Some thoughts on how to use markers in tetraploid rose breeding

M.J.M. Smulders[a], W.E. van de Weg, P.M. Bourke, R.E. Voorrips, P. Arens and C. Maliepaard

Plant Breeding, Wageningen University & Research, Wageningen, The Netherlands.

Abstract

Thanks to recent advances in high-throughput genotyping tools and software for genetic analyses, it is now possible to perform QTL mapping and genome-wide association analysis in tetraploid roses and to establish the effect of favorable SNP marker haplotype(s) ("plus-alleles"). The challenge is how to use these tools and technologies for rose breeding in a cost-effective way. At least four points need to be addressed: (1) which steps of the breeding process may benefit from marker information; (2) what is the value of the identified plus-alleles for the trait; (3) where in the breeding germplasm do these plus-alleles occur, and (4) how can plants carrying these plus-alleles be selected efficiently. Here we discuss for which purposes markers may be applied.

Keywords: *Rosa hybrida*, rose, SNP, genotyping, DNA-informed breeding

INTRODUCTION

In rose the development of genetic and molecular tools for breeding has been slow in the past. Most often, marker development, construction of linkage maps and QTL analyses were carried out in diploid roses, e.g., Spiller et al. (2010) on rose scent volatiles. However, breeding takes place mainly in tetraploid roses. Translating obtained knowledge to the tetraploid level was limited. This has changed recently due to three developments: (i) the use of next generation sequencing to generate large numbers of single nucleotide polymorphism (SNP) markers based on genomic or transcriptomic sequences (e.g., Smulders et al., 2012; Smulders and Arens, 2018); (ii) the development of the WagRhSNP array for rose (Koning-Boucoiran et al., 2015) which enables the detection of tens of thousands of SNP markers, and (iii) software for dosage scoring, which allows efficient assignment of the tetraploid SNP genotype of individuals (Voorrips et al., 2011; Bourke et al., 2018a,b). Using these tools and tools for linkage mapping, ultradense genetic maps have now been produced in tetraploid rose (Vukosavljev et al., 2016; Bourke et al., 2017) that contain at least as many markers as the ones from diploid rose crosses (e.g., Spiller et al., 2011). The maps may be either maps for the separate homologous chromosomes or integrated maps. These maps will help performing QTL mapping (Bourke et al., in prep.; Smulders, pers. commun.) and genome-wide association analysis (Schulz et al., 2016; Nguyen et al., 2017) in tetraploid rose, to identify regions in the genome that are statistically correlated with the trait of interest.

Once such genomic regions have been identified, the finding is often considered to be ready to be used in breeding. In fact, many presentations at conferences and many scientific papers end at this point with a statement that the results will be useful for breeding. This is accurate for single gene-based, dominant traits for which the desired allele is inherited from a single source, such as a disease resistance gene from a wild relative species. Markers associated with such a gene can be used in a breeding program, and below we will discuss how they can be employed.

However, it is not the end of research for traits for which the phenotypic variation is affected by a gene that has multiple alleles with distinctive effects. Whereas this may occur also in inbreeding diploid crops, it is far more common in outbreeding polyploid crops.

[a]E-mail: rene.smulders@wur.nl

Unless one plus-allele is very common in the breeding germplasm, it will be necessary to evaluate the effect of multiple common alleles for the QTL region, evaluate possible interactions between alleles, and determine dosage effects (how many copies of the plus-allele are necessary, is full homozygosity of the favorable allele at the tetraploid level required?). For such studies, alleles can be approximated through the haplotypes they reside in, i.e., each allele is characterized by a unique series of SNPs, so developing methods and tools for haplotyping are an important goal in polyploid genetics as well.

DNA-informed breeding

How to employ the existing knowledge for steps in the breeding of rose in an efficient and cost-effective way? At least four points need to be addressed: which steps of the breeding process may benefit from marker information, what is the value of the identified plus-alleles for the trait, where in the breeding germplasm do these plus-alleles occur, and how can plants carrying the alleles be selected efficiently?

