Disclaimer: unless otherwise agreed by the Council of UPOV, only documents that have been adopted by the Council of UPOV and that have not been superseded can represent UPOV policies or guidance.

This document has been scanned from a paper copy and may have some discrepancies from the original document.

———

Avertissement: sauf si le Conseil de l'UPOV en décide autrement, seuls les documents adoptés par le Conseil de l'UPOV n'ayant pas été remplacés peuvent représenter les principes ou les orientations de l'UPOV.

Ce document a été numérisé à partir d'une copie papier et peut contenir des différences avec le document original.

———

Allgemeiner Haftungsausschluß: Sofern nicht anders vom Rat der UPOV vereinbart, geben nur Dokumente, die vom Rat der UPOV angenommen und nicht ersetzt wurden, Grundsätze oder eine Anleitung der UPOV wieder.

Dieses Dokument wurde von einer Papierkopie gescannt und könnte Abweichungen vom Originaldokument aufweisen.

———

Descargo de responsabilidad: salvo que el Consejo de la UPOV decida de otro modo, solo se considerarán documentos de políticas u orientaciones de la UPOV los que hayan sido aprobados por el Consejo de la UPOV y no hayan sido reemplazados.

Este documento ha sido escaneado a partir de una copia en papel y puede que existan divergencias en relación con el documento original.

E

UPOV

BMT/5/16
ORIGINAL: English
DATE: August 31, 1998

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA-PROFILING IN PARTICULAR

Fifth Session
Beltsville, United States of America, September 28 to 30, 1998

COMPARISON OF AFLP DATA WITH PEDIGREE (AZALEA) OR MORPHOLOGY (FLAX AND LINSEED)

Document prepared by experts from Belgium

COMPARISON OF AFLP DATA WITH PEDIGREE (AZALEA) OR MORPHOLOGY (FLAX AND LINSEED)

Jan De Riek[1], Johan Van Waes[2], Isabelle Everaert[1], Erik Van Bockstaele[1]en Marc De Loose[1]

[1]DvP-CLO: Department for Plant Genetics and Breeding - CLO Gent, Caritasstraat 21, 9090 Melle, Belgium
[2]DFE-CLO: Department for Crop husbandry and Ecophysiology - CLO Gent, Burg. Van Gansberghelaan 109, 9820 Merelbeke, Belgium

1. Introduction

The potential of molecular markers for variety identification and protection is an important research object at DvP-CLO involving different kind of crops (ryegrasses, azalea, flax, sugar beet and many "cases" in ornamentals). Different aspects for application in DUS trials, essential derivation and fraud protection are studied. Here, a comparison was made between molecular marker information coming from AFLP fingerprinting and pedigree (azalea) or morphology (flax). In this study results from 2 quite different crops are reported. Azaleas are flowering pot plants propagated by cuttings. The Belgian hybrids might have a narrow genetic origin but are related to Japanese cultivar types and wild species. In general, criteria for variety description of azaleas are sufficient for variety distinction. Flax and linseed are self-pollinating agricultural crops where, for certain types, crop experts dealing with DUS trials have low morphological variation between candidate varieties. Both azalea and flax are "minor" crops, in a way that little information from molecular marker techniques or genetic maps was available on beforehand. For such crops, fluorescent AFLP using a DNA sequencer and highly automated marker scoring can be a very efficient approach to generate a large set of molecular data in a short time. E.g. the flax study dealing with 170 plants was finished within 2 months time.

2. Materials en Methods

2.1 Plant material, DNA isolation, AFLP reactions and PAGE

2.1.1 Azalea

From the breeders gene pool at DvP-CLO, 75 individual plants were chosen: 4 Hirado, 7 Kurume, 55 Belgian pot azaleas and 9 related Rhododendron species from the Tsutsusi subgenus. DNA isolation was performed as in De Riek et al. (submitted). AFLP reactions were run on an ABI Prism 377 DNA Sequencer using the commercially available kit for fluorescent fragment detection (Perkin-Elmer, 1995). EcoRI and MseI were used for DNA digestion. Selective amplification was done using 3 fluorescent labelled EcoRI-MseI primer combinations with 6 selective bases: EcoRI-ACT/MseI-CTA, EcoRI-ACT/MseI-CAT and EcoRI-AAG/MseI-CTA.

