

New Developments in Molecular Techniques for Breeding in Ornamentals

Ornamental Crops

Smulders, Marinus J.M.; Arens, Paul

https://doi.org/10.1007/978-3-319-90698-0_9

This article is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this article please contact openscience.library@wur.nl

Chapter 9

New Developments in Molecular Techniques for Breeding in Ornamentals



Marinus J. M. Smulders and Paul Arens

Abstract In ornamental crops the development of genetic and molecular tools for breeding has been slow because of the large number of ornamental species, many of which are genetically complicated for breeding, being outbreeding crops, polyploid, and/or having a large genome.

This is changing due to three recent developments: (i) next-generation sequencing can now generate large numbers of single nucleotide polymorphism (SNP) markers based on genomic or transcriptomic sequences, (ii) efficient and automated SNP detection systems render genotyping into an automated and relatively cheap process, and (iii) methods and software now exist to analyse these data, also in polyploid crops, to find associations with traits and to generate tools for marker-assisted breeding. The challenge for the coming years will be to implement these tools to speed up breeding.

When more genome sequences of ornamental species or related species become available, it will also be possible to move from associated markers (for a trait or QTL region) to the underlying variation in the causal genes. Knowledge of the existing variation in functional alleles will make it possible to consider directing biosynthetic or regulatory pathways towards, e.g. different colour or scent combinations.

New plant breeding techniques (also called ‘precision breeding techniques’) add new possibilities to direct the breeding process. Notably, gene editing (also called genome editing) using Crispr/Cas may be used to increase the pool of functional variation, but there are challenges to apply it in ornamentals, in terms of the availability of sequence information for the candidate genes and the existence of transformation and regeneration protocols.

Keywords SNP marker · QTL · Marker-assisted breeding · DNA-informed breeding · Functional variation · Plant breeding

M. J. M. Smulders (✉) · P. Arens
Wageningen University & Research, Wageningen, The Netherlands
e-mail: rene.smulders@wur.nl

9.1 Introduction

In ornamental crops the development of genetic and molecular tools for breeding has been lagging behind large agricultural and horticultural crops, because of a combination of factors. These include: the efforts (discovery research; marker generation, production of genetic maps, and mapping of traits) are spread over many different ornamental species; most species are outbreeders that are vegetatively propagated with high levels of genetic diversity; some have a long juvenile phase; and many are polyploid and/or have large genomes. In addition, morphological characteristics are important traits for selection during ornamental breeding of new cultivars, and these can be assessed without the use of markers (Arens et al. 2012), so the need to invest in these tools has been limited. This is now changing as breeding also has to focus on traits that are difficult to assess and/or are controlled by multiple loci (quantitative traits, QTL), such as disease resistances but also quality characteristics including stem production, time to flowering, and flower size (Smulders et al. 2012).

The potential for developing molecular tools has improved extremely in the past decade due to three developments: (i) the progress in next-generation sequencing technologies that now make it easy to generate large numbers of single nucleotide polymorphism (SNP) markers based on genomic or transcriptomic sequences, (ii) the development of efficient and automated SNP detection systems for genotyping, and (iii) the development of methods and software to analyse these data, also in polyploid crops, to find associations with trait phenotypes and generate tools for marker-assisted breeding. As a result, nowadays it is possible to genotype a large number of samples at a few loci with low costs. The challenge for the coming years will be to implement the use of these tools in ornamental crops. This includes developing markers in crops, finding associations with traits, and implementing the results to speed up breeding. It is likely that many examples of actual use of markers will concern disease resistances, as the use of plant protection agents is under increasing pressure. Disease resistances are expensive or difficult to evaluate and combine in crosses unless supported by DNA markers. An example is fusarium resistance in bulb crops. Whereas fusarium resistance is controlled by nucleotide binding site leucine-rich repeat (NBS-LRR)-type disease resistance genes in tomato, the trait is a multi-QTL trait in the important ornamental bulbs lily and tulip, in which the disease resistance is controlled by at least six QTLs (Shahin et al. 2010; Tang et al. 2015). Markers for each of these QTLs would enable selecting progeny that have inherited a combination of these loci. As this can be done in the seedling stage, this would speed up the breeding process, especially in bulbous ornamentals, which have a long juvenile phase (Smulders et al. 2012). Unfortunately, the large QTL regions and the high genetic diversity present in most ornamentals mean that it is not easy to convert these QTLs into an efficient marker-assisted breeding (MAS) application.

With complete genome sequences of ornamental species or species related to them becoming available, it will also be possible to move from a QTL region or associated marker to the underlying variation in candidate genes. Knowledge of the

existing variation in functional alleles in the genes that form biosynthetic or regulatory pathways will make it possible to consider directing these pathways towards, e.g. different colour or scent combinations but also to develop broad host pathogen resistance (Fu et al. 2016).

Finally, new plant breeding techniques (also called ‘precision breeding techniques’) may add new possibilities to direct the breeding process in ornamentals. Notably, gene editing using Crispr/Cas may offer possibilities to expand the pool of genetic variation available for breeders, but there may be challenges to apply it, not the least at the level of cell biology, as effective transformation but especially regeneration protocols are required. Here, too, the increasing availability of sequence data of crops is a stimulant, as gene sequences are needed for target design.

In this chapter, we will describe these developments and indicate what it may mean for the future of ornamental breeding. Where applicable we will refer to useful literature elsewhere.

9.2 DNA Markers

A ‘marker’ is any difference in the DNA whose inheritance can be followed in crosses. Some may be statistically associated with phenotypic traits. For these traits the marker can be used as a proxy for the trait (in case of a single dominant gene) or for a component of it (in case several QTLs underlie a trait) in marker-assisted breeding.

Sequence differences of a single nucleotide are called ‘single nucleotide polymorphism’ (SNP). They have turned out to be the most versatile form of genetic variation used as markers, as they are present in vast numbers in the genome and their detection can be automated. This is in contrast to the older marker systems, which all required manual scoring. Therefore, we now consider SNPs the marker of choice, even when some of the older marker systems have a higher information content (e.g. microsatellite markers often have multiple alleles, while SNPs generally only have two alleles on one position, which is a limitation in polyploid crops that can have multiple alleles at a single locus).

