Theor Appl Genet (1999) 99:1155–1165 © Springer-Verlag 1999

J. De Riek · J. Dendauw · M. Mertens · M. De Loose
J. Heursel · E. Van Blockstaele

Validation of criteria for the selection of AFLP markers to assess the genetic variation of a breeders' collection of evergreen azaleas

Received: 3 September 1998 / Accepted: 25 March 1999

Abstract Fluorescent AFLP and automated data analysis were employed to assess the genetic conformity within a breeders' collection of evergreen azaleas. The study included 75 genotypes of Belgian pot azaleas (*Rhododendron simsii* Planch. hybrids), Kurume and Hirado azaleas and wild ancestor species from the *Tsutsusi* subgenus. Fluorescent detection and addition of an internal size standard to each lane enabled the automated scoring of each fragment arising from a single AFLP primer combination (PC). The use of three PCs generated an initial data set with a total of 648 fragments ranging from 70 bp to 450 bp. Different marker selection thresholds for average fluorescent signal intensity and marker frequency were used to create eight extra restricted data subsets. Pairwise plant genetic similarity was calculated for the nine data sets using the Simple Matching coefficient (symmetrical, including double-zeros) and Jaccard coefficient (asymmetrical, excluding double zeros). The averages, the ranges and the correlation to one other (Mantel analysis) were compared for the obtained similarity matrices. This revealed the sensitivity of ordinations obtained by both similarity coefficients for the presence of weak or intensive markers or for the degree of polymorphism of the markers. For 34 cultivars, pedigree information (at maximum to the fifth ancestor generation) was available. Genetic similarity by descent (kinship coefficient) was turned into a genetic distance and correlated to the genetic conformity, as revealed by the different selections of AFLP markers (Mantel analysis). Use of a Simple Matching coefficient with no or moderate selection to signal intensity and excluding rare and abundant markers gave the best correlation with pedigree. Finally, the ordination of the studied genotypes by means of dendrograms and principal co-ordinate analysis was confronted with known or accepted relationships based on geographical origin, parentage and morphological characters. Genotypes could be assigned to three distinct groups: pot azaleas, Kurume azaleas and Hirado azaleas. Wild ancestor species appeared to be more related to the Japanese azaleas. Intermediate cultivars could be typified as crossings with Kurume or Hirado azaleas or with wild species.

Key words Similarity Coefficients · Pedigree · Kinship · *Rhododendron simsii* · Mantel's test

Communicated by F. Salamini

J. De Riek (✉) · J. Dendauw · M. Mertens · M. De Loose
E. Van Bockstaele
Department for Plant Genetics and Breeding – CLO Gent, Caritasstraat 21, 9090 Melle, Belgium
e-mail: jderiek@clo.fgov.be, Fax: 32-9-2525075

J. Heursel
Department for Crop husbandry and Ecophysiology – CLO Gent, Burg. Van Gansberghelaan 109, 9820 Merelbeke, Belgium

E. Van Bockstaele
Department of Plant Production,
Faculty of Agricultural and Applied Biological Sciences – RUG, Coupure links 653, 9000 Gent, Belgium

Introduction

Traditionally, plant breeders have always evaluated plant origins and pedigrees as an extra indication of the characteristics that were acquired by a new genotype. With the advent of molecular marker techniques, new tools that have the capacity to reveal (dis)similarities in the genome of related plant species or cultivars directly became available. Polymerase chain reaction (PCR)-based multilocus fingerprinting techniques for which no prior DNA sequence information is required, like randomly amplified polymorphic DNA (RAPD) or amplified fragment length polymorphism (AFLP), have in particular facilitated this considerably. The AFLP technique, based on the selective PCR amplification of restriction fragments from a total restriction digest of genomic DNA (Vos et al. 1995), often yields several DNA polymorphisms amplified by one primer combination. Because of its reproducibility and the high data output per reaction, AFLP is being used more and more as a fast and reliable

Table 1 Variety name, type of hybrid and parentage of the genotypes analysed with AFLP. The 34 plants included in the kinship analysis are preceded by a asterisk (*)

Number[a]	Variety	Type	Parentage[b, c]
5	Ademurasaki	Hirado	
6	*Heiwa-no-hikari	Hirado	
7	Concinna	Hirado	
8	Omurasaki	Hirado	
9	*Kirin	Kurume	
11	*Rex	Kurume	Hinomayo×Kirin
12	*Directeur Van Slycken	Kurume	Excelsior×Rex
15	Cheops	Noria-azalea	
1	Gilbert Mullie	Kurume	
2	Blue Danube	Kurume	
3	Palestrina	Kurume	
4	Stewartstownian	Kurume	
20	*Vervaeneana	Pot azalea	Comte Charles de Kerchove×Koenigin der Weisse
21	*Apollo	Pot azalea	
22	*Avenir	Pot azalea	John T.D. Lewelyn×Madame Royer
23	*Professor Wolters	Pot azalea	
24	*Madame Petrick	Pot azalea	
25	*Laura Ashley	Pot azalea	Mevrouw Jozef Heursel[c]×Herrmann's Superba[c]
26	Niobe	Pot azalea	
27	*Adventsglocke	Pot azalea	Paul Schaeme×Fritz Sander
28	Etoile de Belgique	Pot azalea	
29	*Pink Dream	Pot azalea	
30	*Comtesse de Kerchove	Pot azalea	
31	*Paul Schaeme	Pot azalea	Wilhelm Scheurer×Deutsche Perle
32	*Roi Leopold	Pot azalea	
33	*Hexe	Pot azalea	Herzog Adolf von Nassau×Amoenum
34	*Euratom	Pot azalea	Apollo×Hexe
35	*Knut Erwén	Pot azalea	Paircross 4[c]×Paircross 5[c]
36	*Reinhold Ambrosius	Pot azalea	Hexe×?
37	*Eclaireur	Pot azalea	Eggebrechtii×Etoile de Noèl
38	*Elsa Kaerger	Pot azalea	
39	*Ambrosiana	Pot azalea	
40	Leopold-Astrid	Pot azalea	
41	Perle de Swijnaerde	Pot azalea	
42	*Tempérance	Pot azalea	
43	*Coelestine	Pot azalea	
44	Dame Blanche	Pot azalea	bud sport Roi Leopold
45	*Glaser Nr. 10	Pot azalea	Roi Leopold×Coelestine
46	*Madame P.B. Van Acker	Pot azalea	Madame Petrick×Spit Fire
47	Tamira	Pot azalea	
48	Gloire de Saint Georges	Pot azalea	
49	Reinhild	Pot azalea	
50	*Erich Danneberg	Pot azalea	Madame P.B. Van Acker×Paul Schaeme
51	*Hellmut Vogel	Pot azalea	Ambrosiana×Erich Danneberg
52	*Mevrouw Marcel Vanbelle	Pot azalea	Hellmut Vogel×Mistral
53	Doctor Heimann	Pot azalea	
54	*Friedhelm Scherrer	Pot azalea	Ambrosiana×Paircross 6[c]
55	*Mevrouw André Heungens	Pot azalea	Comtesse de Kerchove×Paircross 1[c]
56	*Rosali	Pot azalea	Madame Petrick×Pink Dream
57	*Kingfisher	Pot azalea	
58	*Flamenco	Pot azalea	Violacea×Reinhold Ambrosius
59	Ostalett	Pot azalea	
60	*Bertina	Pot azalea	Hermann Klussman×Zaailing 18
61	*Schnee	Pot azalea	
62	*Mistral	Pot azalea	Osaka[c]×Heiwa-no-hikari
63	*Lara	Pot azalea	Memoria Sander×Heiwa-no-hikari
64	Aline	Pot azalea	
65	Kosmos	Pot azalea	
66	Dorothy Gish	Pot azalea	
67	Heidi	Pot azalea	bud sport Rosali
68	*James Belton	Pot azalea	Vervaeneana×Schryderi(i)
69	*Schuman	Pot azalea	Paircross 2[c]×Paircross 3[c]
70	*Otto	Pot azalea	Friedhelm Scherrer×?[b]
71	Heide Hanisch	Pot azalea	
74	Nordlicht bloei	Pot azalea	bud sport Hellmut Vogel
75	Nordlicht groei	Pot azalea	bud sport Hellmut Vogel

