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agreed by the Council of UPOV, only documents that have been adopted by the Council of UPOV and that have not been superseded can represent UPOV policies or guidance.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

BMT/4/9
ORIGINAL: English
DATE: January 28, 1997

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

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

Fourth Session
Cambridge, United Kingdom, March 11 to 13, 1997

THE USE OF THE ANALYSIS OF MOLECULAR VARIANCE (AMOVA) FOR DISTINCTION STUDIES

Document prepared by experts from France

Introduction

The analysis of molecular variance (AMOVA) (Excoffier et al, 1992) has been introduced as an extension of the analysis of gene frequencies (Cockerham, 1973, Long, 1986, Weir and Cockerham, 1984) for molecular haplotypes in an haploid system. It basically consists in using the distance matrix between haplotypes to measure the population genetic structure within a species. It is a multilocus approach. It's main interest is the testing procedure, based on a permutational analysis, which requires very few assumptions. Although initially designed for haploid systems, the AMOVA treatment provides a general framework for the analysis of population genetic structure (Michalakis and Excoffier, 1995). It has been applied to multilocus molecular and enzymatic data in diploid populations of buffalograss (Peakall et al, 1995) and may virtually be used on any distance matrix. One of the potential application of AMOVA concerns Plant Breeder's Rights for species commercialized as population or synthetic varieties. In this case, DUS studies have to take into account the variability of the discriminant traits within populations. AMOVA can be extended to perform pairwise comparisons between populations and test for multilocus significant differences. The purpose of this paper is (i) to describe the principle of the analysis of molecular variance, and (ii) to discuss its potential use in distinction studies. The results will be illustrated with an analysis of Ray-grass biochemical data and of Alfalfa molecular data.

Analysis of molecular variance : principles

The analysis of population genetic structure consists in characterizing genetic variation within and between populations. Analyses are established for the case in which data are available in samples from different populations, or different subdivisions of the same population. The genetic variation may be a consequence of the statistical sampling, that results in each sample having a different set of individuals. It may also be a consequence of a genetic sampling process, because each generation of a population is formed by the

union of gametes chosen from among those produced by the previous generation. If the populations sampled have a common evolutionary history, then the process of genetic sampling, or drift, between successive generations will result in intraspecific differentiation. This differentiation is conveniently quantified with the F-statistics of Whright (1951), or the analogous measures of Cockerham (1969,1973). It leads naturally into a treatment of genetic distances.

In the basic model, the populations sampled are considered to derive from the same ancestral population. In the absence of disturbing forces, gene frequencies are expected to remain constant over all generations and all populations, while the variance of gene frequencies is expected to increase over time, due to the differenciation between populations. The comparison of mean frequencies between populations can be performed by an analysis of variance and provides an estimation of the different F-statistics (Cockerham, 1973). It is worth stressing that this measure of between-population differentiation is a consequence of the relatedness of genes within populations.

Similarly, the idea underlying the AMOVA approach is to split gene frequencies into independent random effects and to estimate their variance components.

Haploid data

One locus : analysis of gene frequencies Let x_{ijkl} be a indicator variable corresponding to the jth individual (j=1,...,n_i) in the sample from population i (i = 1,...,r) with $x_{ijkl} = 1$ if individual j has allele k at locus l, and $x_{ijkl} = 0$ else. It may be partitioned into hierarchical components, following a linear model

$$x_{ijkl} = x_{kl} + b_{ikl} + c_{ijkl} \qquad (1)$$

where the effects are b for populations and c for individuals within populations. The effects are assumed to be additive, random, uncorrelated, and to have the associated variance components $\sigma^2_{b_{kl}}$ and $\sigma^2_{c_{kl}}$, respectively. For each allele at each locus, a conventional sum of squared deviations from the mean SST_{kl} may be written and, following the

standard decomposition, we have

$$SST_{kl} = SSP_{kl} + SSW_{kl} \tag{2}$$

where SSP_{kl} is the sum of squared deviations between populations and SSW_{kl} is the sum of squared deviations within populations. The corresponding mean squared deviations (MSP_{kl} and MSW_{kl}) are obtained by dividing each sum of squares by the appropriate degrees of freedom, as given in Table 1. The variance components of each hierarchical level are extracted by equating the mean squares to their expectation. The structure of the analysis is that described for F-statistics (Cockerham, 1969, 1973). An unbiased estimate of the differentiation between populations is given by

$$\begin{aligned} \hat{\theta}_{kl} &= \frac{\hat{\sigma}^2_{b_{kl}}}{\hat{\sigma}^2_{b_{kl}} + \hat{\sigma}^2_{c_{kl}}} \\ &= \frac{MSP_{kl} - MSW_{kl}}{MSP_{kl} + (n_c - 1)MSW_{kl}} \end{aligned} \tag{3}$$

