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agreed by the Council of UPOV, only documents that have been adopted by the Council of UPOV and that have not been superseded can represent UPOV policies or guidance.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/5/5
ORIGINAL: English
DATE: September 8, 1998

NTERNATIONAL 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 GENETIC DISTANCES BETWEEN RAPESEED CULTIVARS CHARACTERIZED BY AFLP MARKERS

Document prepared by experts from France

COMPARISON OF GENETIC DISTANCES BETWEEN RAPESEED CULTIVARS CHARACTERIZED BY AFLP MARKERS

V. Lombard*, C.P. Baril**, P. Dubreuil**, F. Blouet** and D. Zhang*

* GEVES, domaine du Magneraud, BP 52, F-17700 Surgères, France
** GEVES, La Minière, F-78285 Guyancourt cedex, France

Introduction

Since the notion of essential derivation was introduced in the last UPOV convention, the assessment of the genetic relatedness between cultivars has become a crucial issue. At present and for most species, registration or/and protection of a new variety only relies on morphological traits for the establishement of Distinction, Uniformity and Stability (DUS). Many of the traits considered are complex and their expression are influenced by the environmental conditions. Morever, for some crops like rapeseed, only few discriminant characters (about ten) are available for DUS testings and alternative descriptors are required. Because of the high degree of discriminant information, molecular markers have been widely and successfully applied for cultivars identification in many species. In particular, AFLP method (amplified fragment length polymorphism) easily provides a large number of markers on a single gel without requiring sequence information for their development.

The importance of reference collections in DUS testings is a problem where molecular markers could be hepful. At present, a new variety has to be compared with all the cultivars previously registered which implies the use of wide area to make comparisons. Molecular markers could be useful to permit a pre-screening of the closest varieties for direct comparisons with new varieties. For this aspect, we propose to evaluate the utility of the Analysis of Molecular Variance (AMOVA) to test the significance of *a priori* levels of structure in our collection.

In the framework of plant protection, statistical tools are required to evaluate the utility of molecular data for assessing genetic proximity and dependance between cultivars. For this purpose, genetic distances seem to be a suitable approach.

The aim of this study is to (i) evaluate the utility of AFLP markers for the identification of rapeseed cultivars, (ii) apply AMOVA to test levels of structure of a collection of cultivars and (iii) compare several genetic distance estimators in the way they reflect true genetic associations between cultivars.

Materials and methods

Plant material

Eighty-three rape seed cultivars were studied including both spring and winter types. Most of the cultivars were chosen to represent the european genetic diversity of rapeseed. Among these cultivars were pairs of very related lines. Three pairs are near isogenic lines : Darmor and Darmor Nain are different for one nanism gene, B ms and E ms are the mâle sterile forms of B and E, respectively. Three cultivars (Apex, goeland and Lady) were registered in three

different countries but no morphological differences can be observed (now, they are synonimous in the European list).
The remainer cultivars are either ancestral cultivars of historical importance in rapeseed selection, or non-European cultivars that were chosen to enlarge genetic diversity (Table 1).

DNA extraction and AFLP assays

Thirty seedlings of one-week age per cultivar were pooled and frozen in liquid nitrogen. Total DNA was extracted using a CTAB protocole with minor modifications (Rogers and Bendich 1988).

AFLP analysis was performed according to the procedure described by Vos *et al.* (1995), using a commercially available kit (AFLP analysis system I, Gibco BRL) and radiolabelling E*co* RI primers with gamma-[^{33}P]-dATP to detect bands. PCR products were loaded into a 5% polyacrylamide gel. After electrophoresis, gels were dried and exposed to X-ray film for about 5 days. 17 primer combinations involving five E*co*RI and seven M*se*I primers provided clear interpretable patterns. AFLP bands were scored as absence (0) or presence (1).

