



WAGENINGEN
UNIVERSITY & RESEARCH

Development of the WagRhSNP AXIOM SNP Array Based on Sequences from Tetraploid Cut Roses and Garden Roses

Acta Horticulturae

Smulders, M.J.M.; Voorrips, R.E.; Esselink, G.; Santos Leonardo, T.M.; Westende, W.P.C. et al

<https://doi.org/10.17660/ActaHortic.2015.1064.20>

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

Development of the WagRhSNP AXIOM SNP Array Based on Sequences from Tetraploid Cut Roses and Garden Roses

M.J.M. Smulders¹, R.E. Voorrips¹, G.D. Esselink¹, T.M. Santos Leonardo¹, W.P.C. van 't Westende¹, M. Vukosavljev¹, C.F.S. Koning-Boucoiran¹, W.E. van de Weg¹, P. Arens¹, D. Schulz², T. Debener², L. Bellon³, M. Mittmann³, A. Pirani³, T. Webster³, F. Brew⁴, P. Cox⁵ and C. Maliepaard¹

¹ Wageningen UR Plant Breeding, Wageningen, The Netherlands

² Institute for Plant Genetics, Leibnitz Universität Hannover, Germany

³ Affymetrix Inc., Santa Clara 95051, California, USA

⁴ Affymetrix UK Ltd, Wooburn Green, High Wycombe, HP10 OHH, UK

⁵ Roath, Eindhoven, The Netherlands

Keywords: *Rosa hybrida*, rose, SNP, genotyping, tetraploid

Abstract

Rose, as many other important ornamental, vegetable and field crops, is polyploid. This poses constraints in genetic analyses due to the occurrence of multiple alleles at marker and trait loci and the existence of multiple allele dosages. Developments in marker discovery (next generation sequencing), detection (SNP arrays) and analysis (software for dosage scoring) now make it feasible to develop high-density molecular marker maps for the homologous chromosomes in tetraploids separately, and thus perform QTL analysis in tetraploids.

We developed a SNP array for rose to develop genetic maps in tetraploid garden roses and cut roses, which are to be used for inheritance studies and genetic mapping. Here we have indicated the general strategy followed for developing a SNP array and for scoring and using the SNP data generated, and elaborated on the activities undertaken to use the WagRhSNP Axiom array in rose. The array design is not proprietary but can be used by all researchers working in rose.

INTRODUCTION

Rose, as many other important ornamental, vegetable and field crops, is polyploid. This poses constraints in genetic analyses, due to, amongst others, the occurrence of multiple alleles at marker and trait loci, and the existence of multiple allele dosages. In order to have good coverage of all homologous chromosomes/homoeologs of a polyploid crop and for dealing with the more complex inheritance patterns, much larger numbers of markers are necessary than for a diploid. As a consequence, development of genetic and molecular tools for breeding has been slow. Most often, marker development and construction of linkage maps were carried out at the diploid level; translating back this knowledge to the tetraploid level was done only to a limited extent. For instance, several diploid genetic maps have been developed in rose (Debener and Mattiesch, 1999; Crespel et al., 2002; Yan et al., 2005; Dugo et al., 2005; Linde et al., 2006; Hibrand-Saint Oyant et al., 2008; Remay et al., 2009; Spiller et al., 2011), but only few tetraploid maps have been made (Rajapakse et al., 2001; Zhang et al., 2006; Gar et al., 2011; Koning-Boucoiran et al., 2012).

Recently, this situation has changed due to developments in three areas. First, with the advent of next generation sequencing methods it is now feasible to develop large numbers of SNP (single nucleotide polymorphism) markers based on genomic or transcriptomic sequences. Second, SNP array hybridisation enables the detection of thousands of markers simultaneously. Third, software for dosage scoring allows efficient assignment of the tetraploid SNP genotypes of individuals. This opens up various areas of study in polyploids, including QTL mapping of traits, establishing the effect of certain haplotype(s) or 'plus' allele(s), haplotype diversity, and establishing the type of inheritance (disomic/tetrasomic and with/without quadrivalent formations).