To identify SNPs in inbreeding crops, DNA sequences of two or more plants can be compared. Low-depth sequencing is often sufficient for reliable SNP detection and subsequent use in introgression breeding. However, most ornamentals are outbreeding and thus they are heterozygous. That means that even within a single diploid genotype, there will be many sites at which the sequences of the two chromosomes are different. Therefore, an analysis of sequences obtained from even a single plant can be mined to find SNPs between the homologous chromosomes. An analysis of sequences of two plants (e.g. the two parents of a cross) will yield both SNPs within each of the parents and SNPs between the parents. Higher-depth sequencing over multiple genotypes is needed to be able to identify reliable SNPs (i.e. distinct from sequence mistakes) and to distinguish paralogs from alleles by, for instance, transcriptome sequencing (e.g. see Shahin et al. 2012a; Koning-Boucoiran et al. 2015; Van Geest et al. 2017b).

9.3 Developments in High-Throughput Sequencing

Sequencing technology has developed tremendously since the completion of the first human genome sequence (Schmutz et al. 2004). DNA sequencing has become cheaper and faster, thanks to the development of next-generation sequencing techniques. This development is an ongoing process, and novel and significantly cheaper technology opens up new applications, not only for problems for which the technology used to be too expensive but also for completely novel applications that were either done by other technology or were non-existent or impossible before.

Goodwin et al. (2016) explain the various technologies underlying the steady increase in throughput and the trade-offs associated with it, in the form of shorter read lengths and/or reduced accuracy. Along with increased output, there was a steep decrease in costs: to sequence a human genome now costs 1000 dollar in pure sequence expenses versus 100 million for the first genome (<https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/>). However, raw sequencing costs do not say everything, as there also costs associated with storage and analysis (Muir et al. 2016). When talking about ‘cheaper’ technology for sequencing, one usually considers the basic costs of producing base-pair calls, which drop faster than the reduction in costs for computer storage. Fortunately, in the wake of new sequencing technology, bioinformatics has also made significant improvements in genome assembly and annotation and in variant calling as well. This is an important element that facilitates sequencing of heterozygous and/or large genomes. This means that now, for the first time, it is feasible to generate high-quality genome sequences for the most important ornamental species.

For developing SNP markers, sequencing of short reads (a few hundred base-pairs long) using the Illumina technology is currently by far the cheapest method. Longer reads (1000–10,000 s of base-pairs) can be produced at roughly tenfold higher cost using PacBio or MinION (Oxford Nanopore). These reads are used, in combination with short reads, for high-quality genome assemblies (see below), RNA splicing variant analyses, or long-range haplotyping.

For SNP marker development, we only need to sequence a small part of the genome of a species, but we do want to sequence the same part for several genotypes in order to be able to screen for reliable SNPs and to avoid SNPs caused by paralogs (different genes in the same genome). One strategy is to focus on specific genes, e.g. candidate genes for the trait of interest, amplified by PCR. If the purpose is to develop a genome-wide set of markers, high-throughput sequencing developments have made this easy for crops with small genomes. Up to roughly the human haploid genome size (~3 Gb) one can whole genome resequence a few plants for detection of SNPs between them or between pools of plants, as, e.g. in the bulked segregant analysis (BSA) approach. For larger genomes a whole genome resequencing approach is still too expensive, but BSA approaches are feasible.

An easy method for reducing the complexity of the DNA, i.e. to limit the DNA you are interested in to a small subset of the genome, is to sequence cDNA made from mRNAs expressed in one or more tissues, commonly referred to as RNA-seq.

RNA-Seq can generate numerous transcripts with sufficient read depth to guarantee high-quality SNP identification (Shahin et al. 2012a; Kim et al. 2014). The advantages of this method include: (i) the SNPs will be from the gene-containing part of the genome, which is where chromosomal recombinations occur during meiosis; (ii) in genes possibly fewer flanking SNPs occur that may impede genotyping; (iii) it is suitable for building genomic resources in orphan crops (e.g. for candidate gene approaches), as no prior knowledge is necessary. All one needs to do is extract RNA from one or more tissues from two or more plants, synthesise DNA, and send this for next-generation sequencing. Not surprisingly, there are now many examples of RNA-seq studies across plant species that each generated 10,000 s of SNPs. For ornamentals the species include *Alstroemeria* (Shahin et al., in prep.), *Begonia* (Arens et al., in prep.), *Caladium* (Cao and Deng 2017), *Chrysanthemum* (Van Geest et al. 2017b), *Gerbera* (Fu et al. 2016), *Lilium* and *Tulipa* (Shahin et al. 2012b), *Mei* (*Prunus mume*; Zhang et al. 2017), *Phalaenopsis* (Huang et al. 2016), *Rosa* (Smulders et al. 2015; Koning-Boucoiran et al. 2015), *Silene* (Casimiro-Soriguer et al. 2016), and *Zantedeschia* (Wei et al. 2016). Alternative strategies that are being used for complexity reduction include DArTseq, capture hybridisation of a targeted subset of the genome, and SLAF-seq (Specific Locus Amplified Fragment sequencing) as applied in *Chrysanthemum* (Chong et al. 2016).

The assembly of the short reads can be difficult in ornamentals because of the high diversity that exists in most of the crops. Currently most assemblies are done with Trinity. To identify SNPs, software programs have been developed (e.g. QualitySNPng; Nijveen et al. 2013) that will attempt to distinguish ‘true’ from ‘false’ SNPs. False SNPs can be due to sequencing errors, but they can also be generated when two paralogous genes from the same genome are compared. This is a problem in diploid as well as polyploid species. The problem can be alleviated to some extent by sequencing multiple genotypes from either a population or from the germplasm (e.g. see Shahin et al. 2012b; Van Geest et al. 2017b).

9.4 Developments in Genotyping Technology

The ultimate genotyping is sequencing of the complete genomes of our ornamental plants, but for breeding it is sufficient to use markers as a proxy for parts of the genome. For genotyping of SNP markers, various methods have been developed, the choice of which partly depends on the number of SNPs to be detected.