Statistical analysis:

Structure of the genetic diversity
The *a priori* structure of the genetic diversity of the collection was tested with AMOVA (Excoffier *et al.* 1992). This method provides us with an estimate of the fraction of between-population diversity (*i.e.* Φ_{st}), analogous to the Fst which can be tested using a permutational procedure. In the present study, this method has been applied to test the structure (i) among winter and spring cultivars (type of cultivars), (ii) among countries of origin for winter cultivars and (iii) among breeding companies. AMOVAs were performed with a program in Turbo Pascal provided by Christine Dillmann. The results were compared with *a posteriori* structure revealed by a principal component analysis (PCA).

Comparisons of genetic distance estimators
Three common distance estimators were computed between each pair of cultivars: the Jaccard's distance (J) (1908), the Nei & Li's distance (NL) (1979) and the Sokal & Michener's distance (SM) (1958) as,

$$J_{xy} = 1 - \frac{n_{11}}{n_{11} + n_{10} + n_{01}} = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01}} \quad [1]$$

$$NL_{xy} = 1 - \frac{2 \times n_{11}}{2 \times n_{11} + n_{10} + n_{01}} = \frac{n_{10} + n_{01}}{2 \times n_{11} + n_{10} + n_{01}} \quad [2]$$

$$SM_{xy} = 1 - \frac{n_{11} + n_{00}}{n_{11} + n_{10} + n_{01} + n_{00}} = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01} + n_{00}} \quad [3]$$

where n_{11} is the number of bands shared by the cultivars x and y (positive matching), n_{10} is the number of bands present in x and absent in y, n_{01} the number of bands present in y and absent in x, and n_{00} the number of bands absent both in x and y (negative matching).

We also computed a weighted Jaccard's distance (WJ) to take into account the frequency of each marker in the calculation of the distance. This type of weighting has been already proposed by Baril *et al.* (1997). The rationale is as follows: a difference at a very frequent or a very rare marker is a more original information than a difference at a marker which frequency is nearly 0.5. Because of the symmetry between a rare or a frequent marker, we weighted the comparison at a marker by the inverse of its PIC (polymorphic information content). In the case of a biallelic marker (absence or presence of the band),

$$PICj = 1 - \left(p_j^2 + (1 - p_j)^2\right) = 2 \times p_j (1 - p_j),$$

where pj is the frequency of marker j.

$$WJ_{xy} = \frac{1}{n_{11} + n_{10} + n_{01}} \sum_{j=1}^{M} \frac{(B_{xj} - B_{yj})^2}{2 p_j (1 - p_j)}$$

[4]

where B_{xj} (resp. B_{yj}) is the frequency of the band corresponding to the marker j within the cultivar x (resp. y) and $M = n_{11} + n_{10} + n_{01}$.

Remarks:
Several correspondences can be noted between genetic distances used in this study and other dissimilarity coefficients. The Jaccard and Sokal & Michener distances are two particular forms of the Gower coefficient of dissimilarity (1971) when negative matching are or not included in the calculation, respectively. Nei & Li distance is equivalent to the Dice coefficient of dissimilarity and Sokal & Michener distance is also calling simple matching distance (Sneath and Sokal 1973). Note that N x SM is the square Euclidian distance between two cultivars, where N is the total number of markers (Baril *et al.* 1997).

Results

Polymorphism and power of dicrimination revealed by AFLP markers in rapeseed

17 primer combinations revealed a total number of 324 polymorphic bands. The number of markers per primer combination ranged from 12 to 30, with an average of 19.1. The most polymorphic primer combinations were E-AAC+M-CAA, E-AAC+M-CTT and E-AAG+M-CTT, which revealed 30 markers (Table 2).

The frequency of markers among the 83 cultivars varied from 0.060 to 0.988, with an average of 0.54 (results not shown). The individual polymorphic information content (PIC) varied between 0.024 and 0.50 (the theoretical maximum for a biallelic loci) with an average of 0.342. 39.2% of the marker had a PIC greater than 0.4 . The power of discrimination of each primer combination was estimated by computing the number of distinguished cultivars (Table 2). This number ranged from 30 to 77. Two primer combinations were required to distinguish all the cultivars: E-AAC+M-CAA and E-AAC+M-CTT or E-AAC+M-CTT and E-AAG+M-CTT

Structure of the genetic diversity (*a priori* and *a posteriori*)

The two factors (type of cultivar and country of origin) tested with AMOVA were both highly significant (Table 3). 32.9% of the molecular variance is distributed between winter and spring cultivars. Among winter cultivars, 11.4% of the variance is due to the partition between French and German cultivars. For the third factor, the genetic variance was significantly structured between each pair of breeders, except between breeders 2 and 3 who are French and German, respectively (Table 3).

