# Assignment tests for variety identification compared to genetic similarity-based methods using experimental datasets from different marker systems in sugar beet

Crop Science

de Riek, J.; Everaert, I.; Esselink, D.; Calsyn, E.; Smulders, M.J.M. et al

https://doi.org/10.2135/cropsci2006.04.0265

RESEARCH

# Assignment Tests for Variety Identification Compared to Genetic Similarity-Based Methods Using Experimental Datasets from Different Marker Systems in Sugar Beet

J. De Riek,* I. Everaert, D. Esselink, E. Calsyn, M. J. M. Smulders, and B. Vosman

**ABSTRACT**

High genetic variation within sugar beet (*Beta vulgaris* L.) varieties hampers reliable classification procedures independent of the type of marker technique applied. Datasets on amplified fragment length polymorphisms, sequence tagged microsatellite sites, and cleaved amplified polymorphic sites markers in eight sugar beet varieties were subjected to supervised classifiers, methods in which individual assignments are made to predefined classes, and unsupervised classifiers, defined afterward on the similarity in marker composition from sampled individuals. Major issues addressed are (i) which classification method gives the most consistent results when three marker techniques are compared, and (ii) given different classification techniques available, for which marker technique is the output generated least constrained by the way data analysis is performed. Assignment tests showed a higher consistency across classifications independent from the marker technique. A good allocation to the proper variety was obtained, together with a reliable allocation pattern among the other varieties. Both aspects deal with the variation within a variety and the distance to other varieties. Assignment data were transformed into an average similarity measure, similarity by assignment ($Sa_{x,y}$), which is a new genetic distance measure with interesting properties.

J. De Riek, I. Everaert, and E. Calsyn, ILVO-Institute for Agricultural and Fisheries Research, Plant Science Unit, Caritasstraat 21, 9090 Melle, Belgium. D. Esselink, M.J.M. Smulders, and B. Vosman, Plant Research International, P.O. Box 16, 6700 AA Wageningen, the Netherlands. Received 2 May 2006. *Corresponding author (jan.deriek@ilvo.vlaanderen.be).

**Abbreviations:** AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; $D_{Eucl}$, Euclidean distance; $D_{Nei}$, Nei genetic distance; PCR, polymerase chain reaction; $S_{Jacc}$, Jaccard similarity coefficient; $S_{SM}$, simple matching similarity coefficient; STMS, sequenced tagged microsatellite site.

The genetic characterization by means of molecular markers offers an appealing approach to variety registration and protection. Molecular markers can provide a fast and reliable identification tool applicable during all stages of seed production, trading, and agricultural production and processing. These properties converge with the demand of the seed companies for better protection of hybrids and inbred lines. For many crops, the number of informative morphological characteristics is limited. The high variation within varieties hampers fingerprinting molecular markers and the construction of reference databases containing molecular profiles.

Different types of biochemical and molecular markers have been developed and used in sugar beet (*Beta vulgaris* L.). Biochemical markers (i.e., isozymes or protein patterns) are laborious (Jung et al., 1993) and have a low degree of polymorphisms (Schneider et al., 1999). On the other hand, random amplified polymorphic DNA is insufficiently reproducible across years and laboratories (Barzen et al., 1995; Jones et al., 1997). In sugar beet, amplified fragment length polymorphisms (AFLPs) (Barnes et al., 1996; Barzen et al., 1995; Pillen et al., 1992; Schondelmaier et al., 1996; Schumacher et al.,

Published in Crop Sci. 47:1964–1974 (2007).
doi: 10.2135/cropsci2006.04.0265

677 S. Segoe Rd., Madison, WI 53711 USA

1997) and microsatellites (Rae et al., 2000) have been used for mapping and fingerprinting. Cleaved amplified polymorphic site (CAPS) markers have also been identified (Paran and Michelmore, 1993). Sequenced tagged microsatellite site (STMS) markers have been developed for sugar beet (Rae et al., 2000). Microsatellite markers were shown to be effective in variety identification (Esselink et al., 2003; Bredemeijer et al., 2002; Röder et al., 2002).

We analyzed three different marker datasets (AFLP, STMS, and CAPS). Two types of data analysis were compared: supervised classifiers and unsupervised classifiers. Supervised classifiers represent a group of methods in which individual assignments are made to predefined classes. Unsupervised classification classes are defined a posteriori based on the degree of difference or similarity in marker composition from sampled individuals (Guinand et al., 2002). Two major issues are addressed in this study. First, classification methods were evaluated for using the three marker techniques. Second, given the different classification techniques available, marker techniques were compared to find out which one yields the most reliable or the least constrained summarizing output, independent from the way data analysis was being performed. Finally, we discuss the potential of assignment tests for the identification or evaluation of varieties.

## MATERIALS AND METHODS

### Plant Material, DNA Isolation, and Marker Analysis

Eight sugar beet varieties were included in the present study; six were triploid varieties, two diploid: Ariana (KWS Saat AG, Einbeck, Germany), Aurelia (KWS), Fortis ($2n = 2x$; Hilleshög, Syngenta Seeds, Landskorna, Sweden), Princesse (Delitzsch Pflanzenzucht, Winsen, Germany), Sylvester (Vanderhave, Advanta, Rilland, the Netherlands), H66377 (Vanderhave), KWS8123 ($2n = 2x$; KWS), and MK9907 (Kühn & Co. International B.V., Bergen op Zoom, the Netherlands). Seeds were obtained from the Belgian sugar beet research institute KBIVB-Tienen that is also responsible for variety testing. Thirty individual plants per variety were analyzed. DNA isolation and AFLP (Vos et al., 1995) analysis were described in De Riek et al. (2001) using the commercially available AFLP kit from PerkinElmer Life and Analytical Sciences, Inc. (Waltham, MA) for fluorescent fragment detection. *Eco*RI and *Mse*I were used for DNA digestion. Three primer combinations with six selective bases were applied: *Eco*RI-ACA and *Mse*I-CTG; *Eco*RI-ACT and *Mse*I-CAT; and *Eco*RI-AGG and *Mse*I-CTT (De Riek et al., 2001).

Microsatellites were isolated from enriched small-insert genomic libraries (Esselink, unpublished data, 2000); 12 STMS markers were used: Bvv 15, Bvv 17, Bvv 21, Bvv 23, Bvv 30, Bvv 32, Bvv 43, Bvv 51, Bvv 53, Bvv 60, Bvv 61, and Bvv 64. Sequenced tagged microsatellite site primers were amplified in a 20-μL reaction volume containing 20 ng of genomic DNA, 2 to 10 pmol of each primer, 100 μmol $L^{-1}$ of each dNTP, 10 mmol $L^{-1}$ Tris-HCl pH 9.0, 20 mmol $L^{-1}$ $(NH_4)_2SO_4$, 0.01% Tween 20, 1.5 mmol $L^{-1}$ $MgCl_2$, and 0.3 U Goldstar *Taq* DNA polymerase. Amplifications were performed using a PerkinElmer 9600; polymerase chain reaction (PCR) conditions were 94°C for 3 min followed by 30 cycles of 94°C for 30 s, at the calculated annealing temperature for 30 s, 72°C for 60 s, and a final extension at 72°C for 3 min. According to corresponding reaction conditions different multiplex sets were composed, each containing three microsatellite loci labeled with different fluorescent dyes (FAM, HEX, NED).

Amplified fragment length polymorphism and STMS fragments were separated by polyacrylamide gel electrophoresis on an ABI Prism 377 DNA Sequencer (Applied Biosystems, Foster City, CA) on 36-cm gels using 4.25% denaturing polyacrylamide (4.25% acrylamide/bisacrylamide 19:1, 6 mol $L^{-1}$ urea in 1× TBE). GS-500 ROX-labeled size standard (PerkinElmer) was loaded in each lane to facilitate the automatic analysis of the gel and the sizing of the fragments. Genescan 2.1 (Applied Biosystems) was used to estimate detection time, signal peak height, and surface for each fragment. Amplified fragment length polymorphism scoring was conducted as described by De Riek et al. (2001). For STMS analysis, only a selected set was used; null alleles were ignored, and alleles whose frequency was below 1% were excluded.