Cameron Peace (2017) has defined a scheme of five steps for translating the output of genomics research into a routine application that is integrated in a breeding program, a situation which he coined 'DNA-informed breeding'. This term is equivalent to the term marker-assisted breeding, with two differences. First, it does not require that the reader understands what a marker is and how it works, which is an advantage during communication with breeders, other possible users, and the general public. Second, it also includes exploitation of neutral markers, as distinct from markers that are associated with traits, for applications such as determining parentage or checking identity.

Peace (2017) recognizes the following five steps:

1. Establishing a breeder's need (or advantage) for use of DNA information for important traits;
2. Adapting tools to the local breeding situation;
3. Identifying efficient application schemes;
4. Accessing effective services in DNA-based diagnostics (this step is often outsourced, balancing cost-effectiveness with throughput capacity and time needed to obtain the results);
5. Gaining experience in conducting DNA-informed breeding.

VARIOUS GOALS

Cultivar protection

DNA sequences are not used directly for defining or assessing DUS (distinctiveness, uniformity, stability) criteria, but they may be used for selecting the cultivars with which the application should be compared, which can be morphologically similar cultivars or, in case of mutants, the group of mutants derived from one original cultivar. This is done at the start of DUS testing in, e.g., rose and *Phalaenopsis*. For a breeder DNA information may be very helpful in cases of infringement.

Molecular markers provide high power for identifying and recognizing seedling-derived cultivars based on unique genotypes, while grouping rose mutants into groups with identical marker scores. The discerning power of a set of 11 polymorphic microsatellite markers scored for allele presence/absence was sufficient to distinguish 700 cultivars, except those that were mutants (Smulders et al., 2009). A set of thousands of polymorphic SNPs with dosage information will have a much higher power, and nowadays it may be cost-effective to use a SNP array for genotyping them. To reduce the costs of a SNP array sometimes dedicated arrays are designed that contain over ten thousand SNPs for each of multiple species, so that they can be produced in large volume for lower costs. When a large number of plants (hundreds to thousands) need to be screened with a relatively small set (tens to hundreds) of selected SNPs, then alternative techniques such as KASP (e.g., Koning-Boucoiran et al., 2012; Holdsworth and Mazourek, 2015) or Fluidigm (Jung et al., 2017) are more cost-effective.

Identity checks in the breeding program

A relatively common application of markers is identity checking. For instance, a yearly check on the identity of the parents and selections is used in some field crop and vegetable breeding companies to identify possible mix-ups of material and cases of mislabeling. This is cost-effective, as such mistakes are hard to avoid completely, and when detected late in the breeding process, or even after a cultivar has been released, the effects can be long-lasting and potentially very costly. It may also be cost-effective to check the source material before large-scale multiplication.

Genotyping the breeding material or parts thereof can also provide information on the true parentage of offspring. For instance, when we studied a segregating population in garden roses, we discovered that what we assumed was offspring of one cross, actually consisted of offspring of a cross with a pollen donor and offspring that was the result of a selfing of the mother plant (Vukosavljev et al., 2016).

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 than without this information. Parental selection may be based on genome-wide data, and employ genomic selection to optimize the use of genetic information. However, it can be applied for a single trait as well, e.g., for 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 a certain 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, i.e., are the positive effects due to the same allele at the same locus, or might different functional alleles or even different loci be involved? 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 that are also influenced by environmental variation. If it turns out that the alleles are from different genes that are genetically close to each other, one could start searching for a rare recombinant, which would become the donor of a unique combination of alleles/traits.

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. Next to selection for the introgression segment, markers in the rest of the genome are used to select against the introgressed genome among backcross progeny to recover the recurrent genome as soon as possible. This is also possible in an outcrossing polyploid crop like rose, both for reducing linkage drag around the locus and for screening and eliminating complete chromosomes from the introgressed parent.