2.1.2 Flax

From the 1997 DUS trial for linseed and flax (DFE-CLO, Belgium), 17 varieties were sampled (10 plants per variety). Sixteen belonged to the group of blue flowering flax with no ciliation of the bolls, a group which is difficult to distinguish based on morphological traits. 'Belinka' (white flowering) was included because it was used in a cross with 'Ariane' to create 'Escalina' (both included in the study). AFLP conditions were as described above. Selective amplification was done using 4 fluorescent labelled EcoRI-MseI primer combinations with 6 selective bases: EcoRI-AAG/MseI-CAT, EcoRI-AAG/MseI-CTA, EcoRI-AGC/MseI-CTA and EcoRI-AAC/MseI-CTG.

2.2 **Statistical analyses**

Filters for marker selection were set (De Riek et al., submitted) towards average signal peak height and marker frequency. For azalea, analysis was performed using the absence/presence (0/1) scores of the markers as primary data; for flax marker frequencies were used. 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 & Vaudor, 1991) and with the MVSP package (Kovach Computing Services, UK). Pedigree analysis was performed by calculating kinship coefficients (r) using KIN (Tinker & Mather, 1993).

3. Results and discussion

3.1 Azalea: analogy between similarities based on AFLP markers and kinship

Fluorescent detection and addition of an internal size standard to each lane enables the automated scoring of every fragment arising from a single AFLP primer combination. For the azalea data set, the use of 3 PC 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 8 extra restricted data subsets (Table 1). The average fluorescent signal intensity was used as a parameter for the reproducibility of the automated AFLP marker scoring; the frequency of a marker in the given data set was used in stead of Polymorphic Information Content (PIC) as a measure for the degree of polymorphism (De Riek et al., submitted). Pair wise plant genetic similarity was calculated for the 9 data sets using Simple Matching coefficient (symmetrical, including double-zeros) and Jaccard coefficient (asymmetrical, excluding double zeros). The averages, the ranges and the correlation to one other 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 included markers (De Riek et al., submitted).

For 34 cultivars, more or less extended pedigree information was available. To compare classifications, using molecular data, to the known or accepted relationship of the studied genotypes, the Jaccard and Simple Matching similarity matrices obtained according to the marker selection criteria mentioned above, were compared to a pedigree based distance matrix by Mantel analysis. Pedigree analysis was performed for 34 cultivars by defining an "unrelated" ancestor population for which no further pedigree information was available. For most of the cultivars, only 3 to maximally 5 generations could be traced back. As can be supposed from the history of *R. simsii* hybrids, many of these "unrelated" ancestors might be in fact highly related. Genetic similarity by descent was evaluated using kinship coefficients (r). Kinship (r) 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), pair wise kinship coefficients (r) were calculated using KIN (Tinker and Mather, 1993) and were turned into a distance (1 - r). To yield a symmetrical (34 x 34) distance matrix as an overall estimate of the genetic distance based on pedigree, Euclidean distances were calculated on the 34 x 75 partial matrix. Results are also presented as a dendrogram (Fig. 1a). Using the restricted data sets under the different AFLP marker selection conditions, 34 x 34 similarity matrices using Jaccard and Simple Matching similarities were calculated. A standardised Mantel statistic (Smouse et al., 1986) was calculated on each pair of similarity matrices and the pedigree based distance matrix (Table 2). This statistic can be compared to the computation of a Pearson correlation coefficient between the values of the two matrices (diagonals excluded). In general, correlation between similarity based on molecular data and pedigree is low. This was not unexpected due to the limited generations in the pedigrees and the nature of *R. simsii* hybrids (cross pollinators, small genetic basis) which probably caused an overestimation of the distances based on pedigree data. The Mantel statistic was maximally around 0.23 and was obtained using a Simple Matching coefficient with no or moderate selection to signal intensity and excluding rare and abundant markers. The corresponding dendrogram is presented in Fig. 1b. Analysis of Fig. 1 indicates that for 10 pairs or groups, a high similarity expected from pedigree is affirmed by AFLP analysis. It appeared that the secondary structure in the AFLP data from groupings at a lower similarity did not correspond well to the pedigree based ordination. Evaluating pedigrees as described caused that plants were mainly grouped as a function of common ancestors. This was biased by some incomplete pedigrees, preferentially grouping descendants to the most important ancestor. E.g. 'Hellmut Vogel' is closely clustered with 'Erich Danneberg', one of its parents. 'Erich Danneberg' on 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'. The AFLP analysis did better group varieties with similar morphological characters. Some examples: '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' x 'Hellmut Vogel') which share growth habitat (fast growing) and fragrant flowers, are better grouped.