Cleaved amplified polymorphic site markers were selected from a list of codominant markers developed for sugar beet by Schneider et al. (1999; Table 1). Polymerase chain reaction conditions were as specified by Schneider et al. (1999). The fragments were separated on a 2% Tris-borate-EDTA–buffered agarose gel after electrophoresis (4 h, 80 V). A scoring table (1/0) was generated by visual scoring after ethidium bromide staining and UV lighting of the gel. According to a GeneRuler 100bp Plus and a GeneRuler 50bp ladder (Fermentas International Inc., Burlington, ON) fragment size was estimated. For all markers, CAPS alleles were scored as present or absent.

### Statistical Analyses

The Jaccard ($S_{Jacc}$) and simple matching similarity ($S_{SM}$) coefficients between two genotypes, the Euclidean distance ($D_{Eucl}$) and the Nei distance ($D_{Nei}$) between two populations, and bootstrapping procedures were as in De Riek et al. (2001). A "dominant" scoring was used for all marker techniques, that is, presence of marker bands (marker frequencies) instead of allelic composition (allelic frequencies).

**Table 1. Gene localization and restriction enzyme used for the cleaved amplified polymorphic site (CAPS) primer sets taken from Schneider et al. (1999).**

| Code | Name | Restriction enzyme | Localization |
|---|---|---|---|
| pr2 | Rubisco, small subunit | *Rsa*I | Calvin cycle |
| pr3 | Triosephosphate isomerase | *Taq*I | Calvin cycle |
| pr4 | ADPG-pyrophosphorylase, large subunit | *Taq*I | Metabolism of transient starch |
| pr5 | Granule bound starch synthase | *Taq*I | Metabolism of transient starch |
| pr6 | Glucose-6-P/phosphate translocator I/II | *Mse*I | Transport processes |
| pr7 | Sucrose/proton symporter | *Mse*I | Transport processes |
| pr8 | Vacuolar pyrophosphatase 2 | *Taq*I | Transport processes |
| pr9 | Adenylate transporter | *Taq*I | Transport processes |
| pr11 | Pyruvate dehydrogenase, subunit A | *Taq*I | Oxidative decarboxylation |
| pr13 | Aspartate aminotransferase | *Taq*I | Nitrogen metabolism |

Table 2. Key figures of the three molecular methods used.

| Method† | Total dataset | | | | On average per single variety | |
|---|---|---|---|---|---|---|
| | No. loci | No. scored bands | Averaged per locus | | Averaged per locus | |
| | | | No. scored bands | No. allelic phenotypes | No. scored bands | No. allelic phenotypes |
| AFLP | 405 | 405 | 1 | 2 | 1 | 2 |
| Microsatellites | 12 | 53 | 4.42 ± 1.62 | 14.6 ± 8.3 | 3.08 ± 1.25 | 5.33 ± 2.83 |
| CAPS | 10 | 57 | 5.70 ± 2.98 | 14.2 ± 10.1 | 5.56 ± 2.02 | 10.00 ± 3.81 |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site.

Analysis of molecular variance (AMOVA) was applied (Schneider et al., 2000) to the Euclidean distance matrix between individual genotypes to partition genetic variation among ploidy level and varieties. The assignment based on the highest probability of an individual's genotype in any of the populations was calculated using the "Doh" software (http://www2.biology.ualberta.ca/jbrzusto/ verified 7 August 2007) starting from the 0/1 data as described by Paetkau et al. (1995, 1997). This includes the calculation of a matrix of distances (**d**) between each pair of populations, calculated as

$$\mathbf{d}_{x,y} = \left(\mathbf{A}_{x,y} + \mathbf{A}_{y,x}\right)\Big/2 \qquad [1]$$

based on the nonsymmetric matrix $\boldsymbol{A}$ defined by

$$\mathbf{A}_{x,y} = 1/n_x \sum_{ix}\left[\log_{10}\left(\mathrm{Pr}_x\, g_i/\mathrm{Pr}_y\, g_i\right)\right] \qquad [2]$$

where $x$, $y$ are populations, $n_x$ is the size of population $x$, $g_i$ is the genotype of individual $i$, and $\mathrm{Pr}_x$ is the genotype probability calculated in population $x$. $\mathbf{A}_{x,y}$ is a measure of how much more likely genotypes of individuals sampled in population $x$ are in population $x$ than in population $y$. **A** is not symmetric.

For the assignment tests based on the pairwise similarity matrix (De Riek et al., 2001), first a ranking of the 30 most similar partners was made per individual plant. For this, a (240 by 240) similarity matrix was constructed, using $S_{\mathrm{Jacc}}$ or $S_{\mathrm{SM}}$. Second, per individual plant, the origin variety for each of the, for example, 10 individual genotypes that show the highest similarity was recorded (we used the 3, 10, or 30 most highly ranking partners). The partitioning of the origin of these highly ranking partners over all varieties in the dataset is then displayed in an assignment table for each variety under analysis, grouping the assignments of all individuals of the variety under analysis. In an identical way as described by Paetkau et al. (1997) for the derivation of $\mathbf{d}_{x,y}$ from $\mathbf{A}_{x,y,}$ (and shown above) this asymmetric assignment table was converted into a symmetric similarity measure, similarity by assignment ($\mathrm{Sa}_{x,y}$), by making it a relative value and simply averaging the assignment values of $\mathrm{Vass}_{x,y}$ and $\mathrm{Vass}_{y,x}$.

Mantel analysis (Mantel, 1967; Mantel Nonparametric Test Calculator for Windows, Version 2.00, 1999, by Adam Liedloff) computed standardized Mantel's statistics between two similarity matrices. The significance of the statistic was evaluated by permutations (1000×) and expressed as a probability (Smouse et al., 1986).

# RESULTS AND DISCUSSION

## Characteristics of the Experimental Data Sets

In total, 405 AFLP, 12 STMS, and 10 CAPS markers were selected for genotyping this set of varieties. For all marker techniques, the presence or absence of (allelic) bands in the individual plants were scored. For the STMS markers, in total 53 different alleles with a allelic band frequency above 0.01 were scored; for 48 of these the allelic band frequency was above 0.05. Ten CAPS markers taken from Schneider et al. (1999) gave, in total, 57 bands.

In Table 2 some key figures differentiating the power of the three molecular markers used are listed. For comparison of the two codominant techniques, the average numbers of bands per locus for the total dataset or on average per single variety are listed. Power of discrimination between individual plants can also be evaluated by the number of unique allelic phenotypes (Becher et al., 2000; Esselink et al., 2003), representing a unique combination of alleles for a particular codominant marker in a given genotype. This approach was introduced to circumvent the problem of determining the actual genotype in a polyploid species if one cannot exactly determine whether a specific allele is present in two or three copies (Esselink et al., 2004, Nybom et al., 2004). These descriptive statistics (Table 2) make clear that the large number of AFLP markers outcompete the codominant marker datasets in characteristics related to the amount of data points. The datasets for STMS and CAPS are comparable for most statistics, including the number of allelic phenotypes that they can distinguish within the total dataset. However, STMS tend to detect fewer allelic phenotypes within one variety, which suggest that this particular dataset may be superior in revealing differences between varieties.