We have developed an SNP array for rose to be used for inheritance studies and

genetic mapping in tetraploid garden roses and cut roses. Here we describe the general strategy followed for developing a SNP array and for scoring and using the SNP data generated, and elaborate on the activities undertaken to develop and use the WagRhSNP Axiom array for cut roses and garden roses.

STRATEGY

Our general strategy for mapping in polyploid crops, which we initially applied to tetraploid rose and potato but intend to also apply to other ornamentals, including tetraploid *Alstroemeria* and *Begonia* and hexaploid *Chrysanthemum* in the future, consists of the following steps.

- Step 1: Next generation sequencing, SNP detection, array development, data generation.
- Step 2: Data inspection, filtering, genotype calling with dosage scoring using fitTetra, quality assessment, segregation type assignment.
- Step 3: Assignment of markers to homologs, map construction, consistency checks (double reduction, physical map, synteny, grandparents, etc.).
- Step 4: Map integration of all markers across segregation types.
- Step 5: Developing a model and software for polyploid QTL analysis.

Here we will summarize the actions taken in the first three steps in rose.

The terminology in tetraploids is sometimes confusing. “Homolog” is used here to denote the sequence of nucleotides or SNPs along a single homologous chromosome.

In preparation for this work we have developed software (PedigreeSim, Voorrips and Maliepaard, 2012) for simulating marker data in a tetraploid crop, which models both bivalent and quadrivalent formation during meiosis, including – in some scenarios – a simulation of double reduction (the situation where a gamete receives two copies of the same chromosome segment, which is possible in auto-polyploids when multivalent chromosome configurations are formed during meiosis).

RESULTS

Step 1. Development of Large Numbers of SNPs

When trying to identify SNPs one needs to compare the same DNA sequence among different plants representing the germplasm, or between the two parents of a mapping population, or (in outcrossing species) within the parents. As several thousands of SNPs are sufficient, one would like to limit the DNA to be sequenced to a conveniently small subset that can be obtained from each plant. Probably the easiest method for such ‘complexity reduction’ is to sequence cDNA made from the mRNAs that are being expressed in one or more tissues. As genes only represent a small fraction of the genome, and not all genes are expressed in a given tissue, this gives a strong reduction of the number of sequences that need to be generated (Smulders et al., 2012).

Illumina paired-end sequencing is currently a cost-effective method for sequencing. Even when four different garden rose cDNA samples were combined in one lane, on average still more than 50 million reads per cultivar were obtained. Raw reads were first assembled in contiguous sequences (contigs) of overlapping reads that probably represent pieces of the same gene, some 35.000 to 47.000 transcripts per genotype. We also added similar reads of the two parents of the K5 cut rose population (Koning-Boucoiran et al., 2012) and a smaller number of 454 reads from *R. multiflora*.

SNPs were called using the new version of QualitySNP, called QualitySNPng (Nijveen et al., 2013, <http://www.bioinformatics.nl/QualitySNPng/>). This software uses a haplotype approach to distinguish genuine SNPs from sequence errors. One of the criteria is that at least 5 reads must be found for both haplotypes (alleles) before a SNP is called. We used the extra requirement that 35 bp on both sides of a putative SNP should be free of additional SNPs, as these may interfere with probe hybridization. The procedure yielded over 68,000 putative SNPs.

The Axiom Array for SNP Detection

We chose to develop the SNP array using the Axiom array system of Affymetrix (Santa Clara, CA, USA). Advantages of the Axiom array system for SNP detection include:

- Large size.
- Small size of conserved probe sequence.
- Can be produced for 480 samples (5 microtiter plates) or more.
- Uses photolithographic templates to synthesize probes, so that reordering is possible for a small number of plates and the new arrays will be identical to the original ones.

Axiom arrays are currently also being developed for other rosaceous crops, notably a 90K array for octoploid strawberry within the RosBREED project, and a 400K array for apple within the EU project FruitBreedomics.