Table 1 (Continued)

Number[a]	Variety	Type	Parentage[b, c]
13	*R. nakaharai*	Species	
18	*R. simsii*	Species	
19	*R. tashiroi*	Species	
73	*R. indicum*	Species	
10	*R. kiusianum*	Species	
14	*R. noriakianum*	Species	
16	*R. mucronatum*	Species	
17	*R. scabrum*	Species	
72	*R. scabrum*	Species	

[a] Numbers refer to the labels of Fig. 3
[b] Parents indicated with ? are unknown; if no pedigree data are given, both parents are unknown
[c] Pedigree information for plants not analysed with AFLP but included in the kinship analysis: Herrmann's Superba ('Madame P.B. Van Acker'×'Paul Schaeme'); Mevrouw Jozef Heursel ('Violacea'×'Hexe'); Osaka ('Madame Petrick'×'Pink Dream'); Paircross 1 ('Reinhold Ambrosius'×'Madame Petrick'); Paircross 2 ('Osaka'×'Pink Dream'); Paircross 3 ('Osaka'×'Pink Dream'); Paircross 4 ('Hermann Seidel'×'Professor Wolters'); Paircross 5 ('Paul Schaeme'×'Apollo'; Paircross 6 ('Coelestine'×?)

molecular marker technique for assessing genetic conformity for both the study and conservation of natural diversity and for the characterisation of breeding gene pools, e.g. in soybean (Maughan et al. 1996), lettuce (Hill et al. 1996), tea (Paul et al. 1997), sunflower (Hongtrakul et al. 1997) and maize (Ajmone Marsan et al. 1998).

Fluorescent fragment labelling applicable on automated sequencers can add an extra advantage by providing a radioactive-free system. Moreover, if a combination of fluorescent dyes with distinct emission spectra can be multiplexed in one lane, higher throughput per gel and the application of an internal size standard in each lane can be achieved. Such systems can enable automated size determination and scoring of the fragments, resulting in a substantial time-saving mainly due to the reduced post-electrophoresis steps for band scoring. This gain is most evident when one has to score a high number of polymorphic fragments in a large set of lowly related plants.

Due to the large number of markers AFLP might yield, one needs consistent criteria to select the more informative markers. The main objective of the study presented here was to validate the reliability of well-defined marker selection parameters to handle AFLP data output. Instead of selecting a subset of polymorphic bands with the eye, we applied thresholds for marker signal intensity and marker frequencies over the data set. Although their definition was facilitated by fluorescent AFLP on a DNA sequencer, these parameters can be readily applied to any band scoring system capable of quantifying band signal intensity.

A first validation focussed on the sensitivity of the resulting ordination for different marker selection parameter settings: discarding weak bands or imposing the degree of polymorphism of the markers included. Two different resemblance measures, the Simple Matching and the Jaccard coefficient, were evaluated as both represent a different group of measures for binary data: the first being a symmetrical measure rewarding the double absence of a marker in a pair of plants as a fact contributing to similarity, the second an asymmetrical measure neglecting double absences. A second validation, although less statistically quantifiable, could be obtained from the genetic diversity study itself, examining a well characterised breeders' collection of 75 genotypes of Belgian pot azalea (*Rhododendron simsii* hybrids) and other related evergreen azaleas belonging to the *Tsutsusi* subgenus (Chamberlain and Rae 1990). The genetic conformity revealed by AFLP was compared with the known pedigree (34 genotypes), the geographical origin or the supposed relationship based on the morphology of the genotypes studied.

Materials and methods

Plant material

In total 75 azalea genotypes were analysed (Table 1). They were chosen from three distinguishable subgroups within the evergreen azaleas: pot azaleas (*Rhododendron simsii* hybrids; 55 genotypes), Hirado azaleas (*R. scabrum* hybrids; 4 genotypes) and Kurume azaleas (hybrids of *R. kiusianum* var *'kiusianum'* and var *'sataense'* and *R. kaempferi*; 7 genotypes). Apart from *R. simsii*, at least three other species *R. indicum*, *R. mucronatum* and *R. scabrum* are accepted to be the ancestors of pot azaleas (Galle 1987). Therefore, also nine individual genotypes of different *Tsutsusi* species were also included. Within the pot azaleas, some varieties obtained from bud sporting were included (two 'Nordlicht' type of bud sports from 'Hellmut Vogel', one altered in flower colour, the other also in growth behaviour; 'Dame Blanche' bud sport from 'Roi Leopold' and 'Heidi' bud sport from 'Rosali').