Multiple loci Under the neutral model, every allele at every locus provides an estimate of the same quantity and

$$\begin{aligned} \hat{\Phi}_{ST} &= \frac{\sum_{k,l} MSP_{kl} - \sum_{k,l} MSW_{kl}}{\sum_{k,l} MSP_{kl} + (n_c - 1)\sum_{k,l} MSW_{kl}} \\ &= \frac{\hat{\sigma}^2_b}{\hat{\sigma}^2_b + \hat{\sigma}^2_c} \end{aligned} \tag{4}$$

is equal to the weighted average of single-locus ratio estimator defined by Reynolds and al (1983). Note that $\hat{\sigma}^2_b = \sum_k \sum_l \hat{\sigma}_{b_{kl}}$ and $\hat{\sigma}^2_c = \sum_k \sum_l \hat{\sigma}_{c_{kl}}$ are estimates of variance components due to differences among populations (σ^2_b) and within populations (σ^2_c) averaged over the loci.

Analysis of Molecular Variance The analysis of molecular variance is a multilocus extension of the analysis of gene frequencies. It's treatment is based on the similarity between squared deviations from the mean and genetic distances. With haploid data, each individual is represented by one haplotype. Let x_{ij} be the vector of single locus allelic

states corresponding to the jth haplotype in the sample from population i. The euclidian genetic distance between two haplotypes is defined as the sum of squared differences between the x's over alleles and loci, weighted by a coefficient that allows us to deal with possible interactions among loci and unequal locus weighting schemes (Excoffier et al, 1992). If loci are assumed to be independent, they are given equal weight and the euclidian genetic distance between two haplotypes from the same population is given by

$$\delta^2_{(ij,ij')} = \sum_k \sum_l (x_{ijkl} - x_{ij'kl})^2 \tag{5.a}$$

while the euclidian genetic distance between two haplotypes from different populations is given by

$$\delta^2_{(ij,i'j')} = \sum_k \sum_l (x_{ijkl} - x_{i'j'kl})^2 \tag{5.b}$$

Then, using the equivalence between conventional sum of squares and sum of squared differences (Li, 1976), the sum of squared deviations averaged over the loci defined in (4) may be written as sum of genetic distances between and within populations

$$SST = \sum_k \sum_l SST_{kl} = \frac{1}{N} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \sum_{i'=1}^{i} \sum_{j'=1}^{n_{i'}} \delta^2_{(ij,i'j')} \tag{6.a}$$

and

$$SSW = \sum_k \sum_l SSW_{kl} = \sum_{i=1}^{r} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{j'=1}^{j} \delta^2_{(ij,ij')} \right) \tag{6.b}$$

with $N = \sum_{i=1}^{r} n_i$ being the total number of individuals. Hence, it is possible to make a partitioning of the distance matrix into hierarchical components : distances between haplotypes within populations and distances between populations. The analysis of variance performed on the euclidian genetic distances defined as in (5.a) and (5.b) provides us with estimates of the variances components $\hat{\sigma}^2_s$ and $\hat{\sigma}^2_c$, as well as with an estimate of the differenciation between the populations $\hat{\Phi}_{ST}$. The corresponding analysis of variance layout is given in Table 2. It is similar to the analysis of variance performed on the indicator variable x's and averaged over alleles and loci.

Diploid data

Although initially designed for haploid data, the extension of AMOVA to diploid data is straightforward. At the diploid level, there are two levels of relatedness of different genes within populations. The degree of relationship F between genes within individuals measures the total variance of gene frequencies between individuals. The degree of relationship θ between genes of different individuals measures the variance of gene frequencies between populations. Those F-statistics are related to each other according to $(1 - F) = (1 - F_{IS})(1 - \theta)$, with F_{IS} being the correlation of uniting genes relative to that of a random pair of genes from within the same population, classically defined as the inbreeding coefficient. Peakall et al (1995) performed single locus analysis to measure those components on allozyme data in buffalograss. The analysis can also be performed as an AMOVA on genetic distances. In this case, the definition of the genetic distance and the resulting statistics will depend on the markers.