The set of SNPs identified in Step 1 formed the basis of the WagRhSNP Axiom array, which contains 68,893 SNPs probed from both directions, totaling 137,786 probes. Probes from both directions can be used to improve the reliability of the genotype calling (see below). This array was used to genotype the two F₁ populations (a cut rose population and a garden rose population) and a diversity panel of garden roses. Using the tetraploid SNP dosage scores we will develop linkage maps, and subsequently work on QTL mapping of traits including frost tolerance for garden roses and flower color, production-related traits and powdery mildew resistance for cut roses, and perform association mapping. The array design is not proprietary but can be used by all research groups working in rose. In a separate paper (Koning-Boucoiran et al., in preparation) we will give the description of all probes and their putative annotation based on homology to the strawberry genome sequence.

Step 2. Dosage Scoring and Genotype Calling

We have developed software for automated genotype calling of bi-allelic SNPs in tetraploid species (fitTetra; Voorrips et al., 2011). This R package enables scoring allele dosages using a mixture model approach. In a tetraploid plant a bi-allelic marker can be present in five different dosage genotypes, nulliplex to quadruplex for one of the alleles (Table 1). The software was used to determine dosages for each of the SNPs and for both probes per SNP (see example in Fig. 1).

In the next step of building the map all erroneous scores will give rise to many problems. As the number of markers is much higher than ever before, we decided that a sensible approach would be to implement a very stringent selection so that map building would start based on a relatively small set of markers that would be mostly free of errors, and subsequently add other groups of markers that were set aside temporarily, while monitoring for unwanted effects. This very stringent selection combined the following criteria:

- Samples were assigned missing scores for a SNP if the dosage could not be assigned with sufficient confidence or if the total signal intensity was too low.
- SNPs were rejected if they contained more than 40% missing values over all samples, or if they were (nearly) monomorphic.
- For use in mapping, additional checks per F₁ populations included:
 - Assigning a missing value to parents if replicate samples had different dosage scores.
 - Rejecting SNPs that did not conform to one of the 20 expected disomic or tetrasomic segregation patterns, or for which the parental scores did not match the observed segregation pattern.
 - Rejecting SNPs with more than 10% missing values in the population.
 - Requiring that both probes of the same SNP produce the same segregation pattern and have (almost) no conflicting scores in the population.

Step 3. Mapping Strategy

The mapping strategy that we followed was aimed to produce separate maps for

all homologs in each of the parents, so $4 \times 7 = 28$ linkage groups per parent. Once these have been generated, they may be merged into consensus maps per parent or per cross, but the separate maps are useful for QTL mapping while assuming that the QTL effects are different per homologous chromosome.

Of the markers that survived the filtering in Step 2, we started with those that segregated 1:1. Note that this group actually consists of four different genotype crosses, which all give 1:1 segregation:

- Nulliplex \times simplex 0×1 $\rightarrow 1:1:0:0:0$
- Nulliplex \times triplex 0×3 $\rightarrow 0:1:1:0:0$
- Simplex \times quadruplex 1×4 $\rightarrow 0:0:1:1:0$
- Triplex \times quadruplex 3×4 $\rightarrow 0:0:0:1:1$

The latter two crosses are actually mirrored situations of the former two. Note also that, in case there would be absolute preferential pairing of chromosomes, simplex \times duplex and duplex \times simplex crosses will segregate 1:1 if the two duplex alleles form a pair in meiosis.

As one follows one allele from one of the parents, it will have high linkage information with another marker if in coupling phase (i.e., on the same homolog), but little linkage information if in repulsion phase (on two different homologs). So it is possible to use this to group the SNP markers into putative homologous chromosomes by clustering.

An alternative approach is to use JoinMap[®] (Van Ooijen, 2011) as if dealing with a diploid population. Recombination frequency estimates of simplex \times nulliplex markers with the simplex alleles in coupling phase are identical for diploids and tetraploids. The markers for a single set of four homologs will still more or less map together, but with a clear gap when moving from one homolog to the next. Usually the gap was about 30 cM, and there was a clear phase change (coupling phase to repulsion phase). This is because under a situation of random pairing of four homologs the minimum expected estimate of the recombination frequency is $1/3$, not because of recombination but the consequence of random assortment of two out of four homologs that are passed on to the progeny.