Dense genome coverage is obtained using set of thousands of polymorphic SNPs. It is cost-effective to genotype them using a SNP array. Currently two systems are available: Infinium arrays from Illumina (arrays exist, e.g. for maize, rice, potato, and apple) and Axiom arrays from Affymetrix (e.g. for apple, strawberry, rose, and chrysanthemum). Custom arrays can be purchased for the crop of interest as well. With Illumina technology this starts with a minimum of about 1200 samples to be genotyped; Axiom arrays can be produced starting from 488 samples (4 arrays of 96).

SNP arrays have been used to produce high-density genetic maps for QTL mapping in, e.g. rose (Vukosavljev et al. 2016; Bourke et al. 2017) and chrysanthemum (Van Geest et al. 2017a, 2017b). The rose array has also been used for association mapping (Schulz et al. 2016; Nguyen et al. 2017).

Not all applications require such a large number of markers. For identification purposes, a few hundred SNPs will suffice, if the frequency of both alleles is balanced to provide sufficient power (Van Eck et al. in preparation). Once QTLs have been found, a small number of linked markers are sufficient to track the genomic region in offspring of crosses. For selection against chromosomes during introgression breeding, four well-spread markers per chromosome suffice for negative selection.

When up to hundred selected SNPs need to be screened in a large number of plants (hundreds to thousands), techniques such as KASP (e.g. Koning-Boucoiran et al. 2012; Holdsworth and Mazourek 2015) or Fluidigm (Jung et al. 2017) are most cost-effective. Amplicon sequencing recently has advanced in such a way that it can be the bridge between Kompetitive Allele Specific PCR (KASP) markers and SNP arrays in numbers of markers that can be cost-effectively genotyped (e.g. see AgriSeq Targeted Genotyping By Sequencing (Thermo Fisher) or PlexSeq (Agriplexgenomics) or other service providers).

9.5 Methods and Software Development

The output of a SNP genotyping assay is, for each marker, the ratio between the hybridisation of probes with two different fluorescent signals, let's call them A and B, that correspond to the two alleles at the SNP locus. In diploids, the signals from SNP arrays or from, e.g. KASP assays have to be classified into three genotype classes (homozygotes AA and BB and heterozygote AB). Software from the companies is available to do this automatically, although it is advisable to scrutinise the outcome of this automatic process, especially for crucial markers. In highly heterogeneous crops, a considerable number of null alleles may occur, i.e. alleles that do not give one of the two possible signals as one of the probes used in the assay does not hybridise due to flanking SNPs. After recoding part of the information can still be used (Tang et al. 2015). To facilitate detection and automate rescoring, a tool has been developed (Di Guardo et al. 2015). Subsequently, the data can be used for producing genetic maps using existing software for linkage mapping including Joinmap (version 6 can handle more SNPs than previous versions) or MSDmap (Preedy and Hackett 2016, an improved new version of which is being developed that is even faster).

For SNP genotyping in polyploid crops, we also first need to classify the SNP signals in terms of allele dosage (e.g. five classes, AAAA to BBBB in a tetraploid, etc.). Software for this has been developed: FitTetra (Voorrips et al. 2011) for tetraploid crops, FitPoly (Bourke et al., in preparation) for all ploidy levels. Subsequently,

genetic maps can be made in polymapR (Bourke et al. 2018) which uses MSDmap for ordering. Within polymapR integrated maps can be produced as is exemplified for rose and chrysanthemum as tetraploid and hexaploid examples (Bourke et al. 2017; Van Geest et al. 2017a). With these integrated maps for each offspring, the IBD (identity-by-descent) probabilities of parts of the linkage groups can be estimated. For tetraploids specific IBD estimation software has been developed (Zheng et al. 2015); for other ploidy levels, a simpler approach based on only the most informative markers can be implemented.

Then we need to identify QTLs/markers/genomic regions associated with relevant traits and identify markers that are associated with the superior allele. Ultimately, we would like to have pipelines for these steps with easy-to-use interfaces to give breeders access to the data in a user-friendly manner. Advances in this direction have been made in recent years in diploid outbreeding species, where software and tool development are ahead of those for polyploids: approaches are available for pedigree reconstruction (Van de Weg et al. 2018), and software exists for QTL analyses on multiple breeding families and for phasing of SNP markers across pedigreed germplasm (FlexQTL, Bink et al. 2014) and for assignment of haplotypes (PediHaplotyper, Voorrips et al. 2016). An example of the output is shown in Fig. 9.1. Breeders' Information Management Systems are being developed in the RosBREED project (www.rosbreed.org).

9.6 How to Implement Markers

Peace (2017) has defined a scheme for translating the output of genomics research into a routine application that is part of a breeding program, which he calls 'DNA-informed breeding'. This is equivalent to the term 'marker-assisted breeding', but also comprises applications that use the sharing of neutral markers or DNA sequences, such as determining parentage or identity. In addition, it may be easier for communication as it does not require explaining what a marker is and how it works.

He recognises these five steps:

- Establishing a breeder's desire for use of DNA information
- Adapting tools to local breeding utility
- Identifying efficient application schemes
- Accessing effective services in DNA-based diagnostics (this is often outsourced)
- Gaining experience in conducting DNA-informed breeding

The DNA information may be used for a range of purposes. Without the intention of being complete, we briefly list a number of possible applications.