DNA isolation

Young leaf material (approximately 1 g fresh weight) was harvested from 2-year-old potted plants maintained in the greenhouse. At harvest plant material was immediately immersed in liquid nitrogen and subsequently lyophilised for 48 h. The dry material was vacuum-packed for storage at –20°C until DNA extraction. Stored material was ground using a Culatti mechanical mill. Different DNA isolation protocols based on Dellaporte et al. (1983), Doyle and Doyle (1987), Greenwood et al. (1989), Kobayashi et al. (1995) and Weising et al. (1991) were tested with the final result being the following procedure. To 25–75 mg lyophilised ground tissue, 10 ml of a 50 m*M* Tris-HCl (pH 8) containing 5 m*M*

EDTA, 350 m*M* sorbitol, 0.1% β-mercaptoethanol, 10% PEG (MW 8000) and 0.001% BSA were added. This mixture was homogenised for 30 s by inverting the tubes and then filtered through Miracloth (Calbiochem) to remove coarse cell debris. The nuclei fraction was pelleted by centrifugation (2500 *g*, 5 min, 4°C). To the nuclei fraction, 1 ml CTAB extraction buffer (100 m*M* Tris-HCl pH 8, containing 2% CTAB, 20 m*M* EDTA, 1.4 *M* NaCl, 0.5 m*M* $Na_2S_2O_5$, 0.4% β-mercaptoethanol and 1% PVP MW 40000) and RNase (10 U) were added. Samples were incubated for 40 min. at 65°C. Afterwards, samples were homogenised with 1 ml chloroform/isoamylalcohol (24/1) and centrifuged for 15 min. at 10 000 *g*. The supernatant was transferred to a fresh tube and the DNA precipitated with 1 ml of ice cold (–20°C) isopropanol. After centrifugation (5000 *g*; 15 min), the pellet was washed with EtOH (76%) - 0.2 *M* NaOAc, dried and dissolved in water. DNA concentration and quality were constantly checked relative to a standard series of lambda-DNA after electrophoresis on a 1.5% TAE buffered agarose gel.

AFLP reactions and PAGE

AFLP was performed using the commercially available kit from Perkin-Elmer Biosystems for fluorescent fragment detection (Perkin-Elmer 1995). *Eco*RI and *Mse*I were used for DNA digestion. Adapter ligation, preselective and selective amplification were as in the manufacturers' protocols. Selective amplification was done using fluorescent-labelled EcoRI-MseI primer combinations with 6 selective bases. The primer combinations used were: *Eco*RI-ACT/*Mse*I-CTA (PC1), *Eco*RI-ACT/*Mse*I-CAT (PC2), *Eco*RI-AAG/*Mse*I-CTA (PC3). PCR amplifications were performed using a Perkin-Elmer 9600. AFLP fragments were separated by PAGE on a ABI Prism 377 DNA Sequencer on 36-cm gels using 4.25% denaturing polyacrylamide (4.25% acrylamide/bisacrylamide 19/1, 6 *M* urea in 1×TBE). GS-500-ROX labelled size standard (Perkin Elmer) was loaded on each lane in order to facilitate the automatic analysis of the gel and the sizing of the fragments.

Band scoring and statistical analyses

During a run on the ABI 377, the fluorescent signal in each lane is being recorded continuously. For each fluorescent dye electropherograms are extracted that are comparable to the densitometric curves obtained after scanning of an autoradiogram. GENESCAN 2.1 was used to estimate detection time, signal peak height and surface for each fragment. Sizing of the fragments was performed by the Genescan software module by interpolation to the internal lane standard according to the Local Southern algorithm. Only the fragments between 70 bp and 450 bp were used for scoring. Resolution of the gel system (i.e. the capacity to separate 2 subsequent bands in one lane) is 1 bp. However, due to lane-to-lane variation and differences in the interpolation of the standard the scoring of the same band position between different lanes varied within 1 bp. After export of the Genescan data to Microsoft Access, these variations in fragment size were assigned to the corresponding categories and a scoring table (1/0) was generated. Using Access queries, we set different filters towards signal peak height and marker frequency. Calculation of similarity coefficients, construction of dendrograms (UPGMA), Mantel analysis and principal co-ordinates analysis were performed by the modules SIMIL, CLUSTER, MANTEL and PCOORD of the "R package" (Legendre and Vaudor 1991). Mantel analysis computed standardised Mantel's statistics between two resemblance matrices. The significance of the statistic was evaluated by permutations (1000×) and expressed as a probability (Smouse et al. 1986). Pedigree analysis was performed by defining an "unrelated" ancestor population first, i.e. by tracing back the pedigree until all parents were unknown. For all plants in the pedigree, kinship coefficients (r) were calculated using KIN (Tinker and Mather 1993) and were turned into a distance (1–r).

Results and discussion

Effectiveness of the primer combinations for *R. simsii* hybrids

For azalea, only primer combinations with 6 selective bases that had been tested previously were used (De Riek et al. 1997). They generate approximately 70–90 AFLP fragments per reaction. Intense bands, indicative of repetitive fragments, were not observed in the sets used. PC1 generated 219 unique fragments when all plant reactions were scored; PC2, 198 and PC3, 231. A visual check on the electropherograms did not show any polymorphisms between the initial variety 'Hellmut Vogel' and its two bud sports, or between 'Heidi' and 'Rosali' or 'Dame Blanche' and 'Roi Leopold'. From a more extensive study, similar results were obtained using several PC and different bud sports of 'Hellmut Vogel' (De Riek et al. 1997). The parentage of plants with known genitors was confirmed on a visual check. No bands were present in the offspring that could not be traced back in the putative parents.

Sensitivity of the ordination for different marker selection parameter settings

Scoring AFLP gels can be an error-prone task due to the fact that in larger sets of unrelated plants large ranges of fragments differing only 1 bp from each other might occur. Also, there might be a tendency to discard weak, infrequent and/or constant bands. When discarding constant bands, for example, a quite obvious effect is possible: pairwise similarities between plants will decrease. However, whether the plant order in classifications will be changed is mostly unclear. Another uncertainty to deal with is the impact of a change in the quality of AFLP gels (variations in radioactive or fluorescent signal, different dyes or reactions) on the final results. When carrying out automated scoring, these kind of risks might be estimated as being higher because of the lack of a check by eye. Therefore, consistent marker selection criteria are required.