## Unsupervised Classification based on Differences in Similarity and Clustering

### *Genetic Distances between Individual Plants*

For the separate marker datasets, individual pairwise similarity matrices (240 by 240) were constructed using $S_{\mathrm{Jacc}}$ and $S_{\mathrm{SM}}$ and the binary $D_{\mathrm{Eucl}}$; here on, Mantel's tests were applied to evaluate the concordance of the genetic relationships revealed by each of the marker techniques. The Mantel correlation coefficient $r$ ranged from −0.007 for the comparison AFLP and CAPS to 0.14 for the comparisons CAPS to STMS and AFLP to STMS ($P < 0.02$). These values indicated a rather poor correspondence between the data structure of the three matrices at the individual plant level. However, when comparing varieties, the direct relationships between the individual genotypes present within varieties is not the only concern but merely the

overall view on the group of genotypes making a variety. Therefore, we took the analysis to that level. Table 3 gives the average $S_{Jacc}$ and $S_{SM}$ taken over all pairwise comparisons between two accessions for each marker technique. Values closer to 1 indicate higher similarity. Across all pairwise comparisons between plants within a variety, KWS8123 had the highest internal average similarity independently of the similarity measure or the marker dataset. The variety showing the lowest internal average similarity did depend on the dataset: for AFLP, Ariana, MK9907, and Sylvester are among the lowest; for CAPS, Fortis and MK9907; and for STMS, Sylvester, Ariana, and Princesse. When the internal average similarity of a certain variety was compared to the average similarities between this particular variety and the rest, the internal average similarity was always higher indicating that, with no exception, plants belonging to a particular variety are always on average more similar to themselves than to another variety. This indicates that all datasets at least partly reveal the genetic structure of the varieties.

### Classification based on Ordinations from Marker Frequency Data

In Table 4 the standard $D_{Nei}$ between pairs of varieties and its standard errors are given. Using the AFLP dataset the standard $D_{Nei}$ range from 0.02 to 0.08; using the CAPS dataset, from 0.02 to 0.13; and using the STMS dataset, from 0.4 to 0.22. The range as obtained with this AFLP dataset is somewhat higher than reported before (De Riek et al., 2001) using the same AFLP primer combinations. However, in the previous study, 90 individual plants taken from three seed production years were analyzed, which makes the potential variation within a variety larger (and hence the distances between varieties lower).

The range of $D_{Nei}$ indicate that the discriminatory capacity was lowest for the AFLP dataset, despite its including the highest number of data points, and increased for CAPS and STMS. However, standard errors for CAPS and STMS data were much higher (Table 4 above diagonal) making them a less precise estimate.

**Table 3. The average similarity taken over all pairwise comparisons between plants of two accessions for amplified fragment length polymorphism (AFLP), cleaved amplified polymorphic site (CAPS), and sequenced tagged microsatellite site (STMS) markers (below the diagonal Jaccard; above the diagonal Simple Matching coefficients). On the diagonal the average similarity for all pairwise comparisons internal to each variety is given for both coefficients.**

| | Ariana | Aurelia | Fortis | H66377 | KWS8123 | MK9907 | Princesse | Sylvester |
|---|---|---|---|---|---|---|---|---|
| **AFLP** | | | | | | | | |
| Ariana | *0.7077* 0.5406 | *0.7090* | *0.6765* | *0.6793* | *0.6895* | *0.6804* | *0.6968* | *0.6783* |
| Aurelia | 0.5399 | *0.7301* 0.5647 | *0.6959* | *0.6962* | *0.7001* | *0.6945* | *0.7090* | *0.6958* |
| Fortis | 0.4992 | 0.5213 | *0.7211* 0.5521 | *0.6832* | *0.6799* | *0.6754* | *0.6871* | *0.6861* |
| H66377 | 0.5204 | 0.5396 | 0.5222 | *0.7242* 0.5866 | *0.6931* | *0.6933* | *0.6992* | *0.6959* |
| KWS8123 | 0.5258 | 0.5372 | 0.5105 | 0.5430 | *0.7418* 0.5960 | *0.6894* | *0.6993* | *0.6865* |
| MK9907 | 0.5088 | 0.5248 | 0.5006 | 0.5394 | 0.5263 | *0.7088* 0.5468 | *0.6903* | *0.6821* |
| Princesse | 0.5306 | 0.5439 | 0.5159 | 0.5479 | 0.5410 | 0.5251 | *0.7362* 0.5834 | *0.6942* |
| Sylvester | 0.5085 | 0.5282 | 0.5145 | 0.5435 | 0.5251 | 0.5156 | 0.5317 | *0.7040* 0.5433 |
| **CAPS** | | | | | | | | |
| Ariana | *0.7438* 0.6751 | *0.7386* | *0.7204* | *0.6952* | *0.7349* | *0.6940* | *0.7335* | *0.7125* |
| Aurelia | 0.6712 | *0.7654* 0.7015 | *0.7166* | *0.7063* | *0.7476* | *0.6978* | *0.7473* | *0.7404* |
| Fortis | 0.6449 | 0.6425 | *0.7412* 0.6615 | *0.6656* | *0.7265* | *0.6941* | *0.7172* | *0.7131* |
| H66377 | 0.6337 | 0.6467 | 0.5990 | *0.7447* 0.6930 | *0.6896* | *0.6780* | *0.7001* | *0.6929* |
| KWS8123 | 0.6529 | 0.6687 | 0.6384 | 0.6159 | *0.8188* 0.7402 | *0.6791* | *0.7381* | *0.7162* |
| MK9907 | 0.6279 | 0.6338 | 0.6219 | 0.6227 | 0.6002 | *0.7100* 0.6492 | *0.6870* | *0.6962* |
| Princesse | 0.6729 | 0.6888 | 0.6513 | 0.6482 | 0.6672 | 0.6310 | *0.7625* 0.7120 | *0.7281* |
| Sylvester | 0.6504 | 0.6803 | 0.6456 | 0.6397 | 0.6413 | 0.6387 | 0.6749 | *0.7481* 0.6940 |
| **STMS** | | | | | | | | |
| Ariana | *0.7567* 0.4788 | *0.7541* | *0.6968* | *0.6701* | *0.7424* | *0.6363* | *0.7148* | *0.6799* |
| Aurelia | 0.4568 | *0.8011* 0.5181 | *0.7443* | *0.7005* | *0.7833* | *0.6928* | *0.7310* | *0.7075* |
| Fortis | 0.3431 | 0.3956 | *0.8097* 0.4928 | *0.7391* | *0.7810* | *0.7097* | *0.7134* | *0.7225* |
| H66377 | 0.3544 | 0.3779 | 0.4109 | *0.7660* 0.4994 | *0.7220* | *0.7003* | *0.6825* | *0.7014* |
| KWS8123 | 0.3844 | 0.4317 | 0.3983 | 0.3538 | *0.8726* 0.5714 | *0.7346* | *0.7336* | *0.7028* |
| MK9907 | 0.2972 | 0.3517 | 0.3515 | 0.3877 | 0.3581 | *0.7963* 0.5305 | *0.6583* | *0.6873* |
| Princesse | 0.4150 | 0.4202 | 0.3696 | 0.3724 | 0.3690 | 0.3272 | *0.7571* 0.4789 | *0.6789* |
| Sylvester | 0.3630 | 0.3829 | 0.3813 | 0.3965 | 0.3194 | 0.3653 | 0.3458 | *0.7433* 0.4557 |