Comparaisons between genetic distance estimators

Jaccard distance, Nei & Li distance, Sokal & Michener distance and weighted Jaccard distance ranged from 0.070 to 0.758, from 0.036 to 0.610, from 0.038 to 0.599 and from 0.166 to 2.541, respectively. For the four types of distance, Apex – Goeland and Goeland – Lady are the closest cultivars and the biggest distance was between a French winter cultivar and a Canadian spring cultivar. The three pairs of near isogenic lines were also included in the 16 nearest couples (Table 4). The simple coefficients of correlation calculated between the four genetic distances were very high for the six pairwise comparisons and were over 0.945.

Discussion

Polymorphism of AFLP markers and discriminatory power of the primer combination

Our results show that AFLP is a powerful tool to dicriminate rapeseed cultivars. 324 markers are revealed by 17 primer combinations, with an average of 19.1. This number is not very high in comparison with other intraspecific genetic diversity analyses using AFLP - 60 for sunflowers accessions (Hongtrakul *et al.* 1997)-. This result reflect the relative low genetic diversity in the european genetic pool. AFLP markers are higly polymorphic with high values of PIC and only two primer combinations (60 markers) are required to distinguish the 83 cultivars.

Structure of the collection

The significant *a priori* structure of the collection (in term of genetic type, country of origin and breeder) assessed with AMOVA is consistent with the *a posteriori* structure revealed by PCA (Fig. 4). PCA exhibited a clear split between the spring and winter cultivars and high genetic relationships between cultivars from the same breeder. In our study, we can separate most of the French winter cultivars from the other European winter cultivars, showing the genetic originality of this group. These results are similar to associations revealed with RFLP markers reported in Diers *et al.* (1994) and in Lee *et al.* (1996), with RAPD markers in Mailer *et al.* (1994) and with SSR primers in Charters *et al.* (1996). AMOVA seems to be a good tool to test significance of groupings.

Comparison between genetic distances.

Jaccard and Nei & Li are very highly correlated (0.996) and lead to identical rankings of genetic distances. This was expected, because NL can be easily expressed as an increasing function of J: $NL_{xy} = \frac{J_{xy}}{2 - J_{xy}}$. This relation between the two distances clearly implies a high correlation between them. Then, the choice between J and NL depends on the type of molecular markers used to characterize the cultivars. As reported in Piepho and Laidig (TWC/14/8), NL is more appropriate than J for codominant markers like RFLP while J is more appropriate for dominant markers like RAPD or AFLP. In fact, NL distance gives a double weight to the similarity at a band between two cultivars. This weighting is justified for RFLP markers because there are allelic relationships between bands, which is not the case for AFLP markers.

The high correlation between J and SM (0.983) was not obvious. The difference between these distances (formula [1] and [3]) came from negative matches which are taken into account in the denominator of SM distance and not in the J distance. One explanation, supported by *Peltier et al.* (1994), is that in a case of intra-specific study, an allelic relation

exist between presence and absence of a band and a negative matching is an indication of similarity leading to same kind of results with SM and J in our study.

The weighting of Jaccard (WJ) distance by the inverse of the PIC provided similar relationships between cultivars to Jaccard ones. However, WJ lead to take the most « exotic » cultivars away from the remaining cultivars. One interesting information provided by this distance is the proximity of the couple of near isogenic lines (Darmor and Darmor Nain), different for one nanism gene. Those cultivars are expected to be among the first nearest pairs of cultivars. However, this couple was the thenth, the eleventh and the twelveth nearest couple with J, NL and SM, repectively whereras it was the fourth with WJ.