3.2 **Flax: analogy between ordinations based on AFLP markers and morphological DUS traits**

The aim of this study was to compare classifications based on DUS data with those based on AFLP data. This was performed by comparing the analogy (using Mantel statistics) between similarity matrices based on morphological and molecular data using different distance measures. For the morphological evaluation all varieties in the DUS trial were considered first. AFLP analysis was performed on a subset of varieties which are difficult to distinguish based on morphological traits. Sixteen belonged to the group of blue flowering flax with no ciliation of the bolls. 'Belinka' (white flowering) was included because it was used in a cross with 'Ariane' to create 'Escalina' (both included in the study). So, the correlation between

similarities based on AFLP markers and morphological DUS traits was restricted to this particular group of morphologically more similar varieties.

3.2.1 Morphological DUS data for 88 linseed and flax varieties

Morphological data, collected according to the UPOV guidelines, originated from the 1997 DUS-trials at Merelbeke (DFE-CLO). All characters are scored on a 1 to 9 scale; a difference of 2 units is considered to be the minimum distance for distinction for a certain trait. Two varieties are considered to be distinct if at least 2 traits are distinct. For some traits not the full 1 - 9 scale is used. The 19 characteristics scored can then be attributed to 3 types: a.) binary (e.g. presence or absence of hairs): 3 traits, b.) multistate (e.g. colour of the corolla at the bud stage white, pink or purple): 7 traits, or c.) continuous (e.g. plant height): 9 traits. Within the continuous traits 3 traits were the average of measurements on different plants or seeds. For these 3 traits, we used the average values in stead of the class scores. In total, 88 linseed and flax varieties were tested. This data set was used to evaluate the usefulness of different multivariate ordination techniques (clustering, principal component or co-ordinate analysis, detrended correspondence analysis) together with the crop experts.

The choice of the resemblance measure used in some of the above mentioned multivariate ordinations appeared to be most determining for the obtained classification. Using different resemblance measures of the MVSP package, the correlation between the obtained similarity matrices was compared using standardised Mantel statistics (Table 3). For the ease of interpretation and to distinguish more similar groupings, this table has also been presented as a dendrogram (Fig. 2). The correlation between similarity matrices ranged from 0.26 to 1. This indicates a high influence of the distance measure choice when using morphological data under the form they were provided. Without willing to enter a discussion on what measure is more suited, some observations were made. Using Euclidean distance or similar measures, continuous traits that showed major differences (e.g. plant height and branching) were most determining the classification between linseed and flax. To overcome this, Gower similarity can be used. It is the only similarity measure that uses a different algorithm as a function of the data type (binary, multistate or continuous) and is accepted to be best suited for the combination of qualitative, quantitative or semi-quantitative descriptors. Principal co-ordinate analysis based on a Gower similarity matrix allowed to distinguish more or less the groups crop experts are using. PCO grouped linseed apart from flax; within flax white flowering types were separated from blue-purple flowering types. However, inside groups, varieties that are quite similar to a crop expert were very well separated by the Gower measure. So, this calculation technique appeared to be too discriminative to be used as a fast tool for detecting not or very little distinct varieties. This might be due to the fact that all traits were treated as equally important. A crop expert, familiar with the difficulties that can arise when scoring certain traits, will always take different weights for different traits into account.

3.2.2 AFLP data for 17 blue flowering flax varieties with no ciliation of the bolls

Using 4 AFLP primer combinations yielded a primary data set of more than 1300 markers when scoring 170 plants (10 plants per variety). Therefore, the following marker selection thresholds were applied: average signal peak height of a marker > 75; and frequency of a marker in the whole data set (170 plants) > 20/170 or frequency of a marker for a certain variety (10 plants per variety) > 5/10. This selection reduced the data set to 498 markers. For this data set all frequent markers were included in order not to inflate the variability within the

varieties. Different plants from a same flax variety show often very little polymorphisms. Only taking the polymorphic bands into account might cause a severe distortion. To compare the varieties, marker frequencies per variety were used.