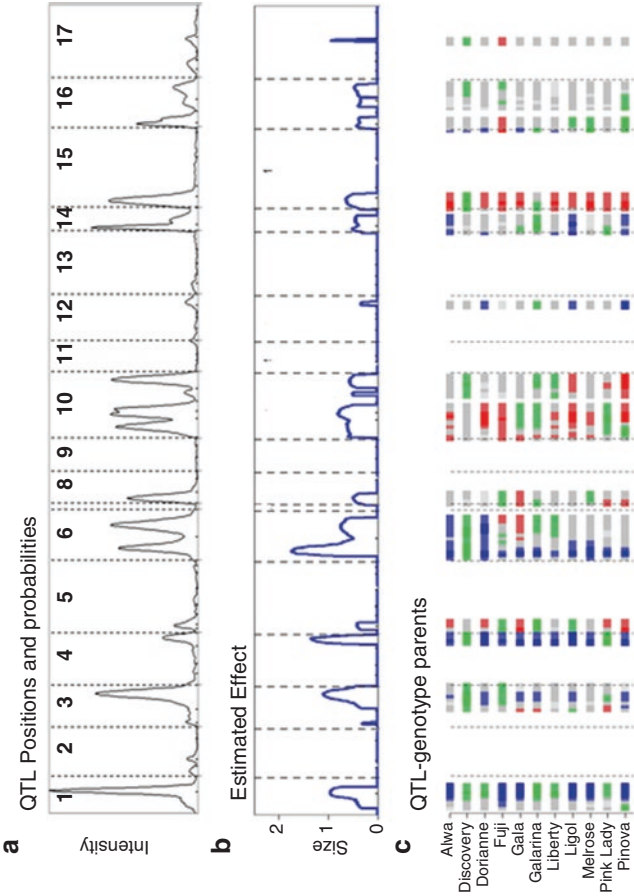


Fig. 9.1 Examples of output of the mapping software FlexQTL™ (www.flexqtl.nl) for a QTL analyses on a horticultural trait using 27 full-sib families of the outcrossing, diploid apple. (a) QTL position and probabilities along the genome (apple has 17 chromosomes). (b) Estimated effects on the scale of observation. (c) Estimated QTL genotypes for the relevant linkage groups for a small part of the parents used. Chromosomes (Linkage groups) are separated by vertical lines. Marker positions are indicated by vertical ticks on the horizontal axis. QTL genotypes are visualized by red, green, and blue horizontal bars: QQ-red, Qq-green, and qq-blue whereby QTL are assumed to be bi-allelic. Q- and q-alleles are correlated with high and low phenotypic values respectively. QTL genotype information may, among others, support the selection of well-combining parents. QTL probabilities are given by the peak area (rather than the peak height) and statistically quantified by Bayes factors (data not shown). In the statistical analyses in this study, microsatellite marker density was limiting rather than population size. Hence, increase of marker density through the use of a SNP array would considerably increase the accuracy and precision by which QTL positions can be determined, thus narrowing down the QTL intervals (i.e. the width of the peaks). Proof of concept of the multi-parent approach and the performance of the software was provided by Bink et al. (2014). Up to the end of 2017, 17 papers on scientific application have been published using this software. (Courtesy: Eric van de Weg)

9.6.1 Variety Protection

For a breeder, DNA information may be very helpful in cases of infringement. In some crops, including rose and *Phalaenopsis*, DNA markers are used during DUS testing for selecting the varieties with which the application should be compared. This may be morphologically similar varieties or, in case of mutants, the group of mutants derived from one original variety. Molecular markers provide high power for distinguishing seedling-derived varieties based on unique genotypes, while grouping mutants into groups with identical marker scores. The power of a set of 11 polymorphic microsatellite markers scored for allele presence/absence was high enough to distinguish 700 rose varieties, except the mutants (Smulders et al. 2009), and the power of 100 or more SNP markers may be expected to be similar or even better in resolution.

9.6.2 Identity Checks in the Breeding Program

A yearly check on the identity of the parents and selections is used in some field crop and vegetable breeding companies to pinpoint possible mix-ups of material and cases of mislabelling. This is cost-effective, as such mistakes are hard to avoid completely, while the effects can potentially be very costly. It may also be cost-effective to check the source material for large-scale multiplication. Such a check requires the same number of markers as for distinguishing varieties (>10 microsatellites or 100+ selected SNPs). Genotyping can also provide information on the correct parentage of offspring. For instance, when we studied a segregating population in garden roses, we discovered that part of the offspring was the result of a selfing of the mother plant (Vukosavljev et al. 2016). Also off-types (resulting from different fathers or mislabelling of seed or seedlings) can be readily detected in this way.

9.6.3 Structure of the Germplasm

DNA information may also provide insight in the structure of the germplasm used, e.g. verifying the historic pedigree records of selections and cultivars. The occurrence of signatures of breeding, i.e. genomic regions for which certain marker alleles have a high prevalence, indicates targets for selection and breeding, even though the underlying trait itself might be unknown. Finally, DNA information gives insight in the available genetic diversity. In rose, this may on the one hand be used to increase the efficiency by which new traits from diploid roses can be introgressed into tetraploid cut roses. On the other hand, DNA information makes it possible to maintain genetic diversity in the material produced. This is especially

important in long-lived, vegetative propagated species such as trees and woody shrubs, where a set of improved varieties should retain as much of the genetic diversity as possible as a prerequisite for being able to withstand global warming and other changes that may occur in the future.

9.6.4 Seedling Selection

An obvious application is to use trait-predictive markers for seedling selection. This is useful for morphologically invisible combinations of alleles. This is, for instance, the case when a cross was intended to produce homozygous resistant plants in offspring for use in further breeding, or to pyramid disease resistance genes from both parents in the offspring. Both the seedlings that inherit two resistance alleles and those that inherit only one resistance allele are resistant, so they cannot be distinguished based on phenotype. Other applications may be for traits that are very expensive or difficult to evaluate or for early selection of traits that can only be evaluated at the production stage.

In order to use seedling selection optimally, one would want to screen the seedlings as early as possible, to avoid spending space and time on offspring that will be discarded later. One short window of opportunity exists when plants are still small seedlings and have not yet been transplanted. A convenient format is to sow the seeds in 8×12 hole trays and fill a corresponding 96-well microtiter plate with leaf punches of the seedlings. These may be analysed in-house or sent to a service providing company that performs the DNA extraction and marker analysis. The results that come back from the provider can then directly be used to cull the seedlings. Only the trays need to be numbered and oriented. For this application throughput time is crucial: the results must be available before the seedlings need to be transplanted.

9.6.5 Parental Selection

Compared to seedling selection, parental selection may often be more efficient and easier to implement. In this case the information about the genetic constitution of all possible parents used in the breeding program is used to optimize the combinations of parents. This may allow a breeder to make better crosses. Parental selection may be based on genome-wide data and employ genomic selection to optimize the use of genetic information. It may also be based on information derived from the pedigree of the parents, as has been elaborated in various woody fruit species using FlexQTL.