Seedling selection

Trait-predictive markers for seedling selection are useful for invisible combinations of alleles, i.e., when the phenotype of the seedlings does not provide clues to make a selection. This is for instance the case when a cross is intended to pyramid disease resistance genes from both parents in the offspring: both seedlings that inherit two genes and those that inherit only one of the genes, have a resistant phenotype, so they cannot be distinguished based on phenotype. Other applications may be for traits that are very expensive or difficult to evaluate otherwise (e.g., resistance against quarantine organisms), 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 money on offspring that are discarded later. In practice, a 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 small trays, and fill a corresponding 96-well microtiter plate with leaf punches. These may be analysed in-house

or sent to a service providing company that carries out 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, no numbering of the individual plants is required. For this application, the throughput time is important: the results must be available before the seedlings need to be transplanted.

Structure of the germplasm

The DNA information generated may also provide insight in the structure of the germplasm used, identifying, e.g., the correct historic pedigree records of selections and cultivars. Signatures of breeding consist of genomic regions for which a certain marker has such a high prevalence that it would seem likely that it has been a target for breeding even though the underlying trait itself might be unknown. These regions can subsequently be introgressed into other types of roses.

DNA information makes it also possible to maintain genetic diversity in the material produced. This is especially important in long-lived, vegetative propagated species such as trees and shrubs including garden roses, where a set of improved cultivars 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 near future.

Genetic diversity is typically stored in genebanks, but for vegetatively propagated plants it is done in the form of living collections of old cultivars (mostly tetraploid) and accessions of crop-wild relative species (tetraploid and diploid). Luckily, roses are maintained in many collections worldwide, but their financing is not always secured, which is a potential threat for the future. The development of core collections could be a way to prevent loss of germplasm whenever rose gardens are closed. They maximise the diversity to be preserved in a relatively small set of accessions, and these are replicated over multiple sites (rose gardens and botanical gardens).

Meiotic behavior 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 with existing (mapping) software. This can now be examined based on co-segregation patterns of marker alleles in full-sib families. For instance, in a cross in tetraploid rose, tetrasomic chromosome pairing was the rule, but one region of chromosome 1 in one parent had disomic pairing (Bourke et al., 2017). As a result, for chromosome 1 not all possible combinations of alleles from the mother will be present in the offspring.

DISCUSSION

Rose cultivars are mainly tetraploid and outbreeding. This creates challenges for the implementation of DNA-informed breeding. We have described some opportunities for the use of DNA markers, and have touched upon a few practical issues in their implementation. For traits that are based on a single dominant gene, such as qualitative disease resistances, and those for which one or a few major QTLs are found, DNA-informed breeding in rose is feasible.

For many genes of interest multiple alleles exist in the rose germplasm, and more than one of these alleles might be present simultaneously in a single individual. QTL mapping approaches that consider the effect of all parental alleles simultaneously (haplotype-informed analyses) offer greater sensitivity to detect these alleles, ultimately helping to untangle the source of both plus and minus alleles from a bi-parental cross. Visualisation can help breeders quickly identify the source of plus-alleles, and in which combinations they give the greatest effect. What needs to be developed to be able to achieve this includes statistical procedures/approaches, software for data analyses, and pipelines with easy to use interfaces that give breeders access to the data in a user-friendly manner. Inspiration might come from advances in diploid outbreeding species, where software and tool development are ahead to polyploids: new approaches for pedigree reconstruction (Van de Weg et al.,

2018), software for QTL analyses on multiple breeding families and for phasing of SNP markers across pedigreed germplasm (FlexQTL; Bink et al., 2014), aggregation of sets of tightly linked single SNP markers into haploblocks (Visual FlexQTL), assignment of haplotypes to such haploblocks (PediHaplotyper; Voorrips et al., 2016), and Breeders Information Management Systems as are being developed in the RosBREED project (www.rosbreed.org; Jung et al., 2017).

One example of such an approach is the study of Verma et al. (2017) in strawberry (which is allooctaploid, but behaves genetically as a diploid) in which they used a SNP array to do fine-resolution QTL mapping in breeding populations for soluble solids content, fruit weight and other traits using a pedigree-based approach. Using marker haplotype analysis, they could establish the functional allele for seasonal flowering and its dosage effect (homozygous recessive), and produce SNPs markers associated with this allele to facilitate marker-assisted selection for this trait.