Choice of a genetic distance estimator for essential derivation

In the framework of plant protection, the choice of the genetic distance is crucial for determining the level of relatedness between cultivars. For the distinctness and without any genetic consideration, J and NL are independent of the cultivar samples because only bands present in x and/or in y are considered. For SM, negative matches are counted and if a new cultivar carries a new band absent in the cultivars previously registered, this becomes a new negative matching for these cultivars and the distance will change. This can be avoid if a set of marker is definitively chosen. The stability of genetic distance is a very attractive quality for breeders because a distance between two cultivars is constant when the number of cultivars in the reference collection increase.

A contrario, the disadvantage of J results in the difficulty of finding statistical distribution of this distance which is important to calculate a confidential interval. This difficulty comes from the denominator, which is not a constant but a random variable. In this case of problematic theoretical computations, bootstrap procedure allows to empirically estimate the sampling variance (Tivang *et al.* 1994) but no statistical tests are available to compare distances between pairs of cultivars.

It is easier to work with euclidian distances like SM or Rogers distance (Rogers 1972). They can be modelled as a binomial variable and their statistical properties are well known (Dillmann *et al.* 1997). This modelisation implies that molecular markers are a random sample of the genome. This assumption is not verified when a subset of markers is chosen for an optimum genome coverage. In this case, Dillmann *et al.* (1997) showed that taking into account the position of molecular markers on a genetic map in the calculation of Rogers distance improves the precision of the estimation of genetic distance in the proportion of 40% between inbred lines that are related by pedigree.

To conclude, this study shows the great discriminatory power of AFLP makers and their capacity to well represent the genetic relationships between rapeseed cultivars consistently with their genetic origin. AMOVA can be successfully applied to test significance of groups in reference collection but methodologies to define groupings must be investigated. The different genetic distance estimators are very correlated and lead to very similar associations between cultivars. However, the rankings are a bit different especially for near cultivars which can lead to different decisions about the genetic dependance between varieties. Then, the choice of a genetic distance in the context of plant protection has important consequences and further investigations are needed about the precision of their estimations and the conditions of their applications (in term of type of molecular markers, genetic structure of the cultivars, diversity of reference collections, breeding programs).

References

Baril CP, Verhaegen D, Vigneron Ph, Bouvet JM, Kremer A (1997) Structure of the specific ability betweentwo species of *Eucalyptus*. I. RAPD data. Theor Appl Genet 94:796-803

Charters YM, Robertson A, Wikinson MJ, Ramsay G (1996) PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp oleifera) using 5'-anchored simple sequence repeat (SSR) primers. Theor Appl Genet 92:442-447

Diers BW, Osborn TC (1994) Genetic diversity of oilseed *Brassica napus* germ plasm based on restriction fragment length polymorphisms. Theor Appl Genet 88:662-668

Dillmann C, Charcosset A, Goffinet B, Smith JSC, Dattée Y (1997) Best linear estimator of the molecular genetic distance between inbred lines. In: Krajewski P, Kaczmarek Z (eds) Advances in biometrical genetics. Proceedings of the tenth meeting of the EUCARPIA section biometrics in plant breeding, 14-16 may 1997, Poznan, pp 105-110

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479-491

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27:857-874

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

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223-270

Lee D, Reeves JC, Cooke RJ (1996) DNA profiling and plant variety registration:2. Restriction fragment length polymorphisms in varieties of oilseed rape. Plant Varieties and Seeds 9:181-190

Mailer RJ, Scarth R, Fristensky B (1994) Discrimination among cultivars of rapeseed (*Brassica napus* L.) using polymorphism amplified from arbitrry primers. Theor Appl Genet 87:697-704

Nei M, Li WH (1979) Mathematical models for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269-5273

Peltier D, Chacon H, Tersac M, Caraux G, Dulieu H, Berville A (1995) Utilisation des RAPD pour la construction de phénogrammes et de phylogrammes chez *Petunia*. In: Techniques et utilisations des marqueurs moléculaires. Coll Les colloques INRA

Piepho H.P., Laidig F.(1996) A review of methods for cluster analysis of marker data. In Technical working party on automation and computer programs, 14th session, Hanover, Germany June 4 to 6 1996, TWC/14/8.