Parental selection may be applied for a single trait as well, e.g. on a few major genes for an important QTL. Sometimes an optimal choice of parents, or the decision to exclude a certain genetic background, may obviate the need to screen the seedlings for that trait later on.

One example of the use of markers is to establish if a trait that segregates from two independent genetic sources has a common origin. This is especially relevant in polyploid crops, in which different QTL constitutions may be difficult to distinguish phenotypically and where it is not trivial to perform large phenotyping trials for traits with a lot of GxE interaction.

Another goal for which parental selection based on DNA information is often used is the acceleration of the introgression of a donor gene into a recurrent background, which is used in ornamental breeding for important single dominant genes, despite the limited possibilities for inbreeding. Examples include disease resistance genes, recurrent flowering, and flower shape mutants such as double flowering. Next to selection for the introgression segment harbouring the trait of interest, markers across the genome are used to select against the remainder of the donor genome among backcross progeny, so that the cultivated genome background is restored as soon as possible. This is also possible in outcrossing polyploid crops, but the serial backcrosses are generally not with a single plant but with plants from the same pool of cultivated plants.

9.6.6 Meiotic Behaviour and Segregation Distortion

The presence of irregularities in chromosome pairing has consequences for the segregation of traits, for the frequencies of desired alleles in the progeny of crosses, for the direction in which crosses can best be made, and for the ability to perform genetic (QTL) analyses. Meiotic behaviour can now be examined based on co-segregation patterns of SNP alleles in full-sib families. For instance, in a cross in tetraploid rose, tetrasomic chromosome pairing was the rule, but one region in one parent had disomic pairing (Bourke et al. 2017). As a result, for chromosome 1 not all theoretical combinations of alleles from the mother were present in the offspring. Van Geest et al. (2017b) established that *Chrysanthemum* behaves as a hexaploid with polysomic inheritance. The occurrence of double reduction events provides possibilities for quick selection for particular chromosome segments with a favourable trait.

9.7 Genome Sequences and Candidate Gene Approaches

Next to using markers associated with traits as a proxy, we can also try to zoom in and assess the underlying functional variation in genes for these traits. Knowledge of the existing variation in functional alleles in the genes that form biosynthetic or regulatory pathways will make it possible to consider directing these pathways towards, e.g. different colour or scent combinations.

9.7.1 *First the Genome Sequence, Then Zoom In*

More and more complete genome sequences of crops become available. The list now also includes a few ornamental species, namely, carnation (Yagi et al. 2014), *Primula veris* (Nowak et al. 2015), the orchid *Phalaenopsis equestris* (Cai et al. 2015), the two parental genomes of *Petunia hybrida* (Bombarely et al. 2016), the ornamental flower and fruit tree *Prunus mume* (Zhang et al. 2012), and rose (Hibrand Saint-Oyant et al. 2018). The huge *Chrysanthemum* genome is being sequenced and assembled in Wageningen (the Netherlands), using the latest long-read sequencing technology (Oxford Nanopore MinION) in combination with short-read sequencing for accuracy (R Finkers, personal communication). Other diploid chrysanthemum species are being sequenced in China and South Korea, while also the production of a tulip genome sequence using long-read MinION sequencing has been announced.

A genome sequence makes it possible to move from a QTL region with associated markers to the underlying variation in candidate genes in that region. An alternative to the genome of the ornamental species itself is to exploit the synteny with genomes of related species to zoom in and identify candidate genes. For instance, Schulz et al. (2016) and Nguyen et al. (2017) performed genome-wide association studies in a set of cultivated, tetraploid garden rose varieties genotyped with the rose high-density SNP array, and they identified regions associated with anthocyanin and carotenoid contents in petals and with the capacity of leaf petioles for direct shoot regeneration. The associated SNP markers were then mapped to the closely related genome of strawberry, which is largely co-linear (Gar et al. 2011; Vukosavljev et al. 2016; Bourke et al. 2017), and the regions in which these markers landed contained known candidate genes for these traits. These genes can be subsequently studied further for validation of functionality and presence of allele variants in rose and their effect on the trait.

9.7.2 *Bulked Segregant Analysis*

Another possible approach is using BSA-seq to search for markers or genes that are different between two groups of offspring from a cross, e.g. sets of resistant and susceptible offspring. These sets, along with the parents, can be analysed using whole genome sequencing. Subsequently regions in the genome can be identified that are associated with resistance, and genes present in that regions can be identified. In this way Hawkins et al. (2016) found a SNP in a MYB gene associated with yellow fruit colour in strawberry.

9.7.3 *Candidate Gene Approaches*

One may also directly test the association of SNPs in certain genes with the trait of interest. This can be done if candidate genes for the trait of interest are available, e.g. based on studies in other species, using the conservation of the coding sequence

of the genes across species, or using the fact that they belong to a gene family of which the genes are highly similar. Screening candidate genes is an effective strategy when a QTL region is so large that it contains hundreds of genes. For instance, Fu et al. (2017) mapped Botrytis resistance in *Gerbera hybrida* and detected 20 QTLs. They also compiled a list of 29 candidate genes from the transcriptome database and developed SNP markers based on homology with known resistance genes against botrytis from other species in literature. These SNPs were then mapped using gene-specific high-resolution melting point markers. Seven of these genes co-localized with the known QTLs (Fu et al. 2016). Two of these genes were tested and showed to be up-regulated upon Botrytis infection. Also, after gene silencing using VIGS, the lesion sizes upon infection were significantly smaller. This indicates that these are possible causal genes.

Kaufmann et al. (2012) isolated and characterized four MLO homologs in rose based on sequence homology to other dicot MLO genes that are involved in plant-powdery mildew interactions. Therefore, they are considered to be candidates for functional MLO genes, and alleles of these genes may confer disease resistance in rose. In pea, tomato, cucumber, apple, and several other crops, recessive mutants of single MLO genes confer full resistance to powdery mildew (e.g. Berg et al. 2017).