Codominant markers with known gametic phase If the multilocus gametic phase is known, the analysis can be performed as an AMOVA on genetic distances between haplotypes. In this case, the distance matrix is to be partitioned into three hierarchical levels : distances between haplotypes within individuals, distances between individuals and distances between populations, with corresponding variance components σ_a^2, σ_b^2 and σ_c^2. The estimation of variance components provides us with two statistics,

$$\hat{\Phi}_{ST} = \frac{\hat{\sigma}_c^2}{\hat{\sigma}_a^2 + \hat{\sigma}_b^2 + \hat{\sigma}_c^2} \tag{7.a}$$

measures the differentiation between populations, and

$$\hat{\Phi}_{IS} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_a^2 + \hat{\sigma}_b^2} \tag{7.b}$$

measures the average correlation between uniting genes, relative to that of a random pair of genes from within the same population. As previously, those two statistics are weighted average of the corresponding single-locus ratio estimators θ and F_{IS}.

Codominant markers with unknown gametic phase If the markers are situated on physically linked loci, diploid individuals may be heterozygous at more than one locus and the gametic phase may be ambiguous. However, the gametic phase need not be known to estimate $\hat{\Phi}_{ST}$. Therefore, Michalakis and Excoffier (1995) proposed to define the gametic phase at random by creating two dummy haplotypes for each individual in each population. The AMOVA can then be performed on genetic distances between dummy haplotypes as in the haploid case described on Table 2, each of the $2N$ haplotypes being considered as one individual. The use of dummy haplotypes will yield correct estimates of population differentiation, even if the loci are statistically linked, as long as the loci are given equal weight (Weir and Cockerham, 1984). The genetic meaning of $\hat{\Phi}_{ST}$ depends on the definition of the genetic distances or, similarly, on the definition of the variables x_{ijkl}. With the x's being indicator variables, $\hat{\Phi}_{ST}$ estimates θ. With microsatellite data, the x's may be defined as allelic sizes, and the genetic distances as the sum of squared differences in allele size over all loci. In this case, $\hat{\Phi}_{ST}$ estimates Slatkin's R_{ST} (1995).

Dominant or multilocus markers With dominant markers, the allelic interpretation is no more possible. The simplest distance metric is the number of bands not shared between two individuals. Let M be the total number of band levels, and x_{ijk} be an indicator variable corresponding to the jth individual in the sample from population i, with $x_{ijk} = 1$ if individual j has band k and $x_{ijk} = 0$ else. Then, the distance between two individuals is equal to the sum of squared differences between the x's over the M bands. The same would apply for markers with unknown genetic determinism. The indicator variable x_{ijk} may be partitioned as previously into hierarchical components, following a linear model, and assuming the effects to be additive, random, uncorrelated. Performing AMOVA on the genetic distances between individuals will provide estimates of variance components σ_γ^2 and σ_ε^2. It is equivalent to performing a classical analysis of variance on bands frequencies. In this case however, the genetic interpretation is no more possible and the Φ-statistics are no more equivalent to the traditional F-statistics. Note that comparable results may be obtained with codominant markers by defining the x's as the

number of copies for each allele and by computing the genetic distances between the N individuals instead than between the $2N$ dummy haplotypes (Peakall et al, 1995).

General case In the general case, an individual I is defined by a vector X_I of M variables. The generalized distance between two individuals I and J is classically defined as

$$\delta_{IJ}^2 = (X_I - X_J)' W (X_I - X_J) \tag{8}$$

where W is a matrix of differential weights for the various variables. If W is the identity matrix, each variable is independent and has the same weight, and (8) reduces to Euclidian distance. If W is a diagonal matrix, the variables are still considered as independent, but are given a different weight. Given this general structure, the M variables may be molecular markers with the loci being given different weights. However, they can be any variable, including morphological traits with the W matrix being the variance-covariance matrix of the M traits. In this case, (8) is the Mahalanobis distance (Mahalanobis, 1936). The rest of the analysis does not depend on which particular form of W has been chosen and AMOVA is performed on the distance matrix partitioned into hierarchical components as in Table 2. The resulting $\hat{\Phi}_{ST}$ is a weighted measure of the corresponding single-variable ratio. It is not clear however wether the degrees of fredom defined in Table 2 are appropriate when W is a non diagonal matrix, *i.e.* when the variables are no more considered as independent.