When we applied this to the SNPs that segregated in the K5 population we could separate 2 to 4 homologs of each of the seven chromosomes per parent. In total, 25 out of 28 linkage groups were generated for parent P540, 26 for parent P867, with up to 119 SNP markers mapped onto a single homolog.

After establishing the maps with only 1:1 segregating markers, markers that segregate 1:2:1 can be added. This group consists of three different types of genotype crosses:

- Simplex \times simplex 1×1 $\rightarrow 1:2:1:0:0$
- Simplex \times triplex 1×3 $\rightarrow 0:1:2:1:0$
- Triplex \times triplex 3×3 $\rightarrow 0:0:1:2:1$

It is not possible to follow individual parental alleles, but the markers do give linkage information between two homologs from the two parents.

The third group that is useful to add are the duplex markers that are absent in the other parent (duplex \times nulliplex, nulliplex \times duplex) as they enable linking two of the homologs within one parent. These simplex \times simplex and duplex \times nulliplex segregating markers also allowed connecting parts of homologs that were still unconnected in Joinmap based on the simplex \times nulliplex markers only.

DISCUSSION

We have developed a 68K rose SNP array and applied it to genotype two segregating tetraploid rose populations and an association panel of tetraploid roses. The SNP array design is available for the rose research community. If the array is used for multiple mapping populations it can be expected that the value of the data will increase, as loci identified in one population can also be tested in others. This would be interesting for QTLs for traits such as scent (Spiller et al., 2010), vigour (Yan et al., 2006), powdery mildew (Linde et al., 2006), flowering time and inflorescence architecture (Kawamura et

al., 2011), and others. We are currently in the process of analyzing the data and generating the maps. Preliminary results show that maps are of high quality, supporting the potential of combining NGS sequencing and high throughput SNP genotyping to identify and use large numbers of SNP markers in genetic analysis. Large numbers of SNP marker data are essential for mapping in polyploids as only a subset of the markers is kept after the stringent filtering process. The data will also be used to study the inheritance in the cut rose and garden rose populations, and macrosynteny with the published genome of diploid strawberry. To facilitate genetic analyses in polyploids we will optimize all necessary steps and combine them into a software pipeline for genetic analysis.

The use of RNAseq as complexity reduction step for sequencing yielded a large number of SNPs that could be identified without adjacent SNP in the flanking (35 nt) regions. The sequences allowed making a direct link to the *Fragaria* genome sequence through blasting the (gene) contigs. Thus, next to generating SNP markers we have also produced a large set of transcripts that will come available for the rose research community, along with other sets of transcripts and ESTs (e.g., Dubois et al., 2012).

An alternative approach to SNP retrieval and array construction is Genotyping-By-Sequencing (GBS), in which SNPs are scored by deep sequencing, thus circumventing the need for SNP array construction. GBS may provide additional information on the number of different alleles present at a particular locus. Given the costs involved in such an approach it is best suited for situations where candidate genes have already been identified. Such a strategy has been applied in tetraploid potato (Uitdewilligen et al., 2013). Complexity reduction by hybridization to baits (SureSelect; Agilent) followed by sequencing was used to determine the five possible alternative allele copy number states for candidate genes. In rose this would be an interesting approach, e.g., for the candidate genes identified for inflorescence architecture (Kawamura et al., 2011).

ACKNOWLEDGEMENTS

This research was partially supported by TTI Green Genetics and the Ministry of Economic Affairs of the Netherlands (TKI-U Polyploids project BO-26.03-002-001). The support of the companies participating in the Polyploids project is gratefully acknowledged.