Studies on the pathways involved in important traits such as dormancy and flowering time have received considerable attention because with RNA-seq the candidate genes for these pathways can be retrieved easily from the ornamental crop of interest. Leeggangers et al. (2013) described possible applications of such candidate genes for flowering time (shortening of the juvenile phase) and vegetative propagation capacity in bulbous species, including tulip (Leeggangers et al. 2017) and lily (Moreno-Pachon et al. 2016).

A caveat of the candidate gene approach is that it can easily lead to false positives. Sometimes many genes are up- or down-regulated in a certain stage or upon a treatment. Therefore, validation steps are needed.

9.8 Precision Breeding Techniques

New plant breeding techniques (also called ‘precision breeding techniques’) are a range of techniques that add new tools to the breeder’s toolbox (Schaart et al. 2016; Van de Wiel et al. 2017), also applicable in ornamental breeding. Notably, directed mutagenesis (also called genome editing or gene editing) using Crispr/Cas or other programmable site-directed nucleases can expand the pool of genetic variation available for breeders. These programmable nucleases make a double-strand break at an exactly defined location in the genome. Mistakes are sometimes introduced while the cell repairs the double-strand break, and plants in which small deletions or insertions (indels) occur at this location in the genome can be selected. From a biological point of view, it is a form of targeted mutagenesis. Gene editing has taken off at a very quick pace, partly because it is technically easy, but also because next-generation sequencing makes it feasible to generate the required sequence information quickly. The examples in the recent literature are mostly about knockout of

genes (e.g. in a colour biosynthetic gene), but it is also possible to mutate elements in promoters of genes, thus changing that spatial and temporal pattern of expression of the gene. This will be very interesting to explore, as many morphological differences are caused by longer or shorter expression of genes. Rodríguez-Leal et al. (2017) performed a systematic screening of combinations of promoter element mutations in three transcription factor genes that are key for the architecture of tomato plants. Their study produced new genetic variation, including plants with novel morphologies.

Gene editing is a potentially non-transgenic approach to directly modify or create, for instance, a disease resistance through a novel mutation in an MLO gene (Debener and Byrne 2014), and the method can produce mutations in all copies of the gene in a single cell, which is required for such a recessive disease resistance. It is therefore particularly suitable for inducing mutations in polyploids. There are many examples of (recessive) deletions that disrupt genes and result in novel functionality, as known from crop domestication and classical mutation breeding (both using radiation and chemical mutagens; Van de Wiel et al. 2017).

Three challenges exist when one considers deploying Crispr/cas: knowledge of the causal gene underlying the trait of interest; availability of a genome sequence or a complete set of transcriptome sequences, so that the guide RNAs needed for targeting can be designed specifically for that particular gene; and, last but not least, effective transformation and especially regeneration protocols, which are needed to regenerate a plant from the cell in which the mutation has been produced. Regeneration protocols are the biggest bottleneck, as many species are amenable to mutations caused by transient expression of the Cas9 protein with a guide RNA present, e.g. in protoplasts or immature embryos, but protocols to regenerate a plant from protoplasts are rare, certainly for ornamentals. It can therefore be expected that gene editing will reinvigorate cell biology and tissue culture studies.

9.9 Conclusion

These are exciting times. The technologies described in this chapter will contribute to new directions for the future of ornamental breeding. Notably, sequencing and genotyping methods are now in place to develop tools to make accelerate breeding, also in crops in which little information is present, which are heterozygous or polyploid, or have a long generation time.

For ornamental breeding companies, the introduction of DNA-informed breeding will have to start by thinking about traits in which (i) there is urgency to solve a problem (e.g. demand from growers for disease-resistant varieties) but little progress can be made with normal crossing and selection schemes, (ii) large cost reduction may be achieved in the selection process (e.g. early selection for a trait that is now only assessable at the end of the breeding cycle), or (iii) elite cultivars can be improved further without having to go through the disruptive phase of crossing.

The next step can be best to contact a research group (university, research institute, or in some cases a specialised breeding research company) as in most cases prior to implementing DNA-informed breeding, a discovery project has to be started to find markers linked to trait(s). In the initial steps, attention must also be paid to plant material (populations) needed for this discovery work, as this often requires higher specification with regard to being true to parents, to population size, and to segregation of (multiple) traits. We highly recommend companies that want to start with DNA-informed breeding to invest in a person that can understand and implement steps within the company and explain and discuss these steps directly with the breeders of the focal crops already in the initial phase.

Acknowledgements This research was partially supported by KB-24-002-017 and the TKI-U Polyploid projects BO-26.03-002-001 and BO-26.03-009-004. The support of the companies participating in the Polyploid projects is gratefully acknowledged.

References

- Arens P, Bijman P, Tang N, Shahin A, van Tuyl JM (2012) Mapping of disease resistance in ornamentals: a long haul. *Acta Hort* (ISHS) 953:231–238. <https://doi.org/10.17660/ActaHortic.2012.953.32>
- Berg JA, Appiano M, Bijsterbosch G et al (2017) Functional characterization of cucumber (*Cucumis sativus* L.) Clade V MLO genes. *BMC Plant Biol* 17:80. <https://doi.org/10.1186/s12870-017-1029-z>
- Bink MCAM, Jansen J, Madduri M et al (2014) Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theor Appl Genet* 127:1073–1090. <https://doi.org/10.1007/s00122-014-2281-3>
- Bombarely A, Moser M, Amrad A et al (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat Plants* 2:16074. <https://doi.org/10.1038/nplants.2016.74>
- Bourke PM, Arens P, Voorrips RE et al (2017) Partial preferential chromosome pairing is genotype dependent in tetraploid rose. *Plant J* 90:330–343. <https://doi.org/10.1111/tpj.13496>
- Bourke P, van Geest G, Voorrips RE et al (2018) polymapR: linkage analysis and genetic map construction from F1 populations of outcrossing polyploids. *Bioinformatics*, bty371. <https://doi.org/10.1093/bioinformatics/bty371>
- Cai J, Liu X, Vanneste K et al (2015) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat Genet* 47:65–72. <https://doi.org/10.1038/ng.3149>
- Cao Z, Deng Z (2017) De novo assembly, annotation, and characterization of root transcriptomes of three *Caladium* cultivars with a focus on necrotrophic pathogen resistance/defense-related genes. *Int J Mol Sci* 18:712. <https://doi.org/10.3390/ijms18040712>
- Casimiro-Soriguer I, Narbona E, Buide ML, del Valle JC, Whittall JB (2016) Transcriptome and biochemical analysis of a flower color polymorphism in *Silene littorea* (Caryophyllaceae). *Front Plant Sci* 7:204. <https://doi.org/10.3389/fpls.2016.00204>
- Chong X, Zhang F, Wu Y et al (2016) A SNP-enabled assessment of genetic diversity, evolutionary relationships and the identification of candidate genes in *Chrysanthemum*. *Genome Biol Evol* 8:3661–3671. <https://doi.org/10.1093/gbe/evw270>
- Debener T, Byrne DH (2014) Disease resistance breeding in rose: current status and potential of biotechnological tools. *Plant Sci* 228:107–117. <https://doi.org/10.1016/j.plantsci.2014.04.005>
- Di Guardo M, Micheletti D, Bianco L et al (2015) ASSIsT: an automatic SNP scoring tool for in- and outbreeding species. *Bioinformatics* 31:3873–3874. <https://doi.org/10.1093/bioinformatics/btv446>