The same range of similarity/distances measures as used above for the morphological traits were applied on the AFLP marker frequencies of the different varieties. Very little variation between the ordinations obtained by different measures was observed. Mantel coefficients between similarity matrices were always above 0.92. A typical result, presented as a dendrogram is shown for Euclidean distance and Pearson Product Moment Coefficient in Fig. 3. The Mantel statistic for this pair of ordinations was 0.92. In both ordinations, the 3 related varieties 'Ariane', 'Belinka' and 'Escalina' were most closely grouped. As indicated (Fig. 3), quite some similar groupings were retained.

Morphological DUS data for this subset of varieties were analysed in a similar way as in 2.1. using different resemblance measures and Mantel analysis. Although less pronounced than in Table 3, as can be expected when studying only a group of quite similar varieties, still the variation between ordinations obtained by different measures was large. This is exemplified by Fig. 4. The Mantel statistic ranged from 0.7 to 1; more or less the same structure according to more similar ordinations as seen from Table 3 was maintained in this reduced data subset.

Fig. 3 and 4 also allow to evaluate the agreement between ordinations based on morphological DUS data and on AFLP data. The corresponding Mantel statistics are shown in Table 4. When evaluating the range for different measures, Euclidean distance based ordination of AFLP data is a good example for a higher Mantel statistic and Pearson Product Moment for a lower one. However, when comparing Fig. 3 and 4, little to no groupings are retained.

4. **Conclusions**

Three types of variety information (pedigree, morphological and AFLP data) were used to classify varieties. Pair wise variety resemblance was calculated using different measures. Analogy between ordinations was checked using dendrograms, multidimensional scaling (PCA, PCO) and Mantel analysis. Here, we wanted to formulate some first conclusions.

Ordinations based on pedigree data can only be of value if detailed pedigree information from different ancestor generations is available. If not, they are likely to end up with different clusters of related plants that are unrelated among each other. Pedigree based ordinations are best suited to compare individuals from a same generation, more or less equally distant from the initial ancestor populations, as is mostly the case with animals. This can be problematic for plant varieties that can have "eternal" life, e.g. for vegetative propagated species.

Ordinations based on AFLP marker data appeared to be highly independent from the number of markers included at least if enough markers were included in the study. For such data sets, ordinations are also highly independent from the resemblance measure used. They show good agreement with pedigree information. Plants being similar by AFLP are more likely to be morphologically similar, in the same way as individuals from the same family are more or less alike. However, it is not true that when individuals are morphologically alike, they should be genetically related.

Ordinations based on morphological DUS traits can be biased by the "nature" of the descriptors used, being binary, in different classes or continuous, and by the range of variation

for each descriptor. This results in a great influence on the ordination of the resemblance measure used. There was only low agreement with ordinations based on AFLP data or pedigree. Plants with similar morphology can have a wide genetic conformity, based on molecular data.

Also some remarks can be formulated towards the different techniques we used for evaluation of the data. The values as such obtained by resemblance measures and Mantel statistics are always to be evaluated in an appropriate context, using known close or not related varieties as a reference. Presentation of results by dendrograms is best adapted for a low number of varieties and to search for close relationships. Multidimensional scaling (PCA, PCO, ...) are valuable for evaluation of a high number of varieties and to search for larger groupings.

References

De Riek J., Mertens M., Dendauw J., De Loose M., Heursel J., and Van Bockstaele E. (submitted to Theor. Appl. Genet.). Genetic diversity in evergreen azaleas assessed by fluorescent AFLP and automated data analysis

Legendre P, and Vaudor A (1991) The R Package: Multidimensional analysis, spatial analysis. Département de sciences biologiques, Université de Montréal, 142 p

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

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

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. Journal of Heredity: 84(3):238

Tables

Table 1: Number of AFLP markers included in the azalea study after selection to marker frequency in the whole data set and to the average signal peak height

Marker frequency selection	Average fluorescent signal intensity (units)		
	50	120	300
no	648	389	166
f > 0.15	245	179	102
0.15 < f < 0.85	195	131	60

Table 2: Standardised Mantel statistic between similarity matrices under different marker selection criteria and pedigree based Euclidean Distance matrix (azalea study)

SIMPLE MATCHING (double zeros)	Average fluorescent signal intensity (units)		
	50	120	300
no	0.094	0.117	0.158
f > 0.15	0.219	0.220	0.181
0.15 < f < 0.85	0.236	0.235	0.199

JACCARD (no double zeros)	Average fluorescent signal intensity (units)		
	50	120	300
no	0.149	0.143	0.155
f > 0.15	0.212	0.194	0.163
0.15 < f < 0.85	0.227	0.193	0.158

