Analysis of a Database of DNA Profiles of 734 Hybrid Tea Rose Varieties

M.J.M. Smulders, D. Esselink, R.E. Voorrips and B. Vosman[a]
Plant Research International
Wageningen UR Plant Breeding
P.O. Box 16, NL-6700 AA Wageningen
The Netherlands

Keywords: BMT, distinctness, DUS, microsatellite, mutant, SSR, UPOV

Abstract

Over 10,000 Hybrid Tea rose varieties have been described. The large number of varieties and the size of the reference collections may cause problems in DUS (Distinctiveness, Uniformity, and Stability) testing. Molecular markers may help to solve these problems by providing high power to identify and recognise seedling-derived varieties based on unique genotypes, while grouping mutants into groups with identical marker scores. Here we describe the use of a set of 11 rose microsatellite markers to generate a database of molecular profiles of Hybrid Tea varieties. The data were analysed with respect to reproducibility, discriminative power, genetic (sub) structure, and correlation between molecular and DUS characteristics. The use of the markers in the DUS context is discussed with respect to the options 2 and 3 as they were formulated by the UPOV-BMT working group. It is concluded that an option 3 approach (granting of PBR based on distinctness observed with molecular markers) is feasible for rose.

INTRODUCTION

The first Hybrid Tea rose was introduced in 1867 and since then more than 10,000 Hybrid Teas have entered the market. The large number of varieties may cause problems in distinctness, uniformity and stability (DUS) testing. One problem is the requirement to compare new varieties to all other varieties in common knowledge. Clearly, strict adherence to this concept is logistically and financially impossible, and DUS testing stations take a somewhat pragmatic view of common knowledge, limiting it to, e.g., varieties that can be grown in similar climatic zones. Nevertheless, this still means that many hundreds of rose varieties may have to be taken into account. Another problem is the reference collection. Maintaining a collection of greenhouse-grown roses is impractical because of the high costs and the occurrence of diseases. This is also the case for garden roses. When maintaining a collection is abandoned, the examination offices need to request reference varieties from the breeders. It is important that the identity of the material submitted can be verified. For this aspect of quality assurance molecular markers are ideally suited, as they are highly discriminating and can be assayed rapidly.

A different use of molecular markers is for granting Plant Breeders Rights (PBR). This issue is heavily debated. Within UPOV the Biochemical and Molecular Techniques Working Group (UPOV-BMT) was set up to examine some of the issues related to the introduction of molecular markers in DUS testing. Three options are being considered (Button, 2006). Option 1 is the use of molecular characteristics as a predictor of traditional characteristics. It has two variants: (a) the use of molecular characteristics which are directly linked to traditional characteristics (gene specific markers), and (b) the use of a set of molecular characteristics which can be used reliably to estimate traditional characteristics; e.g. quantitative trait loci. Option 2 concerns the calibration of threshold levels for molecular characteristics against the minimum distance between two varieties for traditional characteristics. Option 3 is the development of a new system based on molecular markers. In this approach clearly distinguishable differences based on molecular characteristics would be considered as threshold levels for judging distinctness.

[a] ben.vosman@wur.nl

The application of molecular markers for identification purposes was successfully demonstrated in roses (Esselink et al., 2003), tomato (Bredemeijer et al., 2002) and wheat (Röder et al., 2002). Here, we have generated a database containing the molecular profiles of 734 Hybrid Teas. Since for the first time a database of this size has been established, we set out to analyse the molecular data in detail to determine their applicability for identification purposes. Specifically, we have looked at discriminative power, reproducibility, and genetic (sub) structure in the set of varieties analyzed. Furthermore, we analysed the database in the light of an option 2 or 3 approach as proposed by UPOV. To that purpose we have analysed the correlation between molecular and DUS characteristics (option 2) and whether or not candidate varieties would have been granted PBR when only markers are used to show distinctness (option 3).