Selection for the average fluorescent signal intensity of a marker was performed using three thresholds: 50, 120 and 300 units. Before marker selection, all data were collected under a Genescan baseline setting of 50 units (as recommended by the manufacturer) and attributed to a marker category. The average signal intensity of all bands scored as the same marker was then used as a criterion. If the Genescan baseline setting is raised beforehand, it is likely to discard bands that are present but have accidentally a low signal. This is mainly due to the fact that during extraction of the electropherogram only part of the total band signal (three data channels adjacent to a lane tracking line) is being recorded by the software. Peak height and surface can thus be influenced by any shift of the tracking line from the centre of a band.

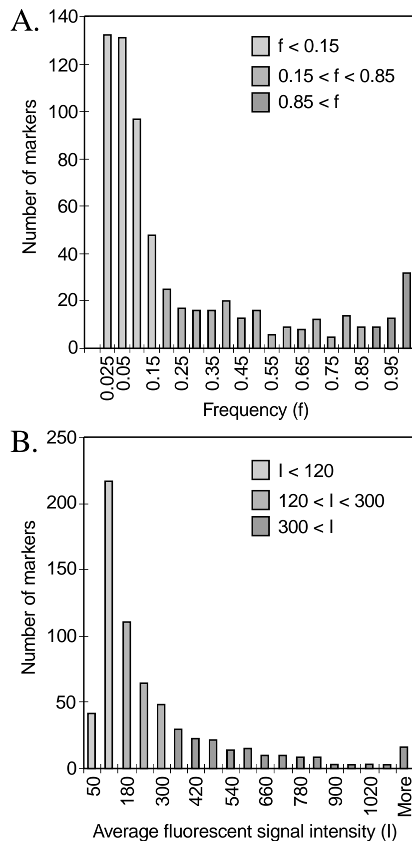


Fig. 1A, B AFLP marker distributions based on: **A** the frequency they appeared in the set of genotypes analysed and **B** their average fluorescent signal intensity in the gels. Different marker selection thresholds are indicated

Selection to marker frequency was performed in two ways: (1) exclusion of all rare markers (f<0.15) and (2) exclusion of all rare and abundant markers (0.15<f <0.85). All fingerprinted plants were taken into account for the frequency calculation. The effect of these thresholds can be deduced from the marker distributions as shown in Fig. 1; Table 2 summarises the number of markers included in the analysis for each of the nine cases. The number of markers decreases drastically when frequency thresholds are applied. Merely rare markers are lost. Without selection to signal intensity, only 50 markers appeared to be abundant. This value was approximately the same under higher signal intensity thresholds (48 abundant markers at 120 units; 42 at 300 units). Therefore, when applying signal intensity selection, one is enriching for abundant markers. The question if abundant markers were also the more homozygous ones could not be evaluated because the Genescan data extraction mode as used (as mentioned before) did not allow the reproducible co-dominant scoring of AFLP fragments based on peak height or surface.

Each of the nine parameter settings (Table 2) was used as an input for the calculation of pair-wise similarity matrices. The use of the Simple Matching (symmetrical) and the Jaccard coefficient (asymmetrical measure) were evaluated. For AFLPs, different alleles in general cannot be distinguished as separate bands, and there might be some distrust in using a symmetrical coefficient. The absence of an AFLP marker in 2 plants can be caused by completely different genetic mutations and should not be considered equally. On the other hand, until it has been checked by segregation in mapping populations or sequencing of the fragments, one is not completely sure that a band present at the same gel position is identical.

Table 2 Number of AFLP markers included in the analyses using different settings for selection for marker frequency and average marker signal intensity

Marker frequency selection	Selection for average fluorescent signal intensity (units)		
	>50	>120	>300
None	648	389	166
f>0.15	245	179	102
0.15<f<0.85	195	131	60

To evaluate the sensitivity of the final ordination for different marker selection parameter settings, we used two approaches: (1) detection of differences on average and range of the similarity matrices, and (2) statistical tests (Mantel test) on the correlation between pairs of matrices (the construction of dendrograms for all cases and the detection of differences in these final ordinations were not considered to be feasible using 75 plants and 18 matrices). Since the results could have been altered by the reduction in the number of data sets when applying different marker selection criteria, the effect of a random reduction of the data set was evaluated first.

By resampling randomly an equal number of markers as for the eight restricted cases, five data sets per case were created. In general, averages and ranges of the similarity matrices for the resampled data sets were close to the values of the initial data set with 648 markers (results not shown). Using the Simple Matching coefficient on the initial data set, the average similarity was 0.82, ranging from a minimum of 0.69 to a maximum of 0.94 (Table 3). The average similarity for all resampled similarity matrices varied between 0.810 and 0.825, a very close range around the value for the initial data set. The minimum similarity varied between 0.55 (resampling of 60 markers) and 0.68 (resampling of 389 markers). The maximum similarity was between 0.95 and 0.98. For the Jaccard coefficient on the initial data set, the average similarity was 0.45, ranging from a minimum of 0.21 to a maximum of 0.78 (Table 3). The averages of all sampled similarity matrices were from 0.43 to 0.47; the minimum from 0.04 to 0.18 and the maximum from 0.77 (resampling of 389 markers) to 0.94 (resampling of 60 markers). In conclusion, if deviations were observed, it was for the minimum and maximum similarities, but only for very reduced data sets (60 markers). Therefore, a minor effect had to be attributed to the reduction of the initial data set when evaluating different marker selection settings.