Table 3: Standardised Mantel statistic between morphological traits based resemblance matrices using different resemblance measures (flax study)

	AVDI	CANB	CHIS	CHOR	EUCL	GOWR	MCD	NOEU	PEAR	PERC	SPEA	SQEU
CANB	0.48											
CHIS	0.70	0.85										
CHOR	0.66	0.89	0.98									
EUCL	1.00	0.48	0.70	0.66								
GOWR	0.53	0.92	0.85	0.86	0.53							
MCD	0.92	0.73	0.89	0.86	0.92	0.78						
NOEU	0.62	0.64	0.85	0.83	0.62	0.70	0.77					
PEAR	0.45	0.61	0.78	0.78	0.45	0.64	0.65	0.91				
PERC	0.89	0.75	0.89	0.87	0.89	0.80	0.98	0.81	0.71			
SPEA	0.34	0.76	0.83	0.84	0.34	0.76	0.62	0.74	0.81	0.64		
SQEU	0.96	0.40	0.63	0.58	0.96	0.44	0.86	0.58	0.41	0.82	0.26	
STEU	0.96	0.86	0.89	0.89	0.96	0.89	0.79	0.79	0.72	0.82	0.74	0.51

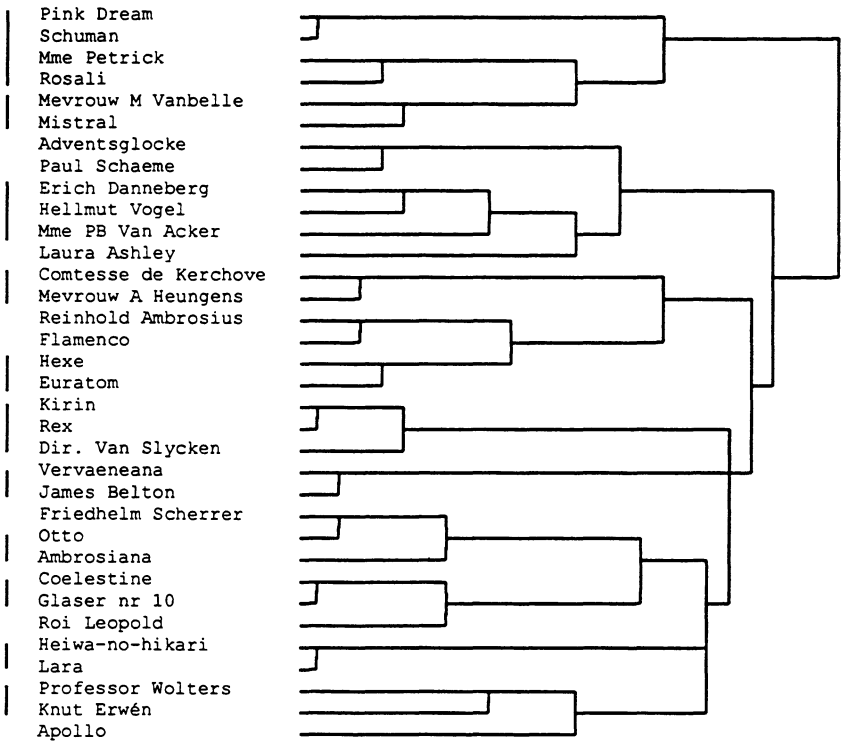
The resemblance measures tested are those of the MVSP package (Kovach Computing Services, Wales)

AVDI: Average Distance
EUCL: Euclidean Distance
STEU: Standardised Euclidean distance
SQEU: Squared Euclidean distance
CANB: Canberra distance
GOWR: Gower distance
CHIS: Chi-squared distances
CHOR: Chord distance
MCD: Mean Character Difference
PERC: Percent similarity
NOEU: Normalised Euclidean Distance (Cosine U)
PEAR: Pearson Product Moment Coefficient
SPEA: Spearman Rank Order Coefficient

Table 4: Standardised Mantel statistic between morphological traits based and AFLP based resemblance matrices (flax study)