Table 4. The standard Nei genetic distance (× $10^{-2}$) between pairs of varieties (below the diagonal) and its standard errors (above the diagonal) for the three marker techniques used.†

| | Ariana | Aurelia | Fortis | H66377 | KWS8123 | MK9907 | Princesse | Sylvester |
|---|---|---|---|---|---|---|---|---|
| **AFLP** | | | | | | | | |
| Ariana | | 0.28 | 0.64 | 0.54 | 0.71 | 0.61 | 0.48 | 0.53 |
| Aurelia | 1.56 | | 0.53 | 0.40 | 0.69 | 0.51 | 0.52 | 0.41 |
| Fortis | 5.54 | 4.71 | | 0.62 | 0.81 | 0.70 | 0.65 | 0.63 |
| H66377 | 5.09 | 3.60 | 6.08 | | 0.75 | 0.46 | 0.67 | 0.40 |
| KWS8123 | 6.88 | 6.55 | 7.55 | 7.49 | | 0.83 | 0.80 | 0.67 |
| MK9907 | 5.97 | 4.72 | 6.95 | 3.99 | 7.37 | | 0.78 | 0.44 |
| Princesse | 4.18 | 4.55 | 4.59 | 6.62 | 7.31 | 7.05 | | 0.57 |
| Sylvester | 4.64 | 3.81 | 5.17 | 3.21 | 6.88 | 3.42 | 4.40 | |
| **CAPS** | | | | | | | | |
| Ariana | | 0.68 | 1.22 | 1.7 | 2.01 | 1.38 | 1.04 | 1.45 |
| Aurelia | 2.22 | | 1.30 | 2.16 | 1.59 | 1.69 | 0.74 | 0.86 |
| Fortis | 2.96 | 4.97 | | 3.04 | 2.42 | 1.37 | 1.61 | 1.52 |
| H66377 | 6.93 | 6.67 | 10.9 | | 3.62 | 1.73 | 2.40 | 2.32 |
| KWS8123 | 5.87 | 5.66 | 7.02 | 12.52 | | 3.21 | 1.99 | 2.14 |
| MK9907 | 4.62 | 5.64 | 4.50 | 6.92 | 11.81 | | 1.65 | 1.34 |
| Princesse | 2.63 | 2.22 | 4.58 | 7.35 | 6.49 | 6.81 | | 1.34 |
| Sylvester | 4.63 | 2.17 | 4.31 | 7.35 | 8.92 | 4.60 | 3.71 | |
| **STMS** | | | | | | | | |
| Ariana | | 1.39 | 3.17 | 3.48 | 2.76 | 5.27 | 1.78 | 2.39 |
| Aurelia | 4.13 | | 2.73 | 2.89 | 3.00 | 4.05 | 2.00 | 2.54 |
| Fortis | 11.91 | 8.78 | | 2.60 | 2.55 | 3.96 | 2.63 | 1.72 |
| H66377 | 14.00 | 12.21 | 6.94 | | 3.47 | 3.01 | 2.81 | 1.86 |
| KWS8123 | 10.42 | 8.42 | 6.67 | 12.91 | | 4.30 | 2.35 | 2.88 |
| MK9907 | 21.64 | 14.77 | 14.92 | 11.53 | 14.07 | | 4.30 | 3.17 |
| Princesse | 6.41 | 7.07 | 7.88 | 10.83 | 8.91 | 17.49 | | 2.26 |
| Sylvester | 10.36 | 9.49 | 7.06 | 7.29 | 11.85 | 13.19 | 9.90 | |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; STMS, sequenced tagged microsatellite site.

Clustering and bootstrapping were used to compare groupings by the three marker techniques (i.e., AFLP, STMS, and CAPS), employing Nei's pairwise distances (Fig. 1). Dendrograms were constructed with the UPGMA-algorithm. Regardless the marker data set used, Ariana, Aurelia, and Princesse clustered together at the highest similarity level. With the CAPS dataset, Sylvester is also attributed to this cluster. However, AFLP and STMS bootstrap values restrict the cluster to Ariane and Aurelia. Clustering of MK9907, Sylvester, and H66377 with the AFLP data set was not supported by the other datasets.

## Supervised Classification Techniques Using AMOVA and Assignment Tests

### *Analysis of Molecular Variance*

The AMOVA procedure (Excoffier et al., 1992) provides a general framework for the analysis of population genetic structure based on any distance matrix. We applied a genetic structure design on Euclidean distances between individual plant genotypes, with allocation of the variation to the ploidy level (diploid versus triploid varieties), and within ploidy level, to varieties. For the three marker datasets used, the major part of the variation could be attributed to variation within varieties ranging from almost 95% for the AFLP to 84% for the STMS datasets (Table 5). Only a small part of the variation was accounted for by ploidy differences.

Neither molecular method was very diagnostic in differentiating between varieties with the STMS method having the highest differentiation values of 14%. The *F*-statistic indices $F_{st}$ (partitioning of all variation), $F_{sc}$ (variation within each ploidy level), and $F_{ct}$ (variation due to the ploidy difference alone) reflect the above observations. $F_{st}$ and $F_{sc}$ estimates were always significant (1023 permutations) for the three data sets used, while $F_{ct}$ estimates were not significant. $F_{st}$ was highest for the STMS (0.15) and lowest for the AFLP dataset (0.056). The population pairwise *F*-statistics matrices revealed a comparable data structure as the use of the standard Nei distance (data not shown).

### *Assignment based on the Highest Probability of an Individual's Genotype in Any of the Populations*

The method described by Paetkau et al. (1995, 1997) starts with the calculation of the probability of the assignment (assignment index) of each individual plant to each variety (data not shown). Assignment tests typically generate asymmetric matrices, showing the number of plants attributed to a certain variety based on this index. For the three marker datasets, plants were in general classified to their original variety (Table 6). The highest assignments were obtained with the STMS dataset. Correlation between assignments based on the different marker datasets was >0.95, significant at $P = 0.001$. The method also generates a derived distance measure $\mathbf{d}_{x,y}$ (Table 6); distances here are to be understood as the ratios of the probability that an individual plant belongs to the original variety compared to the other variety on a logarithmic scale. Only a poor agreement was obtained between