A rose genome sequence would make it easier to move from a marker associated to some trait of interest to the underlying gene affecting the trait. In practical breeding, it is not necessary to know the underlying genes, as a marker is sufficient for marker-assisted breeding. However, once insight in the genes and mechanism behind a trait is gained, it may enable better or novel breeding strategies (Bourke et al., 2017), including for instance pathway engineering of scent or color components. Schulz et al. (2016) used the strawberry genome, which is colinear with the rose genome (Bourke et al., 2017) as a proxy to identify candidate genes for anthocyanin and carotenoid content. Transcriptome data are already available for various stages of development (Dubois et al., 2012; Koning-Boucoiran et al., 2015). The first draft rose sequence has become available recently (Nakamura et al., 2018) and high-quality genomes are expected in 2018.

ACKNOWLEDGEMENTS

This research was partially supported by 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.

Literature cited

Bink, M.C.A.M., Jansen, J., Madduri, M., Voorrips, R.E., Durel, C.E., Kouassi, A.B., Laurens, F., Mathis, F., Gessler, C., Gobbin, D., et al. (2014). Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theor. Appl. Genet. *127* (*5*), 1073–1090 https://doi.org/10.1007/s00122-014-2281-3. PubMed

Bourke, P.M., Arens, P., Voorrips, R.E., Esselink, G.D., Koning-Boucoiran, C.F., Van't Westende, W.P., Santos Leonardo, T., Wissink, P., Zheng, C., van Geest, G., et al. (2017). Partial preferential chromosome pairing is genotype dependent in tetraploid rose. Plant J. *90* (*2*), 330–343 https://doi.org/10.1111/tpj.13496. PubMed

Bourke, P.M., Voorrips, R.E., Visser, R.G.F., and Maliepaard, C. (2018a). Tools for genetic studies in experimental populations of polyploids. Front Plant Sci *9*, 513 https://doi.org/10.3389/fpls.2018.00513. PubMed

Bourke, P.M., van Geest, G., Voorrips, R.E., Jansen, J., Kranenburg, T., Shahin, A., Visser, R.G.F., Arens, P., Smulders, M.J.M., and Maliepaard, C. (2018b). polymapR-linkage analysis and genetic map construction from F_1 populations of outcrossing polyploids. Bioinformatics *34* (*20*), 3496–3502 https://doi.org/10.1093/bioinformatics/bty371. PubMed

Dubois, A., Carrere, S., Raymond, O., Pouvreau, B., Cottret, L., Roccia, A., Onesto, J.P., Sakr, S., Atanassova, R., Baudino, S., et al. (2012). Transcriptome database resource and gene expression atlas for the rose. BMC Genomics *13* (*1*), 638 https://doi.org/10.1186/1471-2164-13-638. PubMed

Holdsworth, W.L., and Mazourek, M. (2015). Development of user-friendly markers for the *pvr1* and *Bs3* disease resistance genes in pepper. Mol. Breed. *35* (*1*), 28 https://doi.org/10.1007/s11032-015-0260-2.

Jung, H.-J., Veerappan, K., Natarajan, S., Jeong, N., Hwang, I., Nagano, S., Shirasawa, K., Isobe, S., and Nou, I.-S. (2017). A system for distinguishing octoploid strawberry cultivars using high-throughput SNP genotyping. Trop. Plant Biol. *10* (*2-3*), 68–76 https://doi.org/10.1007/s12042-017-9185-8.

Koning-Boucoiran, C.F.S., Smulders, M.J.M., Krens, F.A., Esselink, G.D., and Maliepaard, C. (2012). SNP genotyping in tetraploid roses. Acta Hortic. *953*, 351–356 https://doi.org/10.17660/ActaHortic.2012.953.49.