Table 3 Average, minimum and maximum values of the pairwise similarity matrices based on Simple Matching and Jaccard coefficients, derived from AFLP data using different settings for selection for marker frequency and average marker signal intensity

Simple Matching coefficient			
Marker frequency selection	Selection for average fluorescent signal intensity (units)		
	>50	>120	>300
None	0.82 (0.69–0.94)	0.81 (0.63–1)	0.81 (0.60–1)
f>0.15	0.66 (0.50–0.92)	0.69 (0.50–1)	0.74 (0.47–1)
0.15<f<0.85	0.59 (0.43–0.90)	0.58 (0.38–1)	0.60 (0.35–1)
Jaccard Coefficient			
Marker frequency selection	Selection for average fluorescent signal intensity (units)		
	>50	>120	>300
None	0.45 (0.21–0.78)	0.53 (0.26–0.97)	0.65 (0.32–1)
f >0.15	0.53 (0.28–0.87)	0.60 (0.34–1)	0.68 (0.29–1)
0.15<f <0.85	0.36 (0.14–0.81)	0.39 (0.16–1)	0.45 (0.14–1)

The average value of a similarity matrix and ranges (Table 3) can provide different categories of information: (1) the general similarity in the set of plants, (2) the discrimination capacity between genotypes and (3) the robustness of the ordination to genotyping or scoring variability. General similarity for the set of plants was highly influenced by different marker selection parameter settings (Table 3). Exclusion of both rare and abundant markers provoked a decrease in the average values for both coefficients. Ordinations based on the Simple Matching coefficient were more affected as expected from the nature of this coefficient. For Jaccard, exclusion of rare markers induced a slight increase of the general similarity; the opposite was observed for the Simple Matching coefficient. Selection for more intensive markers had little influence or provoked an increase, depending on the setting for selection for marker frequency and the coefficient used.

Regarding the discrimination capacity between genotypes, the generally broader ranges of similarity obtained in Jaccard-based pairwise comparisons allowed better separation of individual genotypes. For both coefficients, imposing any restriction to the markers included in the analysis improved the discriminatory power of the ordination (as evaluated by the ranges). Only little influence was seen on the range by increasing the average fluorescent signal intensity from 120 to 300. The best discrimination capacity was obtained when both rare and abundant markers were excluded from the analysis.

The robustness of the ordination to genotyping or scoring variability could be evaluated because genetically nearly identical genotypes (bud sports) were included that showed no polymorphisms on the electropherograms. For certain settings, the maximum similarity was below 1, and bud sports were wrongly separated as distinct. Two facts can be the cause: (1) artefact bands from imprecise band definition by the Genescan software and (2) variations in signal intensities leading to the underscoring of bands with a peak height around the Genescan detection setting (50 units). Eliminating rare bands, excluding most of the artefact bands, gave no or only slight improvement; raising the average fluorescent signal intensity above the detection setting was much more effective.

As a test for the correlation of the enclosed data configuration, a standardised Mantel's statistic was calculated on pairs of similarity matrices. Two types of tests were aimed for. Firstly, using a Jaccard or Simple Matching coefficient, what is the sensitivity of an ordination for the variation in frequency selection under constant signal intensity settings (Table 4: upper panels) or, vice versa, for a variation in signal intensity selection under constant frequency selection settings (Table 4: lower panels)? Secondly, under the same marker selection settings, what is the correlation between an ordination based on the Jaccard versus the Simple Matching coefficient (Table 5)? The first assay can be considered as a test for the reliability of the classification to improper settings of marker selection criteria; the second as a test for the independence of the classification towards the use of a certain similarity measure. Using the Jaccard coefficient, the correlation between the similarity matrices was not affected a lot by the exclusion of the rare and abundant markers (Table 4, upper panels, values below the diagonal). For the Simple Matching coefficient (values above the diagonal), however, correlation decreased strongly. This effect was stronger when no signal intensity selection was applied. Marker selection to average fluorescent signal intensity did not reveal deviant behaviour between the use of Jaccard and Simple matching similarity (Table 4, lower panels). Correlation decreased when only the more intensive bands were scored. This effect was stronger if the rare and abundant markers were excluded. Correlation between Jaccard and Simple matching similarity matrices (Table 5) decreased when rare and abundant markers were excluded. Marker selection towards signal intensity increased this correlation. This effect was most clear when no frequency selection was applied.

Table 4 Standardised Mantel's statistic[a] between similarity matrices derived from AFLP data using different settings for selection for marker frequency and average marker signal intensity

Variation in frequency selection under constant average fluorescent signal intensity selection (above diagonal: using the Simple Matching coefficient; below diagonal: Jaccard coefficient)

Average signal intensity	Marker frequency selection	None	f>0.15	0.15<f<0.85
>50	None	–	*0.85*	*0.75*
	f>0.15	0.95	–	*0.96*
	0.15<f<0.85	0.90	0.96	–
>120	None	–	*0.90*	*0.77*
	f>0.15	0.96	–	*0.94*
	0.15<f<0.85	0.90	0.95	–
>300	None	–	*0.95*	*0.81*
	f>0.15	0.98	–	*0.90*
	0.15<f<0.85	0.87	0.91	–

Variation in average fluorescent signal intensity selection under constant frequency selection (above diagonal: using the Simple Matching coefficient; below diagonal: Jaccard coefficient)

Marker frequency selection	Average signal intensity	>50	>120	>300
None	>50	—	*0.95*	*0.85*
	>120	0.97	–	0.92
	>300	0.89	0.93	–
f>0.15	>50	–	*0.9*	*0.80*
	>120	0.96	–	*0.88*
	>300	0.88	0.91	–
0.15<f<0.85	>50	–	*0.80*	*0.72*
	>120	0.93	–	*0.81*
	>300	0.83	0.86	–

[a] Probabilities (*P*) evaluated by 1000 permutations were always equal to 0.001