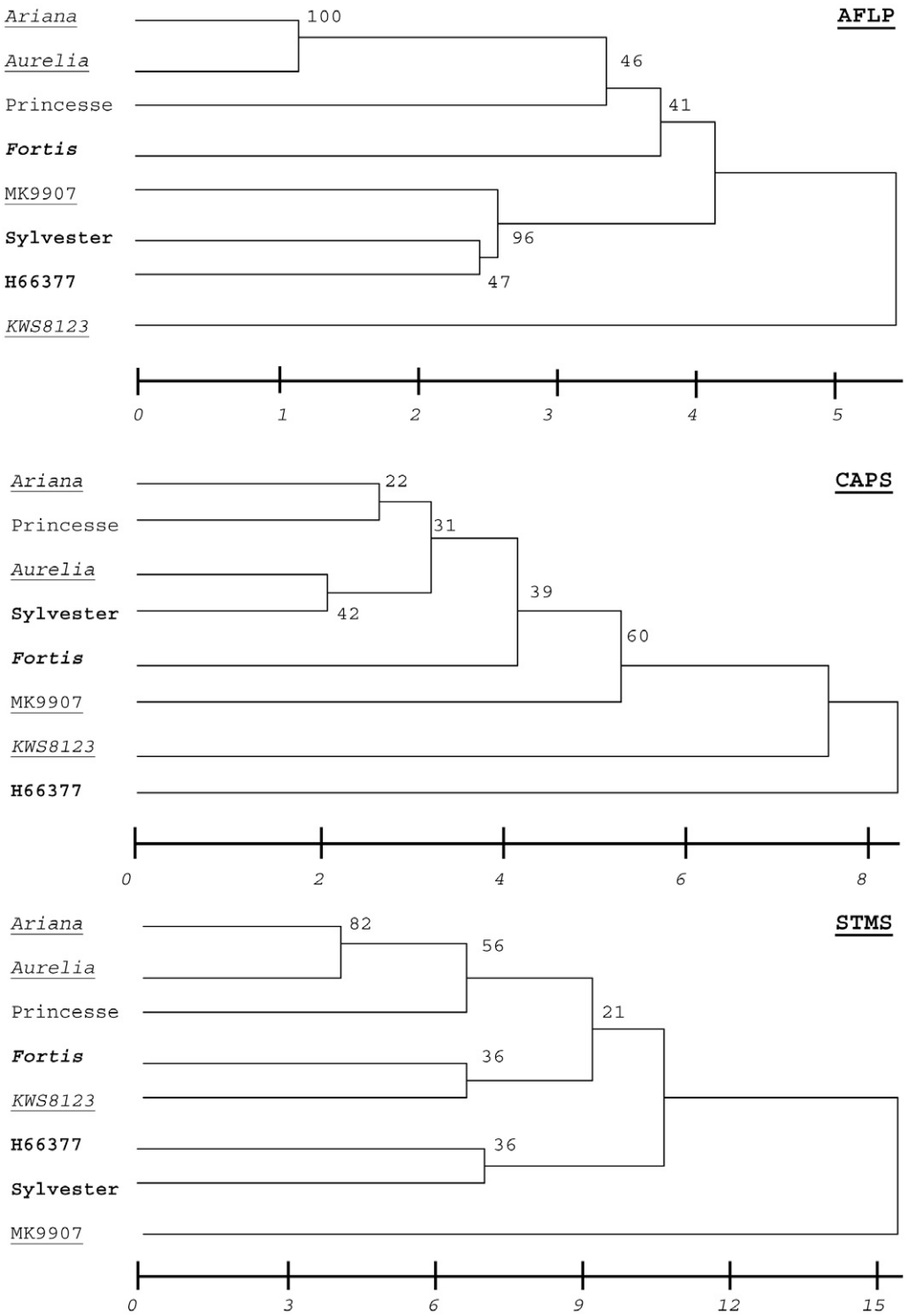


Figure 1. Dendrograms from the standard Nei genetic distances for amplified fragment length polymorphism (AFLP), cleaved amplified polymorphic site (CAPS), and sequenced tagged microsatellite site (STMS) based ordinations (UPGMA-clustering). Bootstrap values are indicated at the nodes.

$\mathbf{d}_{x,y}$ matrices based on the different markers by Mantel testing, indicating that apart from the apparently different levels of the matrix $\mathbf{d}_{x,y}$, also the data structure of the matrices derived from the three marker datasets was different (Tables 6 and 8).

### *Assignment based on the Pairwise Similarity Data for Individual Plants*

Estimates of within variety genetic variation were also directly assessed from the pairwise similarity data for individual plants using $S_{\text{Jacc}}$ and $S_{\text{SM}}$ (Table 7). Compared to the method by Paetkau et al. (1995), this assignment

**Table 5. AMOVA table showing the distribution of the molecular variance according to groups (ploidy level) and populations with indication of derived fixation indices.†**

| Source of variation | df | Sum of squares | Variance components | Percentage of variation |
|---|---|---|---|---|
| **AFLP** | | | | |
| Among groups | 1 | 9.751 | 0.01714 | 0.52 |
| Among populations within groups | 6 | 49.247 | 0.16903*** | 5.09 |
| Within populations | 232 | 727.764 | 3.13691*** | 94.40 |
| Total | 239 | 786.762 | 3.32309 | |
| Fixation indices | | | | |
| $F_{SC}$: 0.05113*** | | | | |
| $F_{ST}$: 0.05602*** | | | | |
| $F_{CT}$: 0.00516 | | | | |
| **CAPS** | | | | |
| Among groups | 1 | 9.936 | 0.04114 | 2.00 |
| Among populations within groups | 6 | 37.398 | 0.14549*** | 7.08 |
| Within populations | 232 | 433.459 | 1.86836*** | 90.92 |
| Total | 239 | 480.792 | 2.05498 | |
| Fixation indices | | | | |
| $F_{SC}$: 0.07224*** | | | | |
| $F_{ST}$: 0.09082*** | | | | |
| $F_{CT}$: 0.02002 | | | | |
| **STMS** | | | | |
| Among groups | 1 | 12.822 | 0.03231 | 1.65 |
| Among populations within groups | 6 | 59.484 | 0.27537*** | 14.05 |
| Within populations | 232 | 383.445 | 1.65278*** | 84.31 |
| Total | 239 | 455.751 | 1.96047 | |
| Fixation indices | | | | |
| $F_{SC}$: 0.14282*** | | | | |
| $F_{ST}$: 0.15695*** | | | | |
| $F_{CT}$: 0.01648 | | | | |

***Significant ($P < 0.001$) at 1023 permutations.

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; STMS, sequenced tagged microsatellite site.

method produces much more dispersion across different varieties. Marker datasets markedly differed in this respect. Similarities based on the AFLP dataset yielded the highest allocation among varieties, CAPS-based similarities were less dispersed. Sequence tagged microsatellite site-based similarities were clearly more distinguishing. This can be evaluated as the number of plants that were traced back to the proper variety. Here, the assignment based on the STMS dataset is much more variety specific indicating that here profiles of varieties are more typical.

We propose that the allocation pattern among varieties of a single variety can be used as a measure for the variety differentiation, particularly as the allocation pattern among varieties was found to be relatively independent of the marker dataset used. For instance, Ariana and Aurelia at one side and KWS8123 at the other can be taken as examples of varieties that are very easily distinguishable from each other and from the remaining varieties in the datasets, although they clearly refer to a common gene pool. This can be seen from the (low) degree with which KWS8123 plants are allocated to Ariana and Aurelia. In addition, Princesse refers to this same breeding pool but to a lesser extent. The same can be observed for Sylvester, H66377, and MK9907.

For these assignments too, the correlation between assignments based on the different marker datasets was high ($r > 0.88$, significant at 0.001).

To better reveal the information content, Table 7 was turned into a symmetric similarity measure ($Sa_{x,y}$) by making it a relative value and simply averaging as introduced for $\mathbf{d}_{x,y}$ in Doh (see Materials and Methods). As such, $Sa_{x,y}$ can be compared to Table 3 as it also reports on variation within and between varieties in one format. In contrast to the similarity values in Table 3, the $Sa_{x,y}$ results have much more internal structure to reveal differences within and between varieties. Note that $Sa_{x,y}$ is a relative measure as it is influenced by the number of plants analyzed in each accession (here, equal numbers were taken) and, more important, by the overall composition of the set of varieties taken into the analysis.

## Comparison of Approaches Used

Summarizing output has been generated for the three marker datasets under the form of different resemblance measures:

- the average Jaccard and simple matching similarity taken over all pairwise comparisons between plants of two accessions (Table 3)
- the standard Nei genetic distances (Table 4) and Euclidean distances between pairs of varieties based on the marker frequencies within each variety
- AMOVA generated population pairwise *F*-statistics matrices (data not shown) from the Euclidean distance matrix between all individuals
- the distance matrix $\mathbf{d}_{x,y}$ (Table 6) from the assignment test by Paetkau et al. (1995)
- the newly defined similarity by assignment, $Sa_{x,y}$ (Table 7)

Two major issues need to be addressed: (i) Which classification method gives the most consistent results when the three marker datasets are compared, and (ii) which marker technique yields the most reliable or least constrained summarizing output? To test the concordance between the different summarizing matrices Mantel's tests were made (Tables 8 and 9). Table 8 describes the comparison of marker techniques for the different summarizing methods. The Mantel's statistic then indicates if, given a summarizing technique, the data structure revealed by the measure is consistent between marker techniques. This can most easily be judged from the correlation coefficients *r* and the corresponding probabilities *P*. The use of the average Jaccard similarity or standard Nei genetic distance generally yields a weakly correlated summarizing output.

Especially, the correlation between AFLP and CAPS is low when the average Jaccard similarity is applied. Significances are low, as can be seen from the *P* values that are seldom below the 1% level. In contrast, with the $Sa_{x,y}$ a good correlation between summarizing output generated from the different marker techniques was obtained: *r* was always above 0.80 and, more importantly, *P* is within the 1% level (or close to 1% for the comparison CAPS to STMS).