Morphology based resemblance	AFLP based resemblance	
	Euclidean Distance	Pearson Product Moment
Euclidean Distance	0.65	0.49
Gower similarity	0.58	0.42

a) Pedigree



b) AFLP

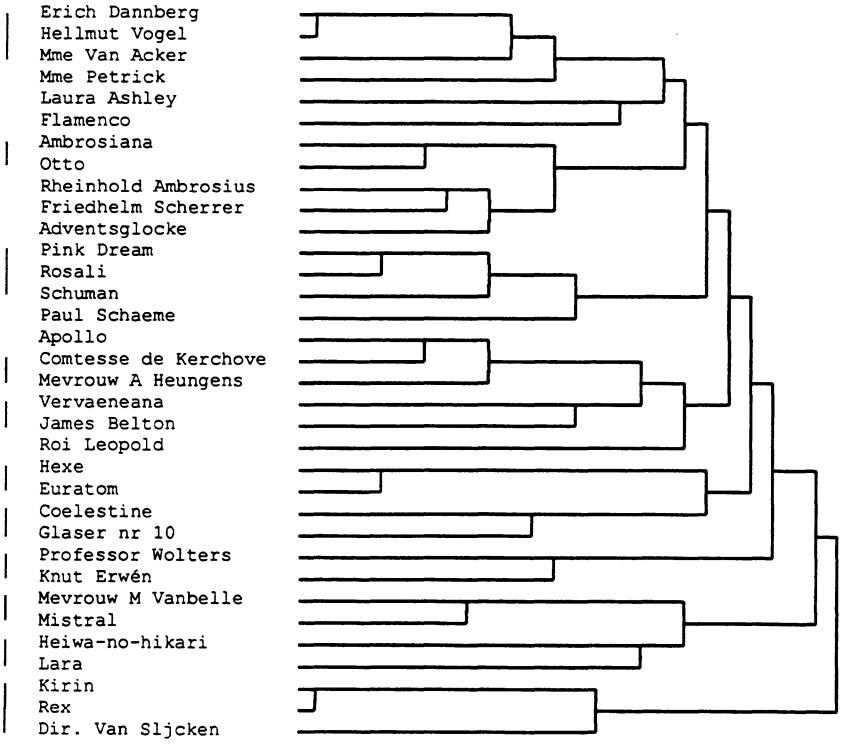


Fig. 1: Ordination of azalea cultivars, based on pedigree data and on AFLP data

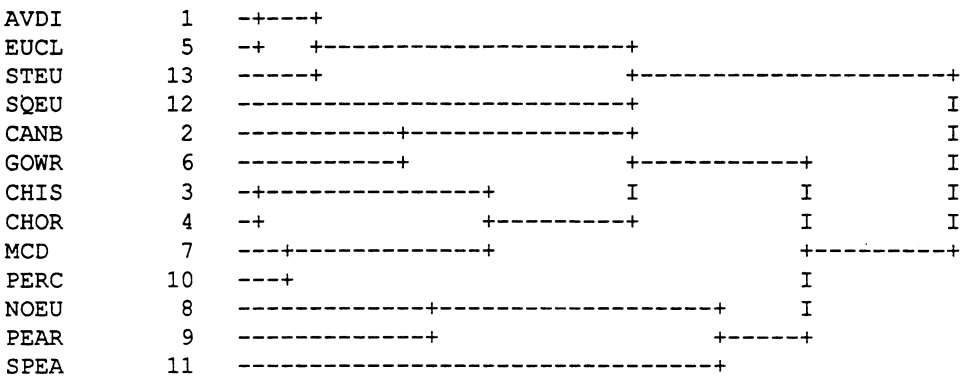
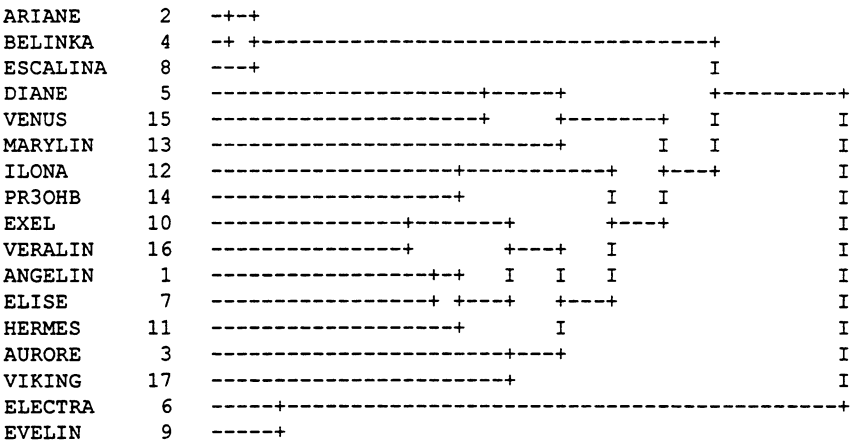