Literature Cited

- Crespel, L., Chirollet, M., Durel, C.E., Zhang, D., Meynet, J. and Gudin, S. 2002. Mapping of qualitative and quantitative phenotypic traits in *Rosa* using AFLP markers. *Theor. Appl. Genet.* 105:1207-1214.
- Debener, T. and Mattiesch, L. 1999. Construction of a genetic linkage map for roses using RAPD and AFLP markers. *Theor. Appl. Genet.* 99:891-899.
- Dubois, A., Carrere, S., Raymond, O., Pouvreau, B., Cottret, L., Roccia, V., Onesto, J.-P., Sakr, S., Atanassova, R., Baudino, S., Foucher, F., Le Bris, M., Gouzy, J., and Bendahmane, M. 2012. Transcriptome database resource and gene expression atlas for the rose. *BMC Genom.* 13:638.
- Dugo, M.L., Satovic, Z., Millan, T., Cubero, J.I., Rubiales, D., Cabrera, A. and Torres, A.M. 2005. Genetic mapping of QTLs controlling horticultural traits in diploid roses. *Theor. Appl. Genet.* 111:511-520.
- Gar, O., Sargent, D.J., Tsai, C.-J., Pleban, T., Shalev, G., Byrne, D.H., Zamir, D. 2011. An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. *PLoS ONE* 6:e20463.
- Hibrand-Saint Oyant, L., Crespel, L., Rajapakse, S., Zhang, L. and Foucher, F. 2008. Genetic linkage maps of rose constructed with new microsatellite markers and locating QTL controlling flowering traits. *Tree Genet. Genomes* 4:11-23.
- Kawamura, K., Hibrand-Saint Oyant, L., Crespel, L., Thouroude, T., Lalanne, D. and Foucher, F. 2011. Quantitative trait loci for flowering time and inflorescence architecture in rose. *Theor. Appl. Genet.* 122:661-675.

- Koning-Boucoiran, C.F.S., Gitonga, V.W., Yan, Z., Dolstra, O., van der Linden, C.G., van der Schoot, J., Uenk, G.E., Verlinden, K., Smulders, M.J.M., Krens, F.A. and Maliepaard, C. 2012. The mode of inheritance in tetraploid cut roses. *Theor. Appl. Genet.* 125:591-607.
- Linde, M., Hattendorf, A., Kaufmann, H. and Debener, T. 2006. Powdery mildew resistance in roses: QTL mapping in different environments using selective genotyping. *Theor. Appl. Genet.* 113:1081-1092.
- Nijveen, H., van Kaauwen, M., Esselink, D.G., Hoegen, B. and Vosman, B. 2013. QualitySNPng: a user-friendly SNP detection and visualization tool. *Nucl. Acids Res.* 41(W1):W587-W590.
- Rajapakse, S., Byrne, D.H., Zhang, L., Anderson, N., Arumuganathan, K. and Ballard, R.E. 2001. Two genetic linkage maps of tetraploid roses. *Theor. Appl. Genet.* 103:575-583.
- Remay, A., Lalanne, D., Thouroude, T., Le Couviour, F., Hibrand-Saint Oyant, L. and Foucher, F. 2009. A survey of flowering genes reveals the role of gibberellins in floral control in rose. *Theor. Appl. Genet.* 119:767-781.
- 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 Hort.* 961:547-551.
- Spiller, M., Berger, R.G. and Debener, T. 2010. Genetic dissection of scent metabolic profiles in diploid rose populations. *Theor. Appl. Genet.* 120:1461-1471.
- Spiller, M., Linde, M., Hibrand-Saint Oyant, L., Tsai, C.-J., Byrne, D.H., Smulders, M.J.M., Foucher, F. and Debener, T. 2011. Towards a unified genetic map of diploid rose. *Theor. Appl. Genet.* 122:489-500.
- Uitdewilligen, J.G.A.M.L., Wolters, A.-M.A., D'hoop, B.B., Borm, T.J.A., Visser, R.G.F. and van Eck, H.J. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355.
- Van Ooijen, J.W. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetics Res.* 93:343-349.
- Voorrips, R.E., Gort, G. and Vosman, B. 2011. Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinform.* 12:172.
- Voorrips, R.E. and Maliepaard, C.A. 2012. The simulation of meiosis in diploid and tetraploid organisms using various genetic models. *BMC Bioinform.* 13:248.
- Yan, Z., Denneboom, C., Hattendorf, A., Dolstra, O., Debener, T., Stam, P. and Visser, P.B. 2005. Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. *Theor. Appl. Genet.* 110:766-777.
- Zhang, L.H., Byrne, D.H., Ballard, R.E. and Rajapakse, S. 2006. Microsatellite marker development in rose and its application in tetraploid mapping. *J. Am. Soc. Hortic. Sci.* 131:380-387.