In addition to the issue raised above, which is relevant when data from different marker origins are to be compared, it is essential to find out which marker technique is the least constrained by different final data analysis. This can be determined from Table 9, in which matrices are compared for the three molecular methods. In general, the correlations for the different methods are high within a marker technique (Table 9). For the summarizing output generated from the three listed methods for CAPS and STMS datasets, the significances of the correlations as seen from *P* are within the 1% level. For the AFLP dataset, there seems to be a less correlated output when the standard Nei genetic distance is being applied.

## CONCLUSIONS

In this paper, we evaluated a number of statistical analyses to identify and characterize sugar beet varieties. As stated before by Manel et al. (2005) in an evaluation of different assignment methods to match different biological questions, it is currently, from a theoretical background, often not possible to say with certainty which of the statistical methods perform best and under which conditions. This strengthens the importance of the present comparative analyses.

A first type of analysis techniques mainly focuses on the differences between varieties and largely ignores the within-variety variation. A commonly used method is the use of marker frequency data. From Table 4 it can be seen that genetic distances between varieties exist; depending on the specific marker dataset, they are more or less significant when compared to

**Table 6. Assignment of individual genotypes per variety and derived distance measure $d_{x,y}$, based on the assignment indices according to Paetkau et al. (1995).†**

| | Assignment based on the highest probability of an individual's genotype | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | To | | | | | | | |
| | Ariana | Aurelia | Fortis | Princesse | Sylvester | H66377 | KWS8123 | MK9907 |
| From | AFLP | | | | | | | |
| Ariana | 18 | 7 | 0 | 3 | 1 | 0 | 0 | 0 |
| Aurelia | 3 | 25 | 0 | 1 | 0 | 0 | 0 | 1 |
| Fortis | 1 | 1 | 27 | 0 | 0 | 0 | 0 | 0 |
| Princesse | 1 | 3 | 0 | 26 | 0 | 0 | 0 | 0 |
| Sylvester | 0 | 4 | 0 | 1 | 21 | 3 | 0 | 1 |
| H66377 | 0 | 1 | 0 | 0 | 0 | 27 | 0 | 2 |
| KWS8123 | 0 | 2 | 0 | 0 | 0 | 0 | 28 | 0 |
| MK9907 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 27 |
| | CAPS | | | | | | | |
| Ariana | 13 | 7 | 1 | 5 | 0 | 1 | 0 | 3 |
| Aurelia | 3 | 14 | 2 | 3 | 6 | 1 | 0 | 1 |
| Fortis | 3 | 1 | 17 | 1 | 3 | 0 | 1 | 4 |
| Princesse | 5 | 1 | 0 | 15 | 6 | 0 | 2 | 1 |
| Sylvester | 0 | 2 | 4 | 2 | 20 | 1 | 0 | 1 |
| H66377 | 0 | 0 | 0 | 3 | 0 | 25 | 0 | 2 |
| KWS8123 | 0 | 2 | 0 | 1 | 1 | 0 | 26 | 0 |
| MK9907 | 2 | 2 | 3 | 0 | 3 | 3 | 0 | 17 |
| | STMS | | | | | | | |
| Ariana | 24 | 5 | 0 | 0 | 0 | 0 | 1 | 0 |
| Aurelia | 1 | 27 | 0 | 0 | 0 | 0 | 1 | 1 |
| Fortis | 0 | 1 | 27 | 1 | 1 | 0 | 0 | 0 |
| Princesse | 1 | 0 | 1 | 27 | 0 | 0 | 1 | 0 |
| Sylvester | 0 | 0 | 2 | 0 | 28 | 0 | 0 | 0 |
| H66377 | 1 | 0 | 1 | 0 | 1 | 26 | 0 | 1 |
| KWS8123 | 0 | 0 | 0 | 0 | 0 | 0 | 30 | 0 |
| MK9907 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 30 |
| | Derived distance measure $d_{x,y}$ | | | | | | | |
| From | AFLP | | | | | | | |
| Ariana | 0 | | | | | | | |
| Aurelia | 4.88 | 0 | | | | | | |
| Fortis | 18.34 | 16.22 | 0 | | | | | |
| Princesse | 13.36 | 15.66 | 16.42 | 0 | | | | |
| Sylvester | 15.55 | 13.26 | 18.44 | 16.01 | 0 | | | |
| H66377 | 16.89 | 12.20 | 20.92 | 23.14 | 11.27 | 0 | | |
| KWS8123 | 26.28 | 25.06 | 29.16 | 28.39 | 27.47 | 28.66 | 0 | |
| MK9907 | 19.98 | 16.07 | 23.76 | 24.68 | 11.87 | 11.15 | 28.01 | 0 |
| | CAPS | | | | | | | |
| Ariana | 0 | | | | | | | |
| Aurelia | 1.04 | 0 | | | | | | |
| Fortis | 1.44 | 2.44 | 0 | | | | | |
| Princesse | 0.85 | 0.98 | 2.01 | 0 | | | | |
| Sylvester | 2.25 | 1.06 | 2.06 | 1.61 | 0 | | | |
| H66377 | 3.31 | 3.83 | 5.26 | 3.53 | 3.93 | 0 | | |
| KWS8123 | 3.57 | 3.57 | 4.42 | 3.67 | 5.04 | 6.82 | 0 | |
| MK9907 | 1.91 | 2.57 | 1.99 | 2.76 | 2.08 | 3.11 | 6.51 | 0 |
| | STMS | | | | | | | |
| Ariana | 0 | | | | | | | |
| Aurelia | 2.25 | 0 | | | | | | |
| Fortis | 6.89 | 5.19 | 0 | | | | | |
| Princesse | 3.97 | 4.71 | 5.82 | 0 | | | | |
| Sylvester | 5.74 | 5.29 | 4.24 | 5.91 | 0 | | | |
| H66377 | 6.94 | 6.54 | 3.75 | 6.43 | 3.89 | 0 | | |
| KWS8123 | 5.94 | 4.25 | 4.89 | 6.49 | 8.48 | 7.55 | 0 | |
| MK9907 | 11.06 | 8.02 | 6.94 | 8.93 | 6.10 | 5.74 | 7.90 | 0 |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; STMS, sequenced tagged microsatellite site.

**Table 7. Assignment of individual genotypes per variety and derived similarity by assignment values ($Sa_{x,y}$) based on the pairwise Jaccard similarity data (top 10 most similar assigned plants for each plant).†**