Testing significance of the variance components and Φ-statistics

Generally, the distribution of the variables is unknown, especially when the x's are binary indicator variables. In this case, it is not possible to test the hypothesis $H_0 : \Phi_{ST} = 0$ with classical methods. In particular, the ratio of the mean squared deviation are not expected to have a F distribution under the null hypothesis. The testing procedure proposed for the AMOVA is permutational analysis of the null distribution for each variance component (Excoffier et al, 1995). It is an approximation of Fisher's exact test.

Under the null hypothesis, samples are considered as drawn from a global population, with variation due to random sampling in the construction of populations. Therefore, the idea is (i) to perform all the possible permutations of haplotypes between population, (ii) for each permutation, to estimate the variance components and the Φ-statistics, and (iii) to compare them to the corresponding statistics obtained with the real samples. The significance of the test is the probability that the statistics exceed the observed value under the null hypothesis. It may be estimated by the percentage of permutations for which the statistics exceeded the observed value. As sample sizes increase, the number of permutation becomes higher and higher and the exact test has to be replaced by the performing of a great number (say 1000) of random permutations. To obtain a null distribution, each haplotype is allocated to a new randomly chosen population, while holding the sample sizes constant at their realised value (Mantel, 1967).

If the individual loci composing the haplotypes are totally linked and the gametic phase is known, then permuting the haplotypes is the correct testing procedure. If the loci composing the haplotypes are statistically independent, the variation under the null hypothesis comes from both random sampling of the loci and random sampling in the construction of the populations. In this case, permuting haplotypes across populations is a conservative procedure and the significance levels will be overestimated. In the general case, the correct testing procedure probably depends on the amount of linkage disequilibrium between the loci and it is still to be found. As a matter of fact, the current testing procedures neglects the variance of linkage disequilibrium due to the statistical sampling of individuals within populations.

Application to distinction between varieties

Pairwise comparisons can be performed to test for a significant difference in gene frequencies between two varieties. Then, AMOVA is a multilocus alternative to the traditional computation of a Chi-square distance. Table 3 compares the two approaches with RAPD data obtained for 8 alfalfa varieties. It can be seen that in this case, AMOVA is slightly

more discriminant than Chi-square. The main advantage of AMOVA that the testing procedure is not sensitive to the existence of rare alleles or rare bands. Moreover, as no genetic interpretation is needed for distinction, it should be possible to compute genetic distances by giving different weights to the different loci or to the different bands. For example, loci situated on the same chromosome may be given a smaller weight, if the information is known. However, as discussed previously, the exact significance levels of the permutation test is generally unknown. This problem could be handled by simulation studies.

References

Cockerham C. C. (1969) Variances of gene frequencies. *Evolution* **23** : 72–84.

Cockerham C. C. (1973) Analysis of gene frequencies. *Genetics* **74** : 679–700.

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

Long J. C. (1986) The allelic correlation structure of Gainj- and Kalam-speaking people. II. The genetic distance between population subdividion. *Genetics* **112** : 629–647.

Mahalanobis P. C. (1936) On the generalized distance in statistics. *Proc. Nat. Inst. Sci. India* **2** : 49–55.

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

Michalakis Y. and L. Excoffier (1995) A generic estimation of population subdivision using distance between alleles with special reference for microsatellite loci. *Genetics* **142** : 1061–1064.

Peakall R., P. E. Smouse and D. R. Huff (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of dioecious buffalograss *Buchloë dactyloides*. *Molec Ecol* **4** : 135–147.

Reynolds J., B. S. Weir and C. C. Cockerham (1983) Estimation of the coancestry coefficient : basis for a short-term genetic distance. *Genetics* **105** : 767–779.

Slatkin M. (1995) A measure of population subdividion based on microsatellite allele frequencies. *Genetics* **141** : 141–154.

Weir B. S. and C. C. Cockerham (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38** : 1358–1370.

Wright S. (1951) The genetical structure of populations. *Ann. Eugen.* **15** : 323–354.

Table 1: Analysis of variance layout for variable indicating allele k at locus l

Source	d.f.	Sum of Squares	Expected Mean Square
Between populations	$r-1$	$\sum_i (x_{i.kl} - x_{..kl})^2$	$\sigma^2_{c_{kl}} + n_c \sigma^2_{b_{kl}}$
Within populations	$\sum_i (n_i - 1)$	$\sum_i \sum_j (x_{ijkl} - x_{i.kl})^2$	$\sigma^2_{c_{kl}}$