Rogers JS (1972) Mesures of similarities and genetic distances. Studies in genetics VII. Univ Texas Publ 7213: 145-153

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. Plant Molecular Biology Manual A6: 1-10

Sneath PHA, Sokal RR (1973) Numerical taxonomy. WH Freeman, San Francisco

Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. Univ Kansas Sci Bull 38:1409-1438

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular data using bootstrap procedure. Theor Appl Genet 89: 259-264

Vos P Hogers R, Bleeker M, Reijans M, Lee T van de, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res Vol 23, No 21: 4407-4414

Table 1 Origin of the plant material with, in parenthesis, the number of cultivars

Type of cultivar	Country of origin	Breeding company
Winter (68)	France (26)	Breeder 1 (14)
		Breeder 2 (9)
		3 other French breeders (3)
	Germany (26)	Breeder 3 (5)
		Breeder 4 (7)
		Breeder 5 (9)
		3 other German breeders (5)
	4 other European countries (15)	
	Japan (1)	
Spring (15)		

★ ★ ★

Table 2 Level of polymorphism and power of dicrimination revealed by the 17 primer combinations

Primer combination	No. of markers	No. of distinguished cultivars
E-ACC+M-CAG	12	30
E-AAC+M-CTA	12	50
E-AAG+M-CAC	17	50
E-ACT+M-CTC	15	51
E-AGG+M-CAT	15	53
E-ACT+M-CAT	16	56
E-AAG+M-CAT	20	57
E-ACT+M-CTT	21	58
E-ACC+M-CTC	12	59
E-AAG+M-CAA	17	60
E-AAG+M-CTC	19	62
E-ACC+M-CAA	16	64
E-AAC+M-CAG	17	65
E-AAC+M-CAT	25	76
E-AAC+M-CAA	30	77
E-AAC+M-CTT	30	77
E-AAG+M-CTT	30	77
Total	**324**	
Mean	**19.1**	**60.1**

The core sequences of primers for the selective amplification were:
E- = 5'-GACTGCGTACCAATTTC-3' for EcoRI primers ;
M- = 5'-GATGAGTCCTGAGTAA-3' for MseI primers.
Each primer contained 3 selective nucleotides at the 3' end. For example, primer E-AAC contained the core sequence plus AAC at its end.

Table 3 Analysis of molecular variance of three factors on molecular data matrix

Factor	Levels	sample size	Φst			
Type	Winter	68	0.329 ***			
	Spring	15				
Country of origin	France	26	0.114 ***			
	Germany	26				
Breeders [1]	1	14		1	2	3	4
	2	9	2	0.112 *			
	3	5	3	0.205 **	0.090 NS		
	4	7	4	0.255 ***	0.196 ***	0.202 **	
	5	9	5	0.207 ***	0.087 **	0.094 *	0.110 **