Koning-Boucoiran, C.F.S., Esselink, G.D., Vukosavljev, M., van 't Westende, W.P., Gitonga, V.W., Krens, F.A., Voorrips,

R.E., van de Weg, W.E., Schulz, D., Debener, T., 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. PubMed

Nakamura, N., Hirakawa, H., Sato, S., Otagaki, S., Matsumoto, S., Tabata, S., and Tanaka, Y. (2018). Genome structure of *Rosa multiflora*, a wild ancestor of cultivated roses. DNA Res. *25* (*2*), 113–121 https://doi.org/10.1093/dnares/dsx042. PubMed

Nguyen, T.H.N., Schulz, D., Winkelmann, T., and Debener, T. (2017). Genetic dissection of adventitious shoot regeneration in roses by employing genome-wide association studies. Plant Cell Rep. *36* (*9*), 1493–1505 https://doi.org/10.1007/s00299-017-2170-8. PubMed

Peace, C.P. (2017). DNA-informed breeding of rosaceous crops: promises, progress and prospects. Hortic Res *4*, 17006 https://doi.org/10.1038/hortres.2017.6. PubMed

Schulz, D.F., Schott, R.T., Voorrips, R.E., Smulders, M.J.M., Linde, M., and 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. PubMed

Smulders, M.J.M., and Arens, P. (2018). New developments in molecular techniques for breeding in ornamentals. In Ornamental Crops, Handbook of Plant Breeding 11, J. Van Huylenbroeck, ed. (Springer International Publishing AG), doi:10.1007/978-3-319-90698-0_9.

Smulders, M.J.M., Esselink, D., Voorrips, R.E., and Vosman, B. (2009). Analysis of a database of DNA profiles of 734 Hybrid Tea Rose varieties. Acta Hortic. *836*, 169–175 https://doi.org/10.17660/ActaHortic.2009.836.24.

Smulders, M.J.M., Vukosavljev, M., Shahin, A., van de Weg, W.E., and Arens, P. (2012). High throughput marker development and application in horticultural crops. Acta Hortic. *961*, 547–551 https://doi.org/10.17660/ActaHortic.2012.961.72.

Spiller, M., Berger, R.G., and Debener, T. (2010). Genetic dissection of scent metabolic profiles in diploid rose populations. Theor. Appl. Genet. *120* (*7*), 1461–1471 https://doi.org/10.1007/s00122-010-1268-y. PubMed

Spiller, M., Linde, M., Hibrand-Saint Oyant, L., Tsai, C.J., Byrne, D.H., Smulders, M.J., Foucher, F., and Debener, T. (2011). Towards a unified genetic map for diploid roses. Theor. Appl. Genet. *122* (*3*), 489–500 https://doi.org/10.1007/s00122-010-1463-x. PubMed

van de Weg, E., Di Guardo, M., Jänsch, M., Socquet-Juglard, D., Costa, F., Baumgartner, I., Broggini, G.A.L., Kellerhals, M., Troggio, M., Laurens, F., 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* (*1*), 5 https://doi.org/10.1007/s11032-017-0755-0.

Verma, S., Zurn, J.D., Salinas, N., Mathey, M.M., Denoyes, B., Hancock, J.F., Finn, C.E., Bassil, N.V., and Whitaker, V.M. (2017). Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (*Fragaria×ananassa*) breeding populations using pedigree-based QTL analysis. Hortic Res *4*, 17062 https://doi.org/10.1038/hortres.2017.62. PubMed

Voorrips, R.E., Gort, G., and Vosman, B. (2011). Genotype calling in tetraploid species from bi-allelic marker data using mixture models. BMC Bioinformatics *12* (*1*), 172 https://doi.org/10.1186/1471-2105-12-172. PubMed

Voorrips, R.E., Bink, M.C.A.M., Kruisselbrink, J.W., Koehorst-van Putten, H.J.J., and van de Weg, W.E. (2016). PediHaplotyper: software for consistent assignment of marker haplotypes in pedigrees. Mol. Breed. *36* (*8*), 119 https://doi.org/10.1007/s11032-016-0539-y. PubMed

Vukosavljev, M., Arens, P., Voorrips, R.E., van 't Westende, W.P., Esselink, G.D., Bourke, P.M., Cox, P., van de Weg, W.E., Visser, R.G., Maliepaard, C., and Smulders, M.J. (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* (*1*), 16052 https://doi.org/10.1038/hortres.2016.52. PubMed