$n_c = \frac{1}{r-1}\left(\sum_i n_i - \frac{\sum_i n_i^2}{\sum_i n_i}\right)$

Table 2: Analysis of molecular variance

Source	d.f.	Sum of Squares	Expected Mean Square
Between populations	$r-1$	$\sum_{i=1}^{r}\sum_{j=1}^{n_i}\sum_{i'=1}^{i}\sum_{j'=1}^{n_{i'}} \frac{\delta^2_{(ij,i'j')}}{\sum_{i=1}^{r} n_i} - \sum_{i=1}^{r} \frac{\sum_{j=1}^{n_i}\sum_{j'=1}^{j} \delta^2_{(ij,ij')}}{n_i}$	$\sigma^2_c + n_c \sigma^2_b$
Within populations	$\sum_i (n_i - 1)$	$\sum_{i=1}^{r} \frac{\sum_{j=1}^{n_i}\sum_{j'=1}^{j} \delta^2_{(ij,ij')}}{n_i}$	σ^2_c

$n_c = \frac{1}{r-1}\left(\sum_i n_i - \frac{\sum_i n_i^2}{\sum_i n_i}\right)$

Distances between alfalfa varieties

	F1		F2		F3		A1		F4		F5		M1		M2	
F1			**0.0424**	(0.007)	0.0079	(0.213)	**0.1659**	(0.000)	0.0003	(0.440)	0.0022	(0.359)	-0.009	(0.764)	-0.007	(0.699)
F2	**14.69**	(0.023)			**0.0332**	(0.015)	**0.2164**	(0.000)	**0.0735**	(0.000)	0.0079	(0.231)	**0.0525**	(0.001)	**0.0685**	(0.002)
F3	7.99	(0.239)	**12.48**	(0.048)			**0.2092**	(0.000)	**0.0381**	(0.010)	0.0061	(0.263)	0.0061	(0.248)	**0.0367**	(0.010)
A1	**40.62**	(0.000)	**53.93**	(0.000)	**50.87**	(0.000)			**0.2371**	(0.000)	**0.1475**	(0.000)	**0.2294**	(0.000)	**0.1969**	(0.000)
F4	2.57	(0.766)	**15.77**	(0.008)	5.77	(0.329)	**44.36**	(0.000)			**0.0383**	(0.014)	-3E-04	(0.415)	0.0195	(0.065)
F5	4.14	(0.529)	6.5	(0.260)	7.34	(0.197)	**31.96**	(0.000)	8.74	(0.068)			**0.0264**	(0.036)	**0.0301**	(0.027)
M1	4.4	(0.620)	**14.07**	(0.029)	7.99	(0.239)	**60.59**	(0.000)	0.739	(0.981)	10.88	(0.054)			0.0022	(0.354)
M2	2.97	(0.812)	**17.75**	(0.007)	**14.2**	(0.028)	**55.84**	(0.000)	2.53	(0.772)	8.09	(0.151)	5.48	(0.484)		

Above diagonal : PHI_{ST}, below diagonal : Chi-square distance.
Significance levels are indicated between brackets.

Analysis of Molecular Variance

1. Introduction
2. Principles
3. Application for Distinction studies
4. Special cases
5. Testing procedure
6. Examples

Analysis of Molecular Variance

✧ Designed by L. Excoffier, P. E. Smouse and J. M. Quattro (1992)

✧ General framework for the study of molecular variation within species

✧ Application to distinction between synthetic or population varieties : Look for significant molecular variation between varieties

Analysis of Molecular Variance Principles

Haplotype x_I (0,0,1,0,0,1,1,0,0,1,1,1,0,1,0,1)

band indicator variable

Generalized squared distance between two haplotypes :

$$d_{IJ}^2 = (x_I - x_J)' W (x_I - x_J)$$

W = weight matrix

Analysis of Molecular Variance Principles

Hierarchical decomposition of the x's :

$$x_{ij} = x + b_i + c_{ij}$$

*i*th variety
*j*th individual

b = random variety effect
c = random individuals within variety effect

Analysis of variance performed on squared genetic distances provides estimates of variance components and Φ-statistics, reflecting the correlation of haplotype diversity at different levels of hierarchical subdivision.

367

Analysis of Molecular Variance layout

Source	df	Sum of Squares	Expected Mean Squares
between varieties	r-1	SSP	$\sigma_c^2 + n_c\sigma_b^2$
within varieties	$\Sigma_i(n_i - 1)$	SSW	σ_c^2

$$n_c = \frac{1}{r-1}\left(\Sigma_i n_i - \frac{\Sigma_i n_i^2}{\Sigma_i n_i}\right)$$

Analysis of Molecular Variance Principles

Equivalence between conventional sum of squares and sum of squared differences :

$$SST = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\sum_{i'=1}^{i}\sum_{j'=1}^{n_{i'}} d^2_{(ij,i'j')}$$

Measure of the differenciation between varieties :

$$\Phi_{ST} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_c^2}$$

Testing procedure :

random permutation of haplotypes between varieties

Analysis of Molecular Variance Application to distinction studies

✧ Two synthetic or population varieties.