MATERIAL AND METHODS

Rose varieties included in the database were based on the list of applications for PBR from the years 1997-2004. DNA was extracted from frozen young leaves using the Qiagen DNA extraction kit. Rose microsatellite (SSR) markers RhE2b, RhAB15, RhAB22, RhD201, RhD221, RhAB40, RhB303, RhM405, RhEO506, RhO517, and RhP519 were analysed as described by Esselink et al. (2003). In the early years (2000-2002), the 11 microsatellite loci were amplified in separate PCR reactions in 96-wells microtiter plates, then combined and analysed in four runs on an ABI 3700. From 2003 onwards the amplifications were done in multiplex format, which greatly reduced the number of handling steps. All samples with problematic data were repeated. Each year, 18 reference and standard varieties were included in the analyses. For reference varieties always the same DNA extraction was used, but standard samples were analysed in duplicate from independent DNA extractions. For error estimation we counted the number of different scores in identical genotypes (which consisted of differences in duplicate varieties included as references, in members of mutant groups, and in samples that were replicated due to poor amplification) separately for 2000-2002 and 2003-2004.

Population genetic analysis is not straightforward for polyploid species, since most programs cannot handle more than two alleles per locus. One approach is to take the presence or absence of each allele as a dominant marker, as an ‘allelic phenotype’ (Esselink et al., 2003). We used the dominant scores to calculate a Jaccard genetic distance with Genstat. As an alternative we also applied SPAGeDi 1.2 (Hardy and Vekemans, 2002), which can handle plants of various ploidy levels. SPAGeDi was also used to calculate the genetic differentiation (Fst) across years and across breeding companies. An overall morphological Euclidean distance was calculated based on 44 DUS trait scores (UPOV guidelines for Rose) without any transformation or normalization, using the Genstat FSIMILARITY command with TEST=euclidean. The presence or absence of structure among the morphological or genetic distances was assessed using a PCO analysis. A specific morphological distance was calculated based only on trait 11730.1 (flower colour) using simple matching: 0 if the two colour scores were equal, 1 if they were different. The correlation between genetic and morphological distances was assessed for both morphological distance measures separately. The association between pairwise genetic distances and pairwise differences in morphological scores was tested by randomization (Mantel test), using 1000 permutations.

RESULTS

The 11 marker loci amplified between 4 (RhM405) and 9 (RhAB40) different alleles (Table 1). The observed number of allelic phenotypes was up to 8 times the number of alleles. The effective number of allelic phenotypes observed was 3.39. Hence, the observed allelic phenotypes for these 11 loci alone can distinguish 20 times as many varieties as exist worldwide. The error rate for 2003 samples was 0.26% of the loci (errors between mutants), for 2004 samples it was 0.30% of the loci (in duplicate samples). Since the average number of alleles per variety across loci was 2.47, this translates into an error rate of about 1 in 1000 for any allele score in the database.

To exclude the possibility that the set of varieties we analyzed contains substructure that needs to be taken into account when making the comparisons between molecular and morphological data, we checked this using the 420 varieties that had been submitted for PBR in the period 2000-2004. Among these there were 407 different genotypes, 13 additional varieties belonged to 12 groups of mutants (consisting of 2-3 identical genotypes each). The genetic differentiation among years for this set was estimated at Fst=0.0007 +/- 0.0005, indicating that every year a similar set of varieties is submitted for PBR. Across the years, we analyzed the differentiation among breeders. There were 45 different breeders, but 12 were present with only one variety, and others with only a few varieties. The varieties from breeders with less than 5 varieties and those from unknown breeders were removed, resulting in 299 varieties grouped in 17 breeding companies. Among these, Fst=0.0056 +/- 0.0011. Apparently, also these 17 companies use basically the same gene pool, although the allele frequencies differ slightly among the companies. Not surprisingly, a PCO analysis of the main variation among the molecular data did not show any obvious structure (result not shown).

A PCO plot (not shown) of the separate DUS characteristics showed three groups of varieties on the first axis (22% explained variation), but not for the second axis. The distinction into three groups was based mainly on the scores of two of the flower colour-related traits: 11732 (spot on the inside) and 11737 (spot on the outside), whose score is either 1 or 9. Two strategies were followed to circumvent this problem of differently scaled measures: an aggregate measure combining all data for the standard set of DUS characteristics, and a focus on the most important traits only.

Using an aggregate morphological distance, we correlated the pairwise genetic distance to the pairwise DUS morphological distance. This produces a large group of data points (Fig. 1). Clearly separate are the mutant pairs, which have identical marker scores. There is a large gap in genetic similarities between mutants and seed-derived varieties, since the latter have a genetic similarity that is always less than 0.95 (Fig. 1). There is no obvious relationship between pairwise genetic and aggregated morphological similarities. This may partly be the result of the way the DUS characteristics were treated, but also due to the use of only 11 marker loci. An overall genetic similarity based on few genomic positions cannot be expected to correlate with an overall similarity based on morphological characteristics.