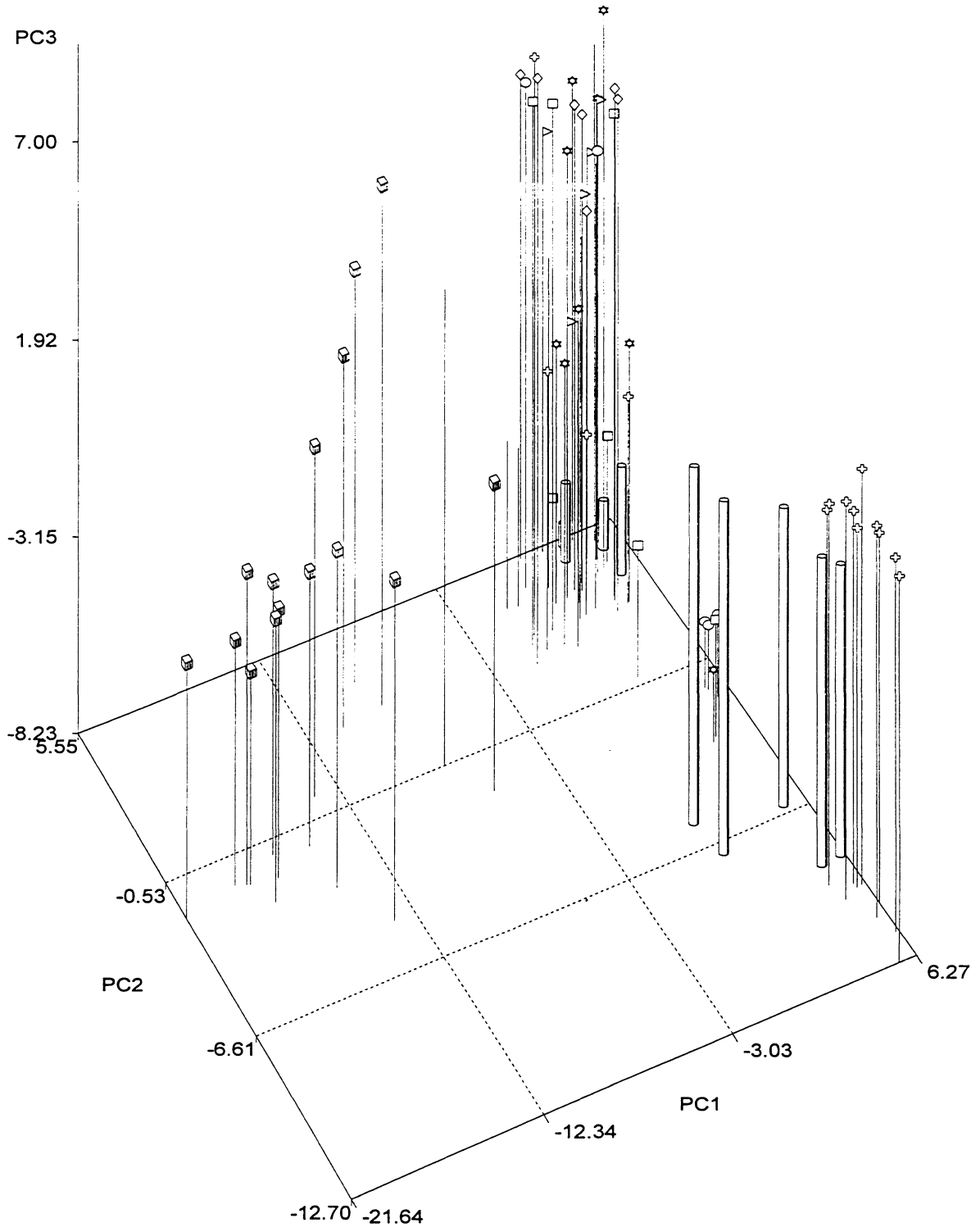
[1]: other breeders are excluded due to the low number of their cultivars in the study
NS: non significant, *: P<5%, **: P<1%, ***: P< 0.1%

★ ★ ★

Table 4 Genetic distances between pairs of related cultivars

Pairs of cultivars	Nei & Li distance		Jaccard distance		Weighted Jaccard distance		Sokal & Michener distance	
	value	rank *	value	rank *	value	rank *	value	rank *
Apex – Goeland	0.036	1	0.070	1	0.166	1	0.038	1
Goeland – Lady	0.044	3	0.084	3	0.254	3	0.045	3
Apex – Lady	0.065	8	0.121	8	0.343	9	0.067	10
Darmor – Darmor Nain	0.069	**11**	0.129	**10**	0.296	**4**	0.076	**12**
A ms – A	0.060	5	0.112	5	0.312	6	0.061	4
B ms – B	0.082	14	0.152	14	0.466	16	0.093	14

* rank among the nearest distances



Symbols

cross: French winter cv. from breeder 1	balloon: German winter cv. from breeder 3
pillar: French winter cv. from breeder 2	diamond: German Winter cv. from breeder 4
square: other French winter cv.	star: German cv. from breeder 5
needle: other winter cv.	flag: other German winter cv.
cube: spring cv.	

Fig. 1 principal component analysis of 83 rapeseed cultivars based on 324 AFLP markers

[End of document]