✧ Test for the hypothesis H_0 = the two varieties are two samples of the same variety :

- Compute genetic distances between haplotypes, within and among varieties.

- Measure Φ_{ST} = correlation of haplotypes of the same variety relative to that of random haplotypes.

- Test for significance.

Analysis of Molecular Variance Special cases, diploid species

The meanings of the Φ-statistics depends on
- The variables of the vector x
  (treatment of the information)
- The definition of the genetic distance

W value	genetic distance
Identity matrix	Euclidian distance
Diagonal	Euclidian with a different weight for each locus
$(\text{variance-cov})^{-1}$	Mahalanobis distance

369

Special cases, diploid species
Treatment of the information

Information		Genetic meaning
Allelic interpretation, 2N dummy haplotypes		
YES/NO per allele, locus	(RFLP, enzymes)	Fst
Allelic size per locus	(microsatellites)	Rst
Band interpretation, N individuals		
YES/NO for each band	(RAPD)	?
Morphological data, N individuals		
Trait value for each trait		multivariate analysis of variance

Testing procedure

Current version of AMOVA :
random permutation of haplotypes (individuals)

- Insensitive to rare classes.

- Extra weight given to genotypic combinations.

- Exact significance level is not known.
  = depends on the statistical linkage between the loci

Alternative : random permutation of bands ?
-correct for independent loci.

Examples

(1) 20 varieties of diploid Raygrass
- English (5 varieties)
- Italian (14 varieties)
- Hybrid (1 variety)

3 enzymes already used for distinction.

(2) 8 varieties of Alfalfa
- European (5 varieties)
- Mediterranean (2 varieties)
- American (1 variety)

8 bands (RAPD)

Conclusion

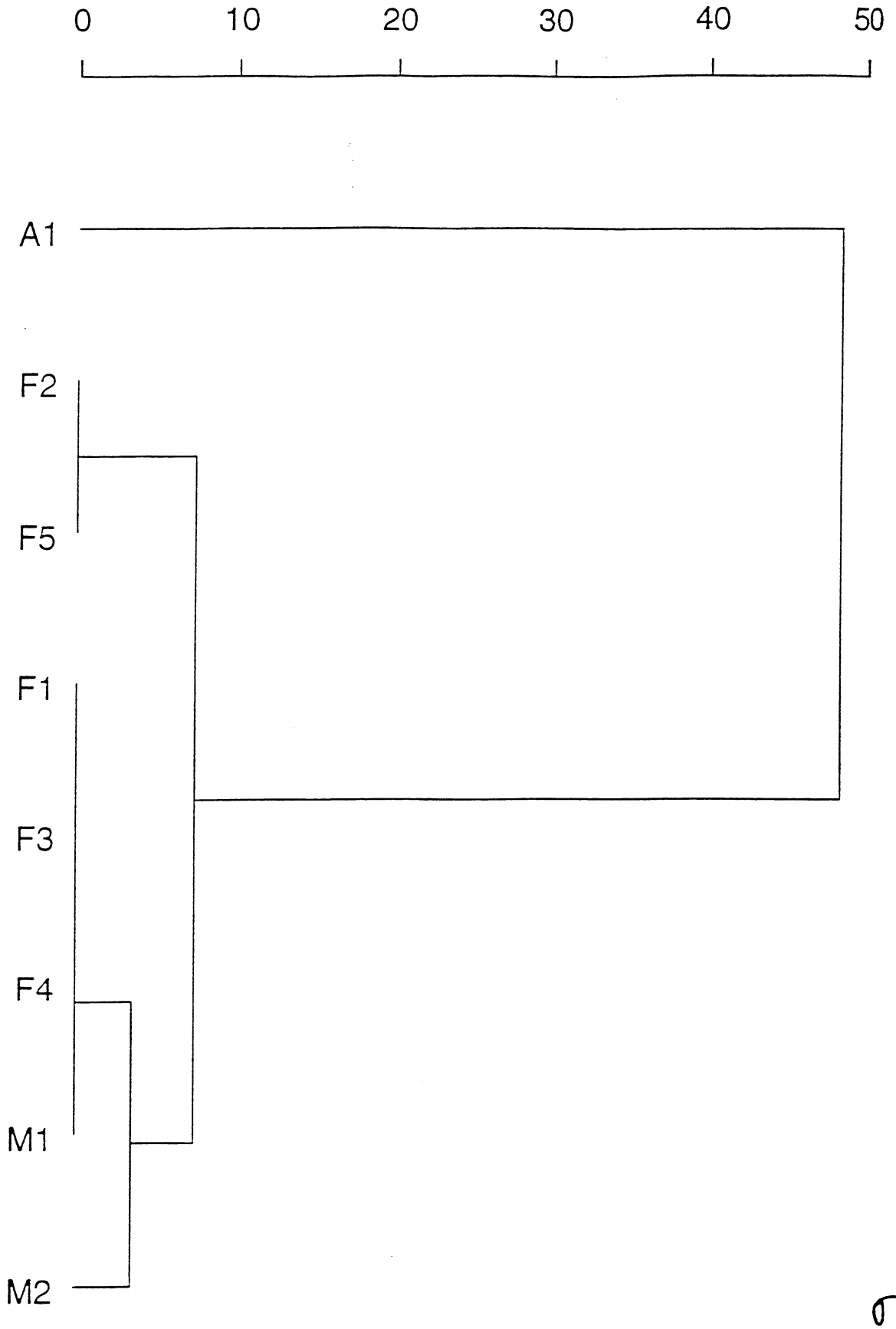
-AMOVA = multilocus alternative to the traditional computation of Chi-square distances