Correlation between similarities based on molecular markers and kinship

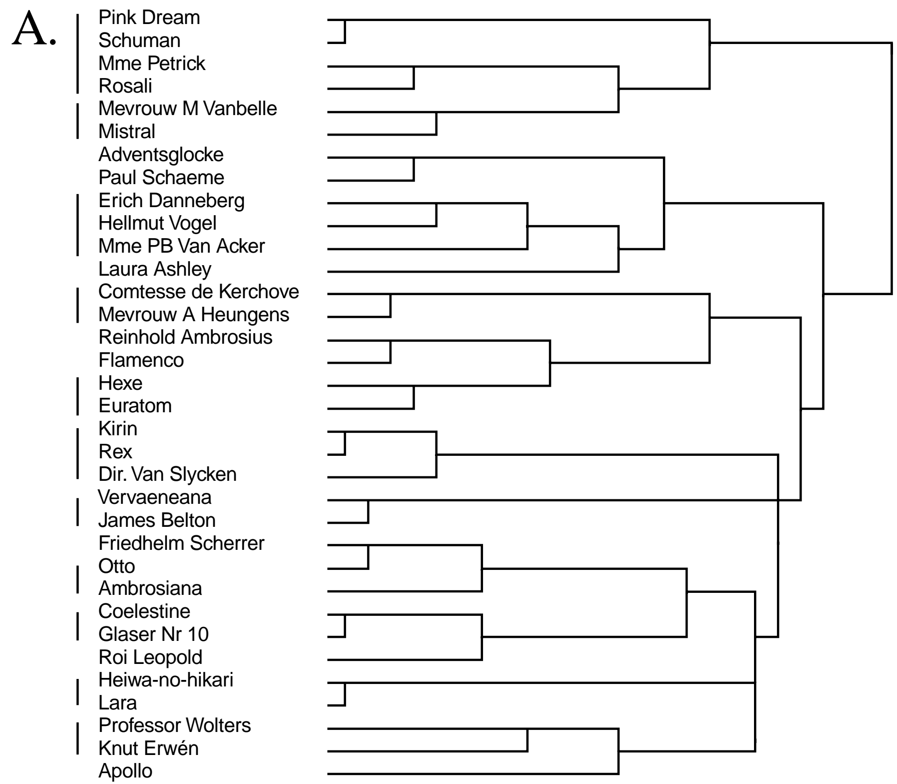
To compare classifications using molecular data to the known or accepted relationship of the genotypes studied, we compared the Jaccard and Simple Matching based similarity matrices obtained according to the marker selection criteria mentioned above to a pedigree-based distance matrix by Mantel analysis. Pedigree analysis was performed for 34 cultivars (Table 1) by defining an "unrelated" ancestor population for which no further pedigree information was available. For most of the cultivars, only three to maximally five generations could be traced back. As can be supposed from the history of *R. simsii* hybrids, many of these "unrelated" ancestors might in fact be highly related. Genetic similarity by descent was evaluated using kinship coefficients. Kinship has been defined as the probability that alleles of a given locus are identical by descent. It can also be considered as an estimate for the degree of genetic similarity between two individuals (Malécot 1948). For all pairs of plants in the pedigree (in total 75), pairwise kinship coefficients were calculated using KIN (Tinker and Mather 1993) and turned into a distance (1–r). To yield a symmetrical (34×34) distance matrix as an overall estimate of the genetic distance based on pedigree, we calculated Euclidean distances on the 34×75 partial matrix. Results are also presented as a dendrogram (Fig. 2). Starting from the AFLP data sets for 75 plants, under the different marker selection conditions, 34×34 similarity matrices using Jaccard and Simple Matching similarities were extracted. A standardised Mantel's statistic was calculated on each pair of similarity matrices and the pedigree-based distance matrix (Table 6). Values for the statistic were, in general, low, however probabilities evaluated by permutation were significant for several marker selection parameter settings. The most significant correlations were obtained using the Simple Matching coefficient with no or moderate selection to signal intensity and ex-

Table 5 Standardised Mantel's statistic between a Simple Matching coefficient based and a Jaccard coefficient-based similarity matrix, derived from AFLP data under identical settings for selection for marker frequency and average marker signal intensity

Average signal intensity	Marker frequency selection	Standardised Mantel's statistic[a]
>50	None	0.96
	f>0.15	0.92
	0.15<f<0.85	0.81
>120	None	0.98
	f>0.15	0.97
	0.15<f<0.85	0.90
>300	None	0.99
	f>0.15	0.98
	0.15<f<0.85	0.89

[a] Probabilities (*P*) evaluated by 1000 permutations were always equal to 0.001

Fig. 2A, B Ordination of azalea cultivars, based on: **A** pairwise kinship coefficients calculated on the pedigree (Euclidean distance; UPGMA-clustering) and **B** AFLP data (marker selection settings: 0.15<f<0.85, I>120; Simple Matching coefficient; UPGMA clustering). Corresponding clusters are indicated



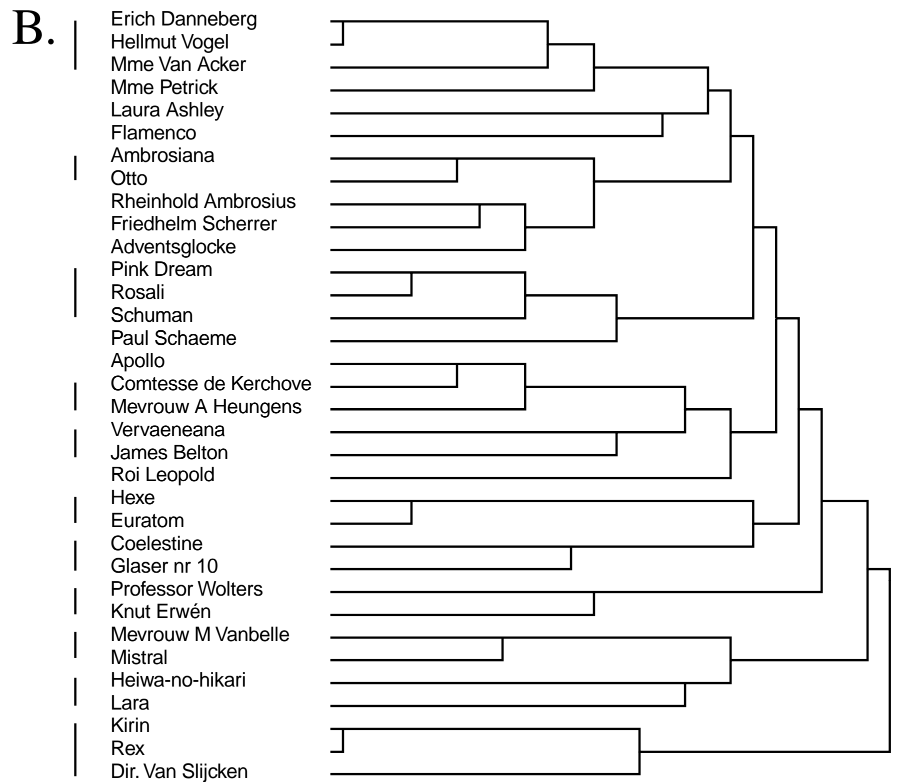


Table 6 Standardised Mantel's statistic (with indication of the probabilities) between similarity matrices based on Simple Matching and Jaccard coefficients, using different settings for selection for marker frequency and average marker signal intensity, and an Euclidean Distance matrix derived from pairwise kinship coefficients calculated on the pedigree