Tables

Table 1. The five genotypes possible for bi-allelic markers in a tetraploid.

Genotype	Allele ratio	Fraction b
aaaa = nulliplex	4:0	0.00
aaab = simplex	3:1	0.25
aabb = duplex	2:2	0.50
abbb = triplex	1:3	0.75
bbbb = quadruplex	0:4	1.00

Figures

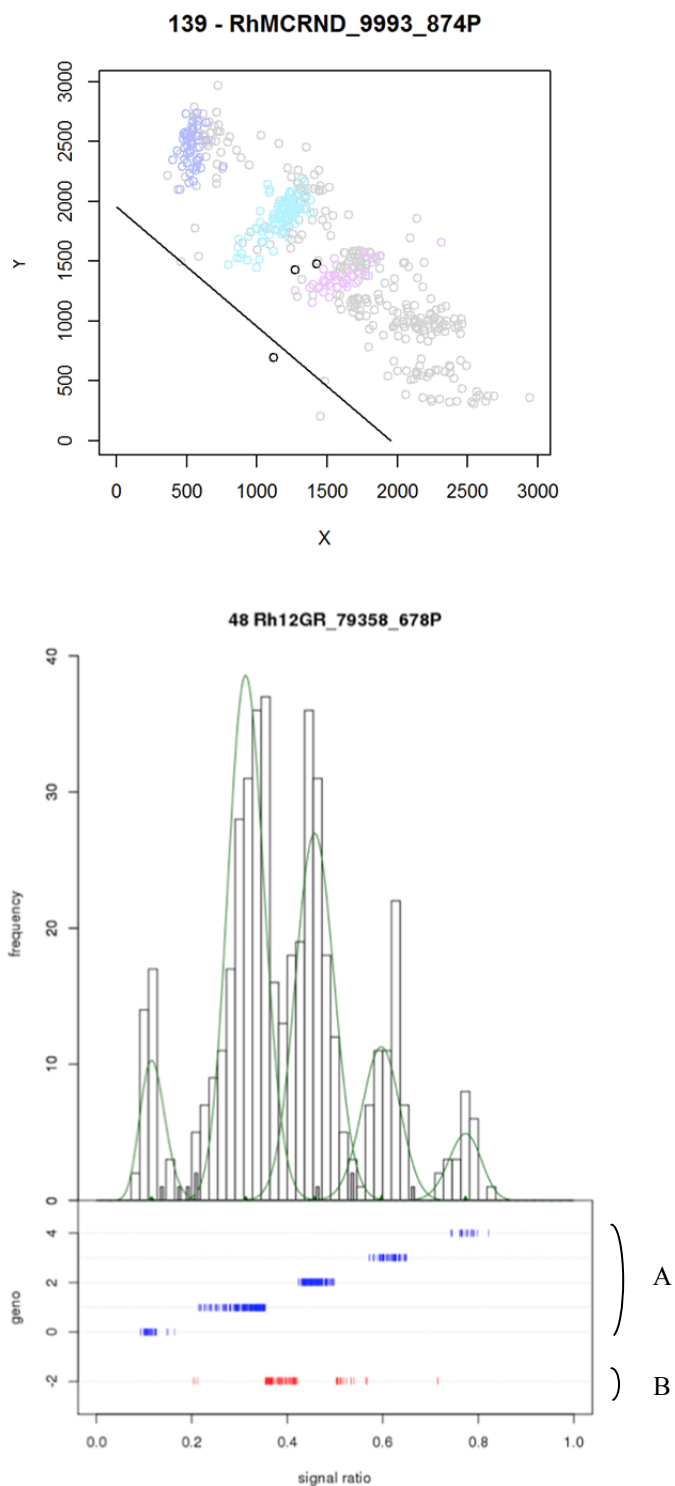


Fig. 1. Top: SNP array signals of individual samples for the two probes used for array hybridisation. The line indicates the minimum signal strength for samples to be analysed. Bottom: fitTetra output with the histogram of signals per sample and the curve fitted to it. Samples for which the genotype was called are indicated with A, while those for which this remains ambiguous are indicated with B.