- Fu Y, Esselink GD, Visser RGF, van Tuyl JM, Arens P (2016) Transcriptome analysis of *Gerbera hybrida* including in silico confirmation of defense genes found. *Front Plant Sci* 7:247. <https://doi.org/10.3389/fpls.2016.00247>
- Fu Y, van Silfhout A, Shahin A et al (2017) Genetic mapping and QTL analysis of Botrytis resistance in *Gerbera hybrida*. *Mol Breed* 37:13. <https://doi.org/10.1007/s11032-016-0617-1>
- Gar O, Sargent DJ, Tsai C-J et al (2011) An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. *PLoS One* 6:e20463. <https://doi.org/10.1371/journal.pone.0020463>
- Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat Genet* 17:333–351. <https://doi.org/10.1038/nrg.2016.49>
- Hawkins C, Caruana J, Schiksnis E, Liua Z (2016) Genome-scale DNA variant analysis and functional validation of a SNP underlying yellow fruit color in wild strawberry. *Sci Rep* 6:29017. <https://doi.org/10.1038/srep29017>
- Hibrand Saint-Oyant L, Ruttink T, Hamama L et al (2018) A high-quality sequence of *Rosa chinensis* to elucidate genome structure and ornamental traits. In: bioRxiv. <https://doi.org/10.1101/254102>
- Holdsworth WL, Mazourek M (2015) Development of user-friendly markers for the pvr1 and Bs3 disease resistance genes in pepper. *Mol Breed* 35:28
- Huang J, Lin C, Cheng T et al (2016) The genome and transcriptome of *Phalaenopsis* yield insights into floral organ development and flowering regulation. *PeerJ* 4:e2017. <https://doi.org/10.7717/peerj.2017>
- Jung H-J, Veerappan K, Natarajan S et al (2017) A system for distinguishing octoploid strawberry cultivars using high-throughput SNP genotyping. *Trop. Plant Biol* 10:68–76. <https://doi.org/10.1007/s12042-017-9185-8>
- Kaufmann H, Qiu X, Wehmeyer J, Debener T (2012) Isolation, molecular characterization, and mapping of four rose MLO orthologs. *Front Plant Sci* 3:244. <https://doi.org/10.3389/fpls.2012.00244>
- Kim JE, Oh SK, Lee JH, Lee BM, Jo SH (2014) Genome-wide SNP calling using next generation sequencing data in tomato. *Mol Cells* 37:36–42. <https://doi.org/10.14348/molcells.2014.2241>
- Koning-Boucoiran CFS, Smulders MJM, Krens FA, Esselink GD, Maliepaard C (2012) SNP genotyping in tetraploid roses. *Acta Hort* (ISHS) 953:351–356. <https://doi.org/10.17660/ActaHortic.2012.953.49>
- Koning-Boucoiran CFS, Esselink GD, Vukosavljev M et al (2015) Using RNA-Seq to assemble a rose transcriptome with more than 13,000 full-length expressed genes and to develop the WagRhSNP 68k Axiom SNP array for rose (*Rosa* L.). *Front Plant Sci* 6:249. <https://doi.org/10.3389/fpls.2015.00249>
- Leeggangers HACF, Moreno-Pachon N, Gude H, Immink RGH (2013) Transfer of knowledge about flowering and vegetative propagation from model species to bulbous plants. *Int J Dev Biol* 57:611–620. <https://doi.org/10.1387/ijdb.130238ri>
- Leeggangers HACF, Nijveen H, Nadal Bigas J, Hilhorst HWM, Immink RGH (2017) Molecular regulation of temperature-dependent floral induction in *Tulipa gesneriana*. *Plant Physiol* 173:1904–1919. <https://doi.org/10.1104/pp.16.01758>
- Moreno-Pachon NM, Leeggangers HACF, Nijveen H et al (2016) Elucidating and mining the *Tulipa* and *Lilium* transcriptomes. *Plant Mol Biol* 92:249–261. <https://doi.org/10.1007/s11103-016-0508-1>
- Muir P, Li S, Lou S et al (2016) The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol* 17:53. <https://doi.org/10.1186/s13059-016-0917-0>
- Nguyen THN, Schulz D, Winkelmann T, Debener T (2017) Genetic dissection of adventitious shoot regeneration in roses by employing genome-wide association studies. *Plant Cell Rep* 36:1493–1505. <https://doi.org/10.1007/s00299-017-2170-8>
- Nijveen H, van Kaauwen M, Esselink DG, Hoegen B, Vosman B (2013) QualitySNPng: a user-friendly SNP detection and visualization tool. *Nucleic Acids Res* 41:W587–W590. <https://doi.org/10.1093/nar/gkt333>