| | Assignment based on the pairwise Jaccard similarity data | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | To | | | | | | | |
| | Ariana | Aurelia | Fortis | H66377 | KWS8123 | MK9907 | Princesse | Sylvester |
| From | AFLP | | | | | | | |
| Ariana | 76 | 69 | 16 | 23 | 42 | 17 | 41 | 16 |
| Aurelia | 43 | 114 | 17 | 27 | 34 | 15 | 30 | 20 |
| Fortis | 11 | 39 | 148 | 27 | 26 | 12 | 22 | 15 |
| H66377 | 10 | 20 | 14 | 138 | 26 | 30 | 37 | 25 |
| KWS8123 | 21 | 28 | 14 | 26 | 169 | 12 | 22 | 8 |
| MK9907 | 14 | 27 | 22 | 51 | 23 | 117 | 30 | 16 |
| Princesse | 22 | 34 | 9 | 32 | 24 | 19 | 150 | 10 |
| Sylvester | 13 | 39 | 22 | 62 | 23 | 28 | 32 | 81 |
| | CAPS | | | | | | | |
| Ariana | 62 | 56 | 38 | 21 | 30 | 21 | 43 | 19 |
| Aurelia | 45 | 88 | 16 | 15 | 37 | 14 | 46 | 39 |
| Fortis | 33 | 37 | 101 | 5 | 31 | 32 | 31 | 30 |
| H66377 | 30 | 26 | 5 | 145 | 15 | 26 | 30 | 23 |
| KWS8123 | 14 | 24 | 24 | 11 | 195 | 1 | 24 | 7 |
| MK9907 | 26 | 29 | 34 | 46 | 12 | 92 | 25 | 36 |
| Princesse | 29 | 48 | 25 | 13 | 23 | 14 | 105 | 33 |
| Sylvester | 21 | 63 | 24 | 18 | 13 | 26 | 58 | 77 |
| | STMS | | | | | | | |
| Ariana | 137 | 88 | 6 | 10 | 14 | | 33 | 12 |
| Aurelia | 62 | 149 | 20 | 12 | 28 | 4 | 16 | 9 |
| Fortis | 3 | 27 | 195 | 24 | 16 | 1 | 14 | 20 |
| H66377 | 4 | 11 | 28 | 194 | 8 | 15 | 13 | 27 |
| KWS8123 | 4 | 9 | 6 | 1 | 275 | 2 | 2 | 1 |
| MK9907 | | 1 | 6 | 19 | 8 | 255 | 2 | 9 |
| Princesse | 45 | 32 | 18 | 10 | 16 | 4 | 161 | 14 |
| Sylvester | 12 | 16 | 28 | 29 | 16 | 14 | 15 | 170 |
| | Similarity by assignment $Sa_{x,y}$ | | | | | | | |
| From | AFLP | | | | | | | |
| Ariana | 0.253 | | | | | | | |
| Aurelia | 0.187 | 0.380 | | | | | | |
| Fortis | 0.045 | 0.093 | 0.493 | | | | | |
| H66377 | 0.055 | 0.078 | 0.068 | 0.460 | | | | |
| KWS8123 | 0.105 | 0.103 | 0.067 | 0.087 | 0.563 | | | |
| MK9907 | 0.052 | 0.070 | 0.057 | 0.135 | 0.058 | 0.390 | | |
| Princesse | 0.105 | 0.107 | 0.052 | 0.115 | 0.077 | 0.082 | 0.500 | |
| Sylvester | 0.048 | 0.098 | 0.062 | 0.145 | 0.052 | 0.073 | 0.070 | 0.270 |
| | CAPS | | | | | | | |
| Ariana | 0.214 | | | | | | | |
| Aurelia | 0.172 | 0.293 | | | | | | |
| Fortis | 0.121 | 0.088 | 0.337 | | | | | |
| H66377 | 0.086 | 0.068 | 0.017 | 0.483 | | | | |
| KWS8123 | 0.075 | 0.102 | 0.092 | 0.043 | 0.650 | | | |
| MK9907 | 0.080 | 0.072 | 0.110 | 0.120 | 0.022 | 0.307 | | |
| Princesse | 0.124 | 0.159 | 0.095 | 0.072 | 0.080 | 0.066 | 0.362 | |
| Sylvester | 0.068 | 0.170 | 0.090 | 0.068 | 0.033 | 0.103 | 0.154 | 0.257 |
| | STMS | | | | | | | |
| Ariana | 0.457 | | | | | | | |
| Aurelia | 0.250 | 0.497 | | | | | | |
| Fortis | 0.015 | 0.078 | 0.650 | | | | | |
| H66377 | 0.023 | 0.038 | 0.087 | 0.647 | | | | |
| KWS8123 | 0.030 | 0.062 | 0.037 | 0.015 | 0.917 | | | |
| MK9907 | 0.000 | 0.008 | 0.012 | 0.057 | 0.017 | 0.850 | | |
| Princesse | 0.130 | 0.080 | 0.053 | 0.038 | 0.030 | 0.010 | 0.537 | |
| Sylvester | 0.040 | 0.042 | 0.080 | 0.093 | 0.028 | 0.038 | 0.048 | 0.567 |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; $Sa_{x,y}$, similarity by assignment; STMS, sequenced tagged microsatellite site.

the standard errors on them. However, the Mantel's analyses (Table 8), indicate rather low correlation coefficients between distance matrices.

When data reduction techniques, such as bidimensional scaling and clustering, are used, an acceptable level of significance was reached for only a limited group of clusters: two clusters in the case of AFLP data and one cluster with the STMS data set (Fig. 1). Taking into account that this study only included eight varieties as an input for clustering, but fingerprinted 30 plants per variety, the results confirmed across the three marker techniques using this approach are rather poor.

AMOVA quantified the within populations variation as ranging from 84% for STMS to 92% for AFLP markers. However, the population pairwise *F*-statistics matrices generated by the AMOVA routine in Arlequin did not outperform the simple $D_{Nei}$ or $D_{Eucl}$ calculated from the direct marker frequency data.

A final set of analyses was based on assignment tests offering the advantage of making use of the multilocus genotype of each individual. The method of Paetkau et al. (1995) yielded assignments that were too unambiguous (Table 6). The correlation between the various marker techniques was high ($r > 0.95$) but, unfortunately, in its derived output much of this equivalence disappeared as both the level and ranges of the derived distance $\mathbf{d}_{x,y}$ were inconsistent, shown by low Mantel's statistics (Tables 6 and 8).

An alternative assignment test based on the pairwise similarity data for individual plants, as introduced by De Riek et al. (2001) for AFLP data, yielded assignments with a more dispersed allocation pattern among varieties than the method by Paetkau et al. (1995) also showing a higher consistency across the marker techniques used (Tables 7 and 8). A good allocation to the proper variety was obtained, together with a reliable allocation pattern among the other varieties. These two aspects represent the variation within a variety and the distance to other

varieties. Although asymmetric in its output, we have shown that it can easily be transformed into an average similarity measure ($Sa_{x,y}$). This index based on assignment tests can be considered as a new genetic distance measure with interesting properties:

1. The assignment tests revealed differences among varieties by the allocation pattern among the other varieties. In particular, this is relatively independent of the marker technique used.
2. The assignments based on the same marker technique but using a different similarity measure were in good agreement.
3. The scales and scopes for the distances measured may be values relatively insensitive to the degree of polymorphism of the marker technique used.
4. The levels of distinction between varieties obtained were much higher (i.e., a higher number of plants is assigned correctly).
5. The measure produced comparable results when calculated using different numbers of best assigned plants (from the top three to the top 30 highest matches for each plant sampled).

As a similarity by assignment is by its nature related to the composition of the dataset (the varieties it is compared with in the assignment test) and to the assignment thresholds imposed (the number of most related plants each individual is attributed to) it should not be treated as an absolute estimate of genetic distance. However, compared to the other analysis techniques in this study, it accomplishes a superior distinction among genetically diverse varieties in a complex cross-pollinating, polyploid crop such as sugar beet. To our knowledge, this is the first time assignment methods were used for variety identification.

**Table 8. Comparison of marker techniques for different summarizing methods by standardized Mantel's statistics *g* (also expressed as a correlation coefficient *r*) between similarity matrices. *P* value was estimated by 1000 random iterations (out of 40,320 possible permutations).†**

| | Avg. Jaccard similarity | Standard Nei genetic distances | $d_{x,y}$ | $Sa_{x,y}$ |
|---|---|---|---|---|
| AFLP vs. CAPS | $g$ = 1.9326 | $g$ = 3.3067 | $g$ = 2.0274 | $g$ = 4.0771 |
| | $r$ = 0.407 | $r$ = 0.6687 | $r$ = 0.6342 | $r$ = 0.8391 |
| | $p$ = 0.3270 | $p$ = 0.0340 | $p$ = 0.0400 | $p$ = 0.0060 |
| AFLP vs. STMS | $g$ = 3.7694 | $g$ = 3.0400 | $g$ = 1.8994 | $g$ = 4.411 |
| | $r$ = 0.6824 | $r$ = 0.5881 | $r$ = 0.4376 | $r$ = 0.9555 |
| | $p$ = 0.0130 | $p$ = 0.1010 | $p$ = 0.0900 | $p$ = 0.0020 |
| CAPS vs. STMS | $g$ = 3.8176 | $g$ = 3.0979 | $g$ = 0.9760 | $g$ = 4.0691 |
| | $r$ = 0.6241 | $r$ = 0.6069 | $r$ = 0.2302 | $r$ = 0.8338 |
| | $p$ = 0.0080 | $p$ = 0.0490 | $p$ = 0.2100 | $p$ = 0.0144 |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; STMS, sequenced tagged microsatellite site.

