PERGAMON Biochemical Systematics and Ecology 29 (2001) 469–483

biochemical systematics and ecology

www.elsevier.com/locate/biochemsyseco

A review of the use of allozyme electrophoresis in plant systematics

Herman van der Bank[a,*], Michelle van der Bank[b], Ben-Erik van Wyk[b]

[a] *Department of Zoology, Rand Afrikaans University, P.O. Box 524, Auckland Park 2006, South Africa*
[b] *Department of Botany, Rand Afrikaans University, P.O. Box 524, Auckland Park 2006, South Africa*

Received 16 September 1999; accepted 10 June 2000

Abstract

The role of electrophoretic data is discussed as it applies to plant taxonomy and systematic studies. Nei's (Am. Nat. 106 (1972) 283–292; Genetics 89 (1978) 583–590) genetic distances calculated for a large number of populations, species and genera were taken from published data. The relation between Nei's genetic identity measures and taxonomic rank (populations, species and genera) are shown graphically. The graphs obtained in this way (from 3021 pairs of plant taxa) differ substantially from previous graphs published by Thorpe (Ann. Rev. Ecol. Syst. 13 (1982) 139–168; in: G.S. Oxford, D. Rollinson (Eds.), Protein Polymorphism: Adaptive and Taxonomic Significance, Academic Press, London, 1983, pp. 131–152) and Thorpe and Solé-Cava (Zool. Scripta 23 (1994) 3–18). These authors suggested that the divergence between the different taxonomic ranks is roughly similar across a wide range of taxa. The latter was based on values for 2664 (Thorpe, 1982) and 8060 (Thorpe, 1983) pairs of animal and plant taxa, but the plant data contributed little to the total. For any given taxonomic rank, we found that plants are genetically more closely related than animals (possibly with the exception of birds). This result is important because the empirical relationships of genetic distance measures, to different levels of taxonomic separation, is often used for distinguishing and identifying cryptic or sibling species where conventional methods are unable to resolve systematic problems.

Keywords: Plant systematics; Electrophoresis; Genetic distance

*Corresponding author. Tel.: +27-11-489-2450; fax: +27-11-489-2286.
E-mail address: fhvdb@na.rau.ac.za (H. van der Bank).

0305-1978/01/$ - see front matter
PII: S0305-1978(00)00086-7

1. Introduction

The ability to classify is a characteristic inherent in all of us. For centuries naturalists have tried to detect, describe, and explain diversity in the biological world. This endeavor is known as systematics (Moritz and Hillis, 1990). Classification is a result of a sequence of operations with the last step involving both the search for possible relationships among taxa together with the subsequent resolving of groups and their ranks. Biologists have shifted their interest to the species problem by generating data from many different disciplines. Botanists have made considerable efforts to integrate the information obtained from new data and to develop a consistent theory of speciation and evolution in plants (Stebbins, 1980). The first and major formulation came in 1950 with the publication of Stebbin's "Variation and evolution in plants" (Stebbins, 1950) which had a major impact on evolutionary theory. After that, more became known about the diversity of organisms, especially at the species level. This had an immense impact on classification and a class could no longer be defined by the invariable presence of certain common characteristics alone. Wittgenstein (1958) and Beckner (1959) later developed this concept, what we now call "polythetic classes". New techniques of data analysis, such as clustering, have changed classificatory ethics and practice in many sciences (Sokal, 1974), including biological classification. Essentialism, or what was known as typology, was substituted with empiricism (Baum, 1989). Unfortunately, many taxonomists who are both well aware of and who accept evolutionary theory still deny its relevance in taxonomy. Many taxonomists do not accept the relevance of evolutionary theory as a basis for the principles of classification, probably due to their predisposition for Aristotelian definitions for species that were never altered (Hull, 1965).

Systematics is so intimately linked to evolution that there should not be any definite boundary between them (Thorpe and Solé-Cava, 1994). However, there is a major distinction to be drawn between the evolutionary geneticist, who studies organisms with a view to understand the role and mechanisms of evolution, and the systematist. Systematists are mainly concerned with the evolution and interrelationships of particular plants insofar that this information will enable them to place populations, species, or other taxa more accurately within an overall systematic arrangement.

For a considerable period, studies of evolutionary genetics were dominated by the use of electrophoresis to investigate the biochemical evolution and the time-dependent structural divergence of protein molecules (Lewontin, 1974; Nei, 1987). More recently, other molecular methods of gathering evolutionary data have also become important, e.g. restriction fragment length polymorphism (RFLP) analysis (Avise and Lansman, 1983; Dowling et al., 1990; Hoelzel, 1992), DNA amplification and sequencing (Moritz and Hillis, 1990; Hoelzel, 1992). The molecular techniques have had a tremendous impact upon systematics and offer enormous potential if they are employed in conjunction with more traditional methods (Thorpe and Solé-Cava, 1994). Although molecular biology is not a cure-all for systematics, molecular systematics can lead to new insights into many problems previously encountered by traditional methods.

Appraisal of electrophoretic results has, necessarily, resulted from comparison with classifications and phylogenies independently derived from morphological, cytological, ecological, and behavioral data. Thorpe (1982, 1983) and Thorpe and Solé-Cava (1994) reviewed enzyme variation, genetic distance (D) and evolutionary divergence in relation to taxonomic separation in mammals, birds, reptiles, amphibians, fish, invertebrates and plants. Genetic identity (I) was employed as the primary measure of similarity among populations, infraspecific taxa, species and confamilial genera. Geneticists have used these measures extensively to define species boundaries (e.g. Van der Bank, 1996; Van der Bank and Kramer, 1996; Viljoen et al., 1996). However, the empirical relationships to different levels of taxonomic separation published by Thorpe and co-workers included limited comparisons of plant studies. Electrophoretic variation within and divergence among populations of only plant taxa will be considered and compared to results reported by the above-mentioned authors.

2. Methods

We have used similar methods as described in Thorpe (1982): (A) The assumption is made that, in most instances, different allozymes encoded by a locus are neutral or nearly neutral. If selection were acting strongly on allozymic diversity, then allozymes would not be nearly as useful systematically as they have proven to be. It is not argued, however, that all allozymes are always selectively equivalent. (B) Conversions of D to I values were made where the latter were not published. (C) Studies where there is taxonomic doubt on other grounds than electrophoresis were omitted. (D) Plants are treated as a single group since there are not enough data for lower plant species; the vast majority of the data are for various angiosperms. (E) Data for probabilities of I values falling into each 0.05 increment of I from 0.0 to 1.0 were plotted. (F) A running mean was used to reduce effects of sampling error (using the probabilities for each 0.05 increment of I as the mean probability over that and the two immediate adjacent increments of I). However, we have focussed on plant data only and compared the indices of Nei's (1972, 1978) I values in the present study. The null hypothesis test for two proportional values (Ferguson, 1976) was used to determine if significant ($P<0.05$) differences between Nei's (1972) and (1978) I values occur.

3. Results and discussion

A review of the advantages and disadvantages of electrophoretic data, and some of the major plant studies which have applied electrophoretic data to systematic problems may allow us to decide whether the advantages outweigh the disadvantages in practice, and whether electrophoresis is, in fact, a valuable systematic tool.

3.1. Advantages of electrophoretic data over more established systematic criteria

1. *Objectivity*: The numeration of alleles and their frequencies are objective determinations, based solely on the mobility of bands on gels. Subjectivity may occasionally enter into some morphological data or behavioral data