Simple Matching coefficient			
Marker frequency selection	Selection for average fluorescent signal intensity (units)		
	>50	>120	>300
None	0.094 (*P*: 0.132)	0.117 (*P*: 0.108)	0.158 (*P*: 0.078)
f>0.15	0.219 (*P*: 0.006)	0.220 (*P*: 0.010)	0.181 (*P*: 0.051)
0.15<f<0.85	0.236 (*P*: 0.002)	0.235 (*P*: 0.132)	0.199 (*P*: 0.015)
Jaccard Coefficient			
Marker frequency selection	Selection for average fluorescent signal intensity (units)		
	>50	>120	>300
None	0.149 (*P*: 0.068)	0.143 (*P*: 0.082)	0.155 (*P*: 0.075)
f>0.15	0.212 (*P*: 0.009)	0.194 (*P*: 0.023)	0.163 (*P*: 0.075)
0.15<f<0.85	0.227 (*P*: 0.002)	0.193 (*P*: 0.016)	0.158 (*P*: 0.038)

cluding rare and abundant markers (Mantel's statistic of 0.23; probability of 0.002). The corresponding dendrogram is presented in Fig. 2. Jaccard-based similarity matrices tended to have a lower correlation with pedigree than those based on the Simple Matching coefficient. Exclusion of rare markers always resulted in a better correlation between similarities based on molecular data and pedigree. This effect was predominant under no or moderate signal intensity selection. If only intensive markers were selected, there was a tendency to a lower correlation with pedigree-based similarity. Analysis of Fig. 2 indicates that for 10 pairs or groups, a high similarity expected from pedigree is affirmed by AFLP analysis. However, the structure in the AFLP-based ordination at a lower level of similarity did not correspond to the pedigree based ordination. Evaluating pedigrees as described resulted in plants being mainly grouped as a function of common ancestors. This was biased by some incomplete pedigrees, preferentially grouping descendants to the most important ancestor. For example, 'Hellmut Vogel' is closely clustered with 'Erich Danneberg', one of its parents. 'Erich Danneberg' in its turn is a direct descendant of 'Madame Pierre B. Van Acker' and 'Paul Schaeme' in the same cluster. However, 'Ambrosiana', the other parent of 'Hellmut Vogel', is separately clustered to its daughter 'Friedhelm Scherrer' and its granddaughter 'Otto'. This might be the reason for the low Mantel's statistics between pedigree and AFLP-based ordinations. The limited number of generations in the pedigrees caused an overestimation of the distances between groups where pedigree information for common ancestors is missing but where they may be highly related. The AFLP analysis grouped varieties with similar morphological characters better. Some examples are as follows. 'Ambrosiana', 'Otto', 'Reinhold Ambrosius', 'Friedhelm Scherrer' and 'Adventsglocke' are cultivars with similar growth habit, dark green elliptic leaves, mid season flowering and carmine red double flowers. In the pedigree based analysis, 'Pink Dream' is most closely clustered to 'Schuman', an inbred between two half-sibs of 'Pink Dream'. But phenotypically it shares light-green leaves, single light-pink-coloured flowers and a late-flowering time with its daughter 'Rosali'. In the AFLP-based analysis, Hirado azaleas and close relatives 'Heiwa-no-hikari', 'Lara', 'Mistral' and 'Mevrouw Marcel Vanbelle' ('Mistral'×'Hellmut Vogel'), which share growth habit, fast growth and fragrant flowers, are grouped better.

Classification of the breeders' gene pool

Final classification of the breeders' gene pool was performed using a data set with 179 markers (Table 1; f>0.15; average fluorescent signal intensity: 120 units). The choice of this particular set was based on the validation study performed: a fairly large data set with moderate exclusion of markers with a lower intensity, no exclusion of abundant markers, which resulted in a good correlation with pedigree. However, as has been previously discussed, other settings are possible that correlate to the results obtained by this particular one. Using this data set classification in dendrograms appeared to be independent of the use of Jaccard or Simple Matching coefficient or Euclidean distance (results not shown). Principal co-ordinate analysis was used to produce a two-dimensional ordination for the breeders' gene pool (Fig. 3). The current pot azalea assortment was clearly separated from the wild *Rhododendron* species. The latter clustered together with the Kurume azaleas. Hirado azaleas were separately grouped together with *R. scabrum*. Both Kurume and Hirado azaleas were grouped with their most important ancestor. Within the *R. simsii* hybrids two subgroups may be distinguished. The first group contains the archetype of a Belgian pot azalea – globular shape, dark-green leaves, early flowering and carmine red double flowers – although it has never been created so far in all its details as a single genotype. It is best characterised by members as 'Madame Petrick' (1880), 'Paul Schaeme' (1890), 'Ambrosiana' (1948) and 'Rein-

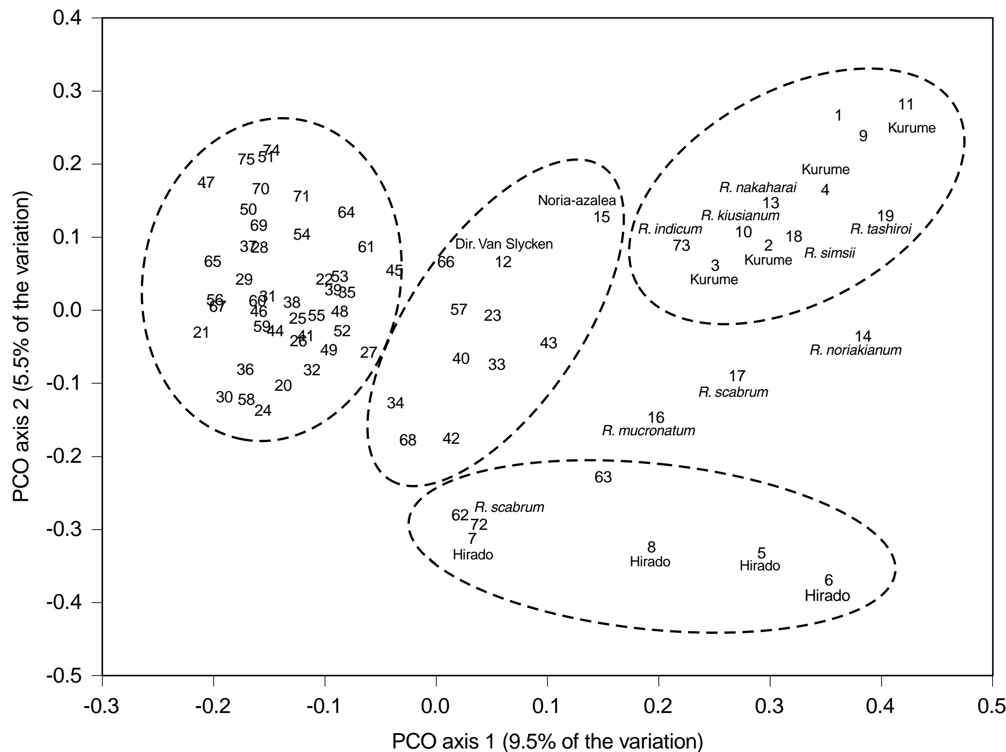


Fig. 3 PCO ordination for the breeders' collection of evergreen azalea, based on an AFLP data set with 179 markers (marker selection settings: f>0.15, I>120; Simple Matching coefficient). See Table 1 for variety names of ID-numbers