-Higher weight given to genotypic combinations

-Seems to be slightly more discriminant

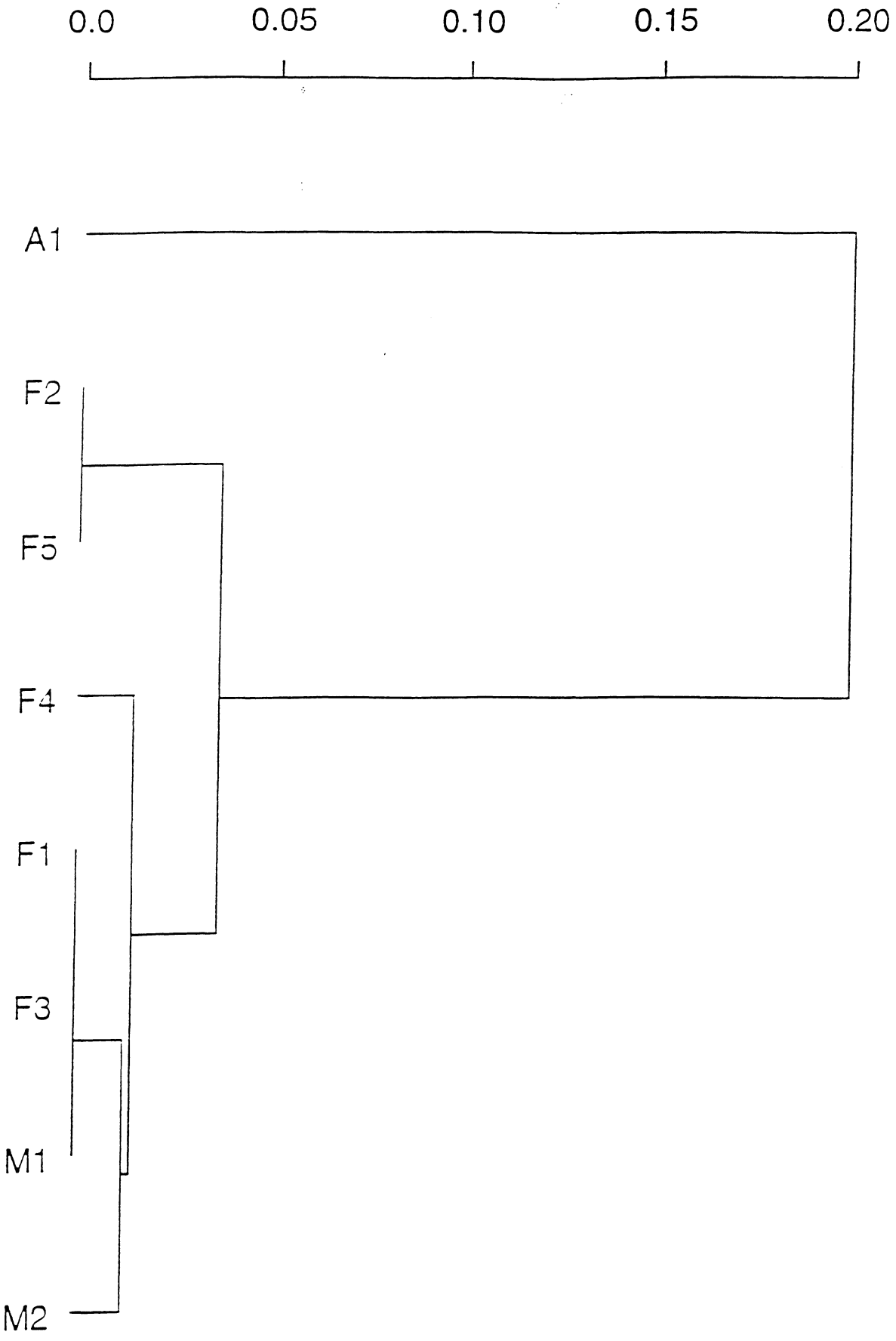
-Current testing procedure is not satisfying