Fig. 2: Corresponding dendrogram based on Table 3

Euclidean Distance



Pearson Product Moment

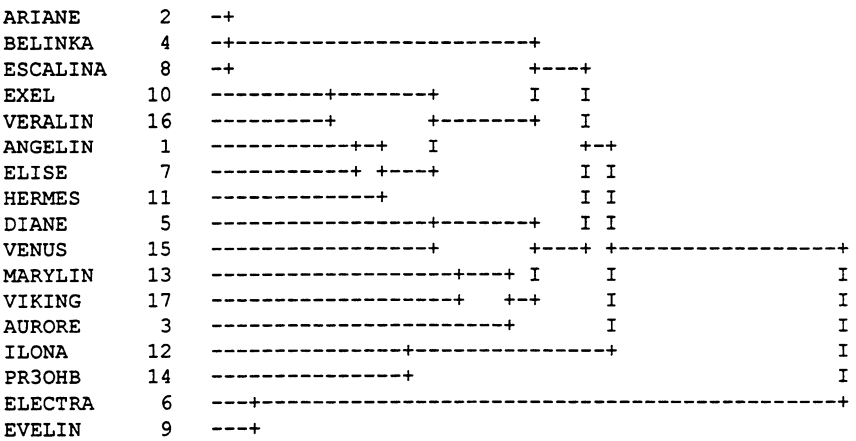
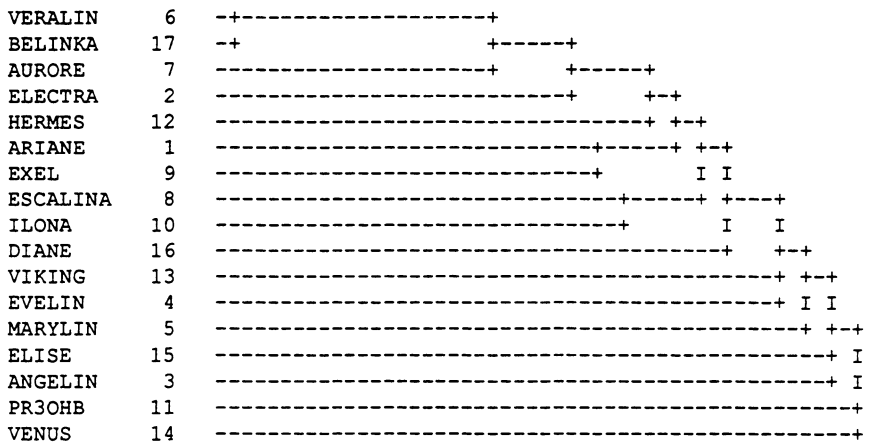


Fig. 3: Ordination of flax cultivars, based on AFLP data

Gower similarity



Euclidean Distance

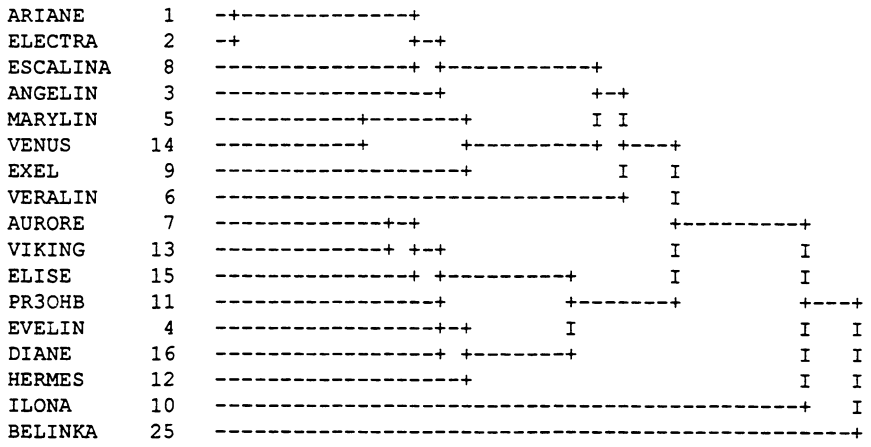


Fig. 4: Ordination of flax cultivars, based on morphological data

[End of document]