The most important distinguishing trait in cut rose is flower colour, which is scored in classes (1-19, 34, 40, 46-47, 50; UPOV colour grouping according to the RHS Colour Chart 2001). We determined whether a higher genetic similarity between two varieties increases the probability that these varieties are in the same colour class. Above 0.7 genetic similarity the number of colour matches was significantly higher than expected by chance (Mantel test, $p<0.001$). This may be an indication that these variety pairs have a common ancestry. Alternatively, some colours may occur only in a specific genetic background. However, predicting the colour of a variety based on its genetic similarity with another variety is not reliable as even above 0.7 genetic similarity only 17% of the variety pairs have matching colour, which is hardly useful although higher than the overall frequency of matches (8%). Further, only 0.8% of all pairs have a similarity between 0.7 and 1.0. For (colour) mutants (similarity=1.0) it obviously cannot be used.

DISCUSSION

The microsatellite markers used show a high discriminative power. All seedling varieties could be distinguished and the genetic similarity between the pairs of varieties was always lower than 0.90. This is in line with expectations as Distinctness is usually not a problem with seedling varieties of rose (personal communication Joost Barendrecht, former Technical Expert, Dutch Plant Variety Board). Original varieties and mutants thereof show a genetic similarity of 1.0. So, mutant families can easily be detected. Reliability of the data stored in the database is high, with an error rate of about 1 in 1000. This indicates that for variety identification purposes the database can be a very efficient

tool. When DUS testing stations would completely abandon living reference collections and obtain plant material for comparison from the breeders, they could easily check the identity of the material. What remains is that original varieties and mutants thereof show an identical DNA fingerprint and accidental mixing of mutants would only become apparent upon flowering. In DUS testing of mutant varieties, comparisons are often made with other members of that particular mutant family. A DNA fingerprint made when plant material has been submitted for DUS testing would be sufficient to spot such mutants or mutant groups right away, and based on that result one could include these mutant varieties for comparison. Also in this respect the database would be very helpful.

Of the three options for implementation of molecular markers in variety registration that UPOV considers, only the options 2 and 3 are relevant in the context of this paper. With respect to option 2, we found no correlation between genetic similarities based on morphological characters and molecular characters, except for some correlation in the case of high genetic similarities (above 0.7), but this only refers to a small number of variety pairs. We only used 11 polymorphic microsatellite markers, as that was already sufficient to distinguish all seedling varieties. In order to obtain some correlation, many more marker loci distributed across the genome are probably necessary. In maize, 36 and 51 markers were used to provide some correlation (BMT/10/14). However, an option 2 approach would not be feasible in rose even with much larger number of markers, as mutants will never show such a correlation.

The situation for an option 3 approach is completely different. All seedling varieties showed a unique DNA fingerprint. When granting of PBR would have been based solely on molecular markers, the decisions made would have been identical to those based on traditional DUS testing on seedling varieties. As mutant varieties have a fingerprint that is identical to that of the variety they were derived from, in a system based on DNA markers we would still need to evaluate the mutant varieties for morphological distinctness. This, however, can be very efficiently done as the related mutants or mutant groups are readily identified using the molecular markers. As rose is a vegetatively propagated crop few problems are to be expected with respect to Uniformity and Stability, and these characters can be evaluated in the breeder's premises.

CONCLUSIONS

Based on the data presented here for rose and also on data available from research in Grapevine (BMT11_16) and potato (BMT11_9) an option 3 approach appears feasible. In these three cases use was made of a small set of well-defined markers only. When granting of PBR would have been solely based on markers, only in the case of mutants a different decision would have been taken. The advantages of using markers are evident: decisions on granting PBR can in most cases be taken within a few days after material is submitted for PBR testing, and the costs will be considerably lower than the costs of present day testing using morphological characteristics. Finally, the DNA fingerprint that is made during the application can be used by the breeder to effectively enforce PBR afterwards.