371



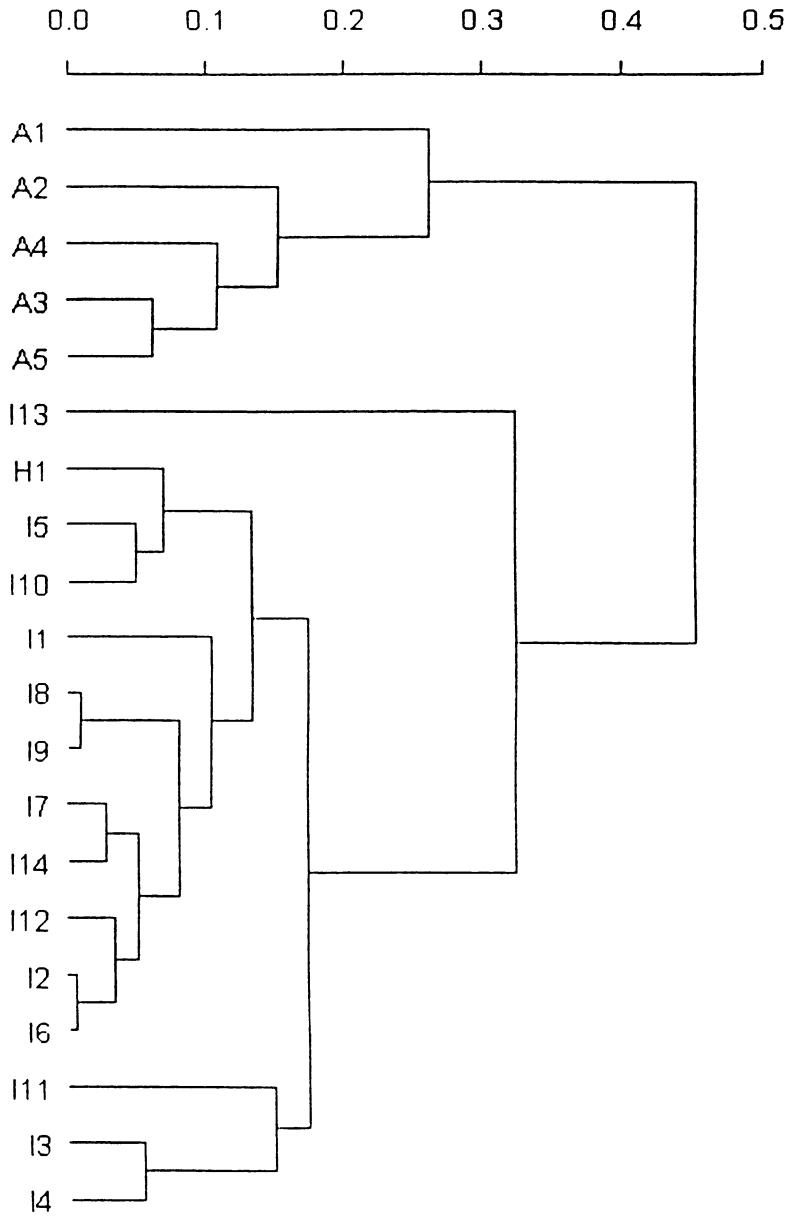
Alfalfa
Chi-square distance
6 bands

372 -



Alfalfa 1st 8 bares

Ray Grass



[End of document]