**Table 9: Comparison of summarizing methods for a marker technique by standardized Mantel's statistics *g* (also expressed as a correlation coefficient *r*) between similarity matrices. *P* value was estimated by 1000 random iterations (out of 40,320 possible permutations).†**

| | AFLP | CAPS | STMS |
|---|---|---|---|
| Avg. Jaccard similarity vs. Nei genetic distances | $g$ = 3.234 | $g$ = 3.9472 | $g$ = 4.8875 |
| | $r$ = 0.6376 | $r$ = 0.9039 | $r$ = 0.9291 |
| | $p$ = 0.1131 | $p$ = 0.0010 | $p$ = 0.0010 |
| Avg. Jaccard similarity vs. $Sa_{x,y}$ | $g$ = 4.6635 | $g$ = 4.2963 | $g$ = 4.0368 |
| | $r$ = 0.8259 | $r$ = 0.7463 | $r$ = 0.8513 |
| | $p$ = 0.0010 | $p$ = 0.0010 | $p$ = 0.0010 |
| Nei genetic distances vs. $Sa_{x,y}$ | $g$ = 4.1989 | $g$ = 4.6896 | $g$ = 4.1486 |
| | $r$ = 0.7697 | $r$ = 0.8264 | $r$ = 0.7863 |
| | $p$ = 0.0030 | $p$ = 0.0010 | $p$ = 0.0010 |

†AFLP, amplified fragment length polymorphism; CAPS, cleaved amplified polymorphic site; STMS, sequenced tagged microsatellite site.

## Acknowledgments

The authors wish to thank the KBIVB-IRBAB for the good cooperation and Laurence Desmet, Veerle Buysens, and Veerle Cools for their efforts in doing DNA preparations and microsatellite analysis. The research was funded by the former Belgian Ministry of Small Enterprises, Traders and Agriculture DG4 and by the Netherlands' Ministry of Agriculture, Nature and Food Safety through the Method Development scheme of CGN.

## References

Barnes, S., G. Massaro, M. Lefèbvre, M. Kuiper, and E. Verstege. 1996. A combined RFLP and AFLP genetic map for sugar beet. p. 555–560. *In* Proc. 59th Congr. of the IIRB, Brussels. February 1996. Int. Inst. for Beet Res., Brussels.

Barzen, E., W. Mechelke, E. Ritter, E. Schulte-Kappert, and F. Salamini. 1995. An extended map of the sugar beet genome containing RFLP and RAPD loci. Theor. Appl. Genet. 90:189–193.

Becher, S.A., K. Steinmetz, K. Weising, S. Boury, D. Peltier, J.-P. Renou, G. Kahl, and K. Wolff. 2000. Microsatellites for cultivar identification in *Pelargonium*. Theor. Appl. Genet. 101:643–651.

Bredemeijer, G.M.M., R.J. Cooke, M.W. Ganal, R. Peeters, P. Isaac, Y. Noordijk, S. Rendell, J. Jackson, M.S. Röder, K. Wendehake, M. Dijcks, M. Amelaine, R. Wickaert, L. Bertrand, and B. Vosman. 2002. Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor. Appl. Genet. 105:1019–1026.

De Riek, J., E. Calsyn, I. Everaert, E. Van Bockstaele, and M. De Loose. 2001. AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet. 103:1254–1265.

Esselink, D., H. Nybom, and B. Vosman. 2004. Assignment of allelic configuration in polyploids using the MAC–PR (microsatellite DNA allele counting–peak ratios) method. Theor. Appl. Genet. 109:402–408.

Esselink, D., M.J.M. Smulders, and B. Vosman. 2003. Identification of cut-rose (*Rosa hybrida*) and rootstock varieties using robust sequence tagged microsatellite markers. Theor. Appl. Genet. 106:277–286.

Excoffier, L., P. Smouse, and J. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics 131:479–491.

Guinand, B., A. Topchy, K.S. Page, M.K. Burnham-Curtis, W.F. Punch, and K.T. Scribner. 2002. Comparisons of likelihood and machine learning methods of individual classification. J. Hered. 93:260–269.

Jones, C.J., K.J. Edwards, S. Castiglione, M.O. Winfield, F. Sala, C. Van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vazquez, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol. Breed. 3:381–390.

Jung, C., K. Pillen, L. Frese, S. Fähr, and A.E. Melchinger. 1993. Phylogenetic relationships between cultivated and wild species of the genus *Beta* revealed by DNA "fingerprinting." Theor. Appl. Genet. 86:449–457.

Manel, S., O.E. Gaggiotti, and R.S. Waples. 2005. Assignment methods: Matching biological questions with appropriate techniques. Trends Ecol. Evol. 20(3):136–142.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Nybom, H., D. Esselink, G. Werlemark, and B. Vosman. 2004. Microsatellite DNA marker inheritance indicates preferential pairing between two highly homologous genomes in polyploid and hemisexual dogroses, *Rosa* L. sect. *Caninae*. Heredity 92:139–150.

Paetkau, D., W. Calvert, I. Sterling, and C. Strobeck. 1995. Microsatellite analysis of population structure in Canadian polar bears. Mol. Ecol. 4:347–354.

Paetkau, D., L.P. Waits, P.L. Clarkson, L. Craighead, and C. Strobeck. 1997. An empirical evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. Genetics 147:1943–1957.

Paran, I., and R.W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985–993.

Pillen, K., G. Steinrücken, G. Wricke, R.G. Herrmann, and C. Jung. 1992. A linkage map of sugar beet (*Beta vulgaris* L.). Theor. Appl. Genet. 84:129–135.

Rae, S.J., C. Aldam, I. Dominguez, M. Hoebrechts, S.R. Barnes, and K.J. Edwards. 2000. Development and incorporation of microsatellite markers into the linkage map of sugar beet (*Beta vulgaris* spp.). Theor. Appl. Genet. 100:1240–1248.

Röder, M.S., K. Wendehake, V. Korzun, G. Bredemeijer, D. Laborie, L. Bertrand, P. Isaac, S. Rendell, J. Jackson, R.J. Cooke, B. Vosman, and M.W. Ganal. 2002. Construction and analysis of a microsatellite-based database of European wheat varieties. Theor. Appl. Genet. 106:67–73.

Schneider, K., D.C. Borchardt, R. Schäfer-Pergl, N. Nagl, C. Glass, A. Jeppsson, C. Gebhardt, and F. Salamini. 1999. PCR-based cloning and segregation analysis of functional gene homologues in *Beta vulgaris*. Mol. Gen. Genet. 262:515–524.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, Univ. of Geneva, Switzerland.

Schondelmaier, J., G. Steinrücken, and G. Jung. 1996. Integration of AFLP markers into a linkage map of sugar beet (*Beta vulgaris* L.). Plant Breed. 115:231–237.

Schumacher, K., J. Schondelmaier, E. Barzen, G. Steinrücken, D. Borchardt, W.E. Weber, C. Jung, and F. Salamini. 1997. Combining different linkage maps in sugar beet (*Beta vulgaris* L.) to make one map. Plant Breed. 116:23–38.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of Matrix correspondence. Syst. Zool. 35:627–632.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Fijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP™: A new technique for DNA fingerprinting. Nucleic Acids Res. 21:4407–4414.