ACKNOWLEDGEMENTS

We thank Yolanda Noordijk for skillfully handling DNA extraction and PCR reactions, and Gerrit van de Wardt for help with analysis of the DUS characteristics. This research was funded under the Method development scheme for Plant Breeders Rights research in The Netherlands. This paper is based on UPOV BMT document 9_12.

Literature Cited

BMT/10/14. Possible use of molecular techniques in DUS testing on maize. How to integrate a new tool to serve the effectiveness of protection offered under the UPOV system.

BMT11_16. The assessment of essential derivation in grapevine.

BMT11_9. Construction of an integrated microsatellite and key morphological

characteristic database of potato varieties on the EU common catalogue Part 1: Discussion of morphological and molecular data.
Bredemeijer, G.M.M., Cooke, R.J. et al., 2002. Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor. Appl. Genet. 105:1019-1026.
Button, P. 2006. New Developments in the International Union for the Protection of New varieties of plants. Acta Hort. 714:195-210.
Esselink, D., Smulders, M.J.M. and Vosman, B. 2003. Identification of cut-rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite markers. Theor. Appl. Genet. 106:277-286.
Hardy, O.J. and Vekemans, X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol. Ecol. Notes 2:618-620.
Röder, M.S., Wendehake, K. et al., 2002. Construction and analysis of a microsatellite-based database of European wheat varieties. Theor. Appl. Genet. 106:67-73.

Tables

Table 1. Power of discrimination of markers in a set of 407 genetically different varieties.

Marker	Number of alleles	Number of allelic phenotypes	PIC value based on allelic phenotypes
RhAB15	6	28	0.72
RhAB201	4	15	0.67
RhAB22	7	23	0.52
RhAB40	9	79	0.76
RhB303	6	37	0.76
RhD221	6	32	0.67
RhE2b	7	32	0.54
RhEO506	6	34	0.72
RhM405	4	9	0.73
RhO517	5	27	0.77
RhP519	6	32	0.71

Figures

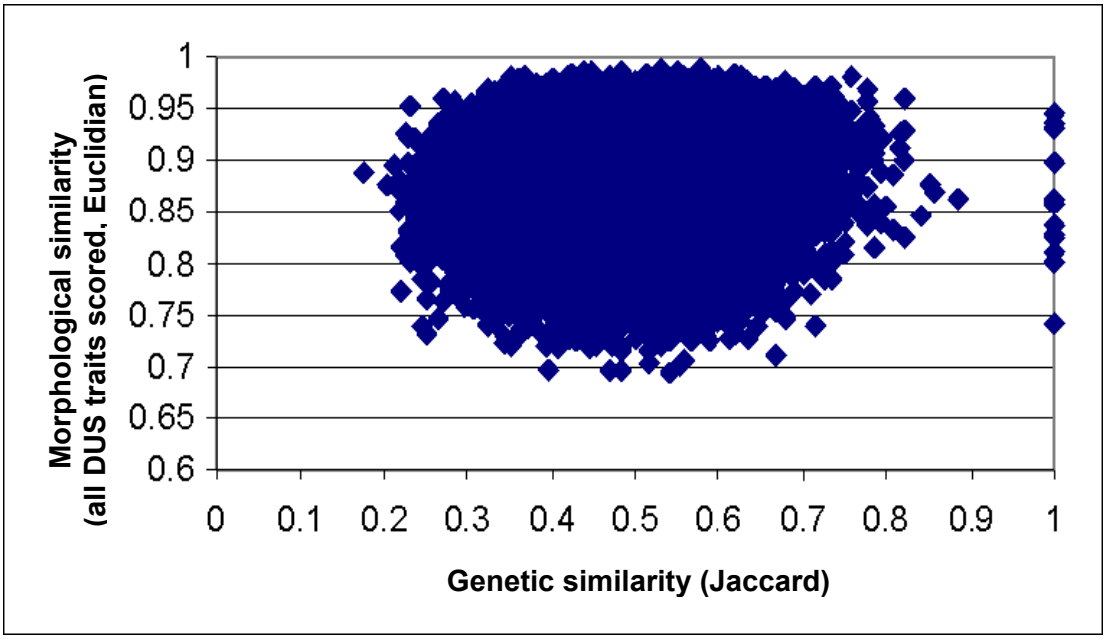

Fig. 1. Genetic versus overall morphological similarity.