hold Ambrosius' (1930), which before 'Hellmut Vogel' (1967) were the most cultivated commercial pot azaleas. The second subgroup is a more loose group of cultivars, intermediate to the other groups. They can be generally typified as flowering late and having single or half-double flowers. A lot of them are older cultivars, e.g. 'Coelestine', 'Professor Wolters', 'Leopold-Astrid' or 'Tempérance'. Some are cultivars that originate from intermediate crosses: 'Cheops', a new pyramidal azalea created from *R. noriakianum*, 'Lara' and 'Mistral', crosses between Hirado azaleas and *R. simsii* hybrids; 'Directeur Van Slycken', a cross between Kurume azalea and a *R. simsii* hybrid. 'Coelestine', a very old cultivar of unknown origin, is also probably an intermediate form. It shares its globular shape, single flowers and late-flowering date with *R. kiusianum* and the Kurume azaleas. Part of this character is also present in its daughter 'Glaser Nr. 10'. 'Dorothy Gish', 'Kingfisher', 'Hexe' and 'Euratom' show a hose-in-hose type of flower, i.e. a flower of which the calyx is enlarged and coloured as a second corolla. This trait is typical for Kurume azaleas as 'Kirin' and 'Rex'. It is believed to be a mutation. The habit of 'Tempérance' and 'James Belton' is very atypical for *R. simsii* hybrids: both are late-flowering, very woody and erect: 'Tempérance' shows very uncommon lilac to blue flowers, 'James Belton' has tiny, green, hairy leaves.

Conclusions

In this relatively large azalea gene pool study fluorescent AFLP resulted in a primary data set with a high number of fragments. The selection of markers to be included in ordination analysis was performed by imposing filter thresholds to marker frequency and signal intensity. Validation of marker selection parameters showed that a range of settings can be found that give ordinations that are highly correlated to each other, that are merely independent of the use of Jaccard or Simple Matching coefficient and that show good agreement with known plant relationships based on pedigree or cultivar origin. Although it is demanding because of the technical infrastructure (DNA sequencer and computer equipment), fluorescent detection of AFLPs facilitates the generation of sufficient numbers of polymorphic markers. The number of PCs to be tested in a gene pool study can be restricted because automated fragment scoring can supply the capacity to register each fragment arising from a single AFLP primer combination. The use of consistent marker selection criteria like marker frequency and fluorescent signal intensity can then be of great help.

Acknowledgements The authors wish to thank Romain Uytterhaegen for his hands-on experience in azalea breeding and cultivation, Veerle Buysens and Laurence Desmet for their efforts in doing DNA preparations and AFLP reactions and Dr. Isabel Roldán-Ruiz for helpful discussions on data analysis and statistics. The whole staff of the biotech lab are very much appreciated for their support and skillful assistance. This research was funded by the Belgian Ministry of Small Enterprises, Traders and Agriculture – DG6 Research and Development D 1/2 – 5781 A – section 1. The experiments comply with current Belgian laws.

References

Ajmone Marsan P, Castiglioni P, Fusari F, Kuiper M, Motto M (1998) Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor Appl Genet 96:219–227

Chamberlain DF, Rae SJ (1990) A revision of Rhododendron IV Subgenus Tsutsusi. Edinburgh J Bot 47:89–200

Dellaporte SL, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–23

De Riek J, Dendauw J, Mertens M, Van Bockstaele E, De Loose M (1997) Use of AFLP for variety protection: some case studies on ornamentals. Med Fac Landbouww Univ Gent 62: 1459–1466

Doyle J, Doyle JL (1987) Genomic Plant DNA preparation from fresh tissue – CTAB method. Phytochem Bull 19:11

Galle FC (1987) Azaleas. Timber Press, Portland Ore.

Greenwood MS, Hopper CA, Hutchison KW (1989) Maturation in Larch. I. Effect of age on shoot growth, foliar characteristics and DNA methylation. Plant Physiol 90:406–412

Hill M, Witsenboer H, Zabeau M, Vos P, Kesseli R, Michelmore R (1996) PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in *Lactuca* spp. Theor Appl Genet 93:1202–1210

Hongtrakul V, Huestis GM, Knapp SJ (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among inbred lines. Theor Appl Genet 95:400–407

Kobayashi N, Takeuchi R, Handa T, Takayanagi K (1995) Cultivar identification of evergreen Azalea with RAPD method. J Jpn Soc Hortic Sci 64:611–616

Legendre P, Vaudor A (1991) The R Package: Multidimensional analysis, spatial analysis. Département de sciences biologiques, Université de Montréal, Montréal, Canada

Malécot G (1948) Les mathématiques de l'hérédité. Masson and Cie, Paris

Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance and near-isogenic line analysis. Theor Appl Genet 93:392–401

Paul S, Wachira FN, Powell W, Waugh R (1997) Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. Theor Appl Genet 94:255–263

Perkin-Elmer (1995) AFLP™ Plant Mapping Kit: Protocol

Rowland J, Nuygen B (1993) Use of PEG for purification of DNA from woody plants. Biotechniques 14:735–736

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of Matrix correspondence. Syst Zool 35:627–632

Tinker NA, Mather DE, (1993) KIN: Software for computing kinship coefficients. J Hered 84:238

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

Weising K, Beyerman B, Ramser J, Kahl G (1991) Plant DNA fingerprinting with radioactive and digoxigenated oligonucleotide probes complementary to simple repetitive DNA sequences. Electroforesis 12:159–169