso-Blanco C, Borgwardt K, Schmid KJ, Weigel D (2011) Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* 43:956–U960
72. He GH, Prakash C (2001) Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. *Genet Resour Crop Evol* 48:347–352
73. Hyten DL, Song QJ, Zhu YL, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci U S A* 103:16666–16671
74. Levi A, Thomas CE, Keinath AP, Wehner TC (2001) Genetic diversity among watermelon (*Citrullus lanatus* and *Citrullus colocynthis*) accessions. *Genet Resour Crop Evol* 48:559–566
75. Song K, Osborn TC (1992) Polyphyletic origins of *Brassica napus* – new evidence based on organelle and nuclear RFLP analyses. *Genome* 35:992–1001
76. Chen S, Nelson MN, Chevre AM, Jenczewski E, Li ZY, Mason AS, Meng JL, Plummer JA, Pradhan A, Siddique KHM, Snowden RJ, Yan GJ, Zhou WJ, Cowling WA (2011) Trigenomic bridges for *Brassica* improvement. *Crit Rev Plant Sci* 30:524–547
77. Yu FQ, Gugel RK, Kutcher HR, Peng G, Rimmer SR (2013) Identification and mapping of a novel blackleg resistance locus *LepR4* in the progenies from *Brassica napus* x *B. rapa* subsp. *sylvestris*. *Theor Appl Genet* 126:307–315
78. Hayward A, McLanders J, Campbell E, Edwards D, Batley J (2012) Genomic advances will herald new insights into the *Brassica*: *Leptosphaeria maculans* pathosystem. *Plant Biol* 14:1–10
79. Lysak MA, Koch MA (2011) Phylogeny, genome, and karyotype evolution of crucifers (Brassicaceae). In: Schmidt R, Bancroft I (eds)

- Genetics and genomics of the Brassicaceae. Springer, New York, pp 1–31
80. Hu Z, Huang S, Sun M, Wang H, Hua W (2012) Development and application of single nucleotide polymorphism markers in the polyploid *Brassica napus* by 454 sequencing of expressed sequence tags. *Plant Breed* 131: 293–299
 81. Schranz ME, Song BH, Windsor AJ, Mitchell-Olds T (2007) Comparative genomics in the Brassicaceae: a family-wide perspective. *Curr Opin Plant Biol* 10:168–175
 82. Trick M, Long Y, Meng JL, Bancroft I (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing. *Plant Biotechnol J* 7:334–346
 83. Mayer KFX, Waugh R, Langridge P et al (2012) A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491:711–716
 84. Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
 85. Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, Vanderzande S, Hellens RP, Kumar S, Cestaro A, Velasco R, Main D, Rees JD, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner SE, Bassil N, Peace C (2012) Genome-wide SNP detection, validation, and development of an 8K array for apple. *PLoS One* 7:e31745
 86. Verde I, Bassil N, Scalabrini S, Gilmore B, Lawley CT, Gasic K, Micheletti D, Rosyara UR, Cattonaro F, Vendramin E, Main D, Aramini V, Blas AL, Mockler TC, Bryant DW, Wilhelm L, Troggio M, Sosinski B, Aranzana MJ, Arus P, Iezzoni A, Morgante M, Peace C (2012) Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS One* 7:e35668
 87. You FM, Deal KR, Wang J, Britton MT, Fass JN, Lin D, Dandekar A, Leslie CA, Aradhya M, Luo MC, Dvorak J (2012) Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. *BMC Genomics* 13:354
 88. Bundock PC, Elliott FG, Ablett G, Benson AD, Casu RE, Aitken KS, Henry RJ (2009) Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnol J* 7:347–354
 89. Iorizzo M, Senalik DA, Grzebelus D, Bowman M, Cavagnaro PF, Matvienko M, Ashrafi H, Van Deynze A, Simon PW (2011) De novo assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* 12:389
 90. van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, van der Poel H, van Oeveren J, Versteegen H, van Eijk MJT (2007) Complexity reduction of polymorphic sequences (CRoPS (TM)): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One* 2:e1172
 91. Hendre PS, Kamalakannan R, Varghese M (2012) High-throughput and parallel SNP discovery in selected candidate genes in *Eucalyptus camaldulensis* using Illumina NGS platform. *Plant Biotechnol J* 10:646–656
 92. Kharabian-Masouleh A, Waters DL, Reinke RF, Henry RJ (2011) Discovery of polymorphisms in starch-related genes in rice germplasm by amplification of pooled DNA and deeply parallel sequencing. *Plant Biotechnol J* 9:1074–1085
 93. You FM, Huo N, Deal KR, Gu YQ, Luo MC, McGuire PE, Dvorak J, Anderson OD (2011) 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* 12:59
 94. Berkman PJ, Lai KT, Lorenc MT, Edwards D (2012) Next-generation sequencing applications for wheat crop improvement. *Am J Bot* 99:365–371
 95. Berkman PJ, Skarszewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai KT, Campbell E, Kubalakova M, Simkova H, Batley J, Dolezel J, Hernandez P, Edwards D (2012) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. *Theor Appl Genet* 124:423–432
 96. Hernandez P, Martis M, Dorado G, Pfeifer M, Galvez S, Schaaf S, Jouve N, Simkova H, Valarik M, Dolezel J, Mayer KFX (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. *Plant J* 69:377–386
 97. Lai K, Berkman PJ, Lorenc MT, Duran C, Smits L, Manoli S, Stiller J, Edwards D (2012) WheatGenome.info: an integrated database and portal for wheat genome information. *Plant Cell Physiol* 53:e2

Plant Genotyping

Methods and Protocols

Batley, J. (Ed.)

2015, XI, 315 p. 55 illus., 34 illus. in color., Hardcover

ISBN: 978-1-4939-1965-9

A product of Humana Press