- Nowak MD, Russo G, Schlapbach R et al (2015) The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biol* 16:12. <https://doi.org/10.1186/s13059-014-0567-z>
- Peace CP (2017) DNA-informed breeding of rosaceous crops: promises, progress and prospects. *Hortic Res* 4:17006. <https://doi.org/10.1038/hortres.2017.6>
- Preedy KF, Hackett CA (2016) A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theor Appl Genet* 129:2117–2132. <https://doi.org/10.1007/s00122-016-2761-8>
- Rodríguez-Leal D, Lemmon ZH, Man J, Bartlett ME, Lippman ZB (2017) Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171:470–480.e8. <https://doi.org/10.1016/j.cell.2017.08.030>
- Schaart JG, van de Wiel CCM, Lotz LAP, Smulders MJM (2016) Opportunities for products of new plant breeding techniques. *Trends Plant Sci* 21:438–449. <https://doi.org/10.1016/j.tplants.2015.11.006>
- Schmutz J, Wheeler J, Grimwood J et al (2004) Quality assessment of the human genome sequence. *Nature* 429:365–368
- Schulz DF, Schott RT, Voorrips RE, Smulders MJM, Linde M, Debener T (2016) Genome-wide association analysis of the anthocyanin and carotenoid contents of rose petals. *Front Plant Sci* 7:1798. <https://doi.org/10.3389/fpls.2016.01798>
- Shahin A, Arens P, Van Heusden AW et al (2010) Genetic mapping in *Lilium*: mapping of major genes and quantitative trait loci for several ornamental traits and disease resistances. *Plant Breed* 130:372–382. <https://doi.org/10.1111/j.1439-0523.2010.01812.x>
- Shahin A, van Gurp T, Peters SA, Visser RGF, van Tuyl JM, Arens P (2012a) SNP markers retrieval for a non-model species: a practical approach. *BMC Res Notes* 5:79. <https://doi.org/10.1186/1756-0500-5-79>
- Shahin A, van Kaauwen M, Esselink D et al (2012b) Generation and analysis of expressed sequence tags in the extreme large genomes *Lilium* and *Tulipa*. *BMC Genomics* 13:640. <https://doi.org/10.1186/1471-2164-13-640>
- Smulders MJM, Esselink D, Voorrips RE, Vosman B (2009) Analysis of a database of DNA profiles of 734 hybrid tea rose varieties. *Acta Hortic* 836:169–174. http://www.actahort.org/books/836/836_24.htm
- Smulders MJM, Vukosavljev M, Shahin A, van de Weg WE, Arens P (2012) High throughput marker development and application in horticultural crops. *Acta Hortic (ISHS)* 961:547–551. <https://doi.org/10.17660/ActaHortic.2012.961.72>
- Smulders MJM, Voorrips RE, Esselink GD et al (2015) Development of the WagRhSNP Axiom SNP array based on sequences from tetraploid cut roses and garden roses. *Acta Hortic* 1064:177–184. <https://doi.org/10.17660/ActaHortic.2015.1064.20>
- Tang N, van der Lee T, Shahin A et al (2015) Genetic mapping of resistance to *Fusarium oxysporum* f. sp. *tulipae* in tulip. *Mol Breed* 35:122. <https://doi.org/10.1007/s11032-015-0316-3>
- Van de Weg E, Di Guardo M, Jänsch M et al (2018) Epistatic fire blight resistance QTL alleles in the apple cultivar ‘Enterprise’ and selection X-6398 discovered and characterized through pedigree-informed analysis. *Mol Breed* 38:5. <https://doi.org/10.1007/s11032-017-0755-0>
- Van de Wiel CCM, Schaart JG, Lotz LAP, Smulders MJM (2017) New traits in crops produced by genome editing techniques based on deletions. *Plant Biotechnol Rep* 11:1–8. <https://doi.org/10.1007/s11816-017-0425-z>
- Van Geest G, Bourke PM, Voorrips RE et al (2017a) An ultra-dense integrated linkage map for hexaploid chrysanthemum enables multi-allelic QTL analysis. *Theor Appl Genet* 130:2527–2541. <https://doi.org/10.1007/s00122-017-2974-5>
- Van Geest G, Voorrips RE, Esselink D, Post A, Visser RGF, Arens P (2017b) Conclusive evidence for hexasomic inheritance in chrysanthemum based on analysis of a 183k SNP array. *BMC Genomics* 18:585. <https://doi.org/10.1186/s12864-017-4003-0>
- Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinformatics* 12:172

- Voorrips RE, Bink MCAM, Kruisselbrink JW et al (2016) PediHaplotyper: software for consistent assignment of marker haplotypes in pedigrees. *Mol Breed* 36:119. <https://doi.org/10.1007/s11032-016-0539-y>
- Vukosavljev M, Arens P, Voorrips RE et al (2016) High-density SNP-based genetic maps for the parents of an outcrossed and a selfed tetraploid garden rose cross, inferred from admixed progeny using the 68k rose SNP array. *Hortic Res* 3:16052. <https://doi.org/10.1038/hortres.2016.52>
- Wei Z, Sun Z, Cui B, Zhang Q, Xiong M, Wang X, Zhou D (2016) Transcriptome analysis of colored calla lily (*Zantedeschia rehmannii* Engl.) by Illumina sequencing: de novo assembly, annotation and EST-SSR marker development. *PeerJ* 4:e2378. <https://doi.org/10.7717/peerj.2378>
- Yagi M, Kosugi S, Hirakawa H et al (2014) Sequence analysis of the genome of carnation (*Dianthus caryophyllus* L.). *DNA Res* 21:231–241. <https://doi.org/10.1093/dnares/dst053>
- Zhang Q, Chen W, Sun L et al (2012) The genome of *Prunus mume*. *Nat Commun* 3:1318. <https://doi.org/10.1038/ncomms2290>
- Zhang J, Zhao K, Hou D et al (2017) Genome-wide discovery of DNA polymorphisms in Mei (*Prunus mume* Sieb. et zucc.), an ornamental woody plant, with contrasting tree architecture and their functional relevance for weeping trait. *Plant Mol Biol Report* 35:37–46. <https://doi.org/10.1007/s11105-016-1000-4>
- Zheng C, Boer MP, Van Eeuwijk FA (2015) Reconstruction of genome ancestry blocks in multiparental populations. *Genetics* 200:1073–1087. <https://doi.org/10.1534/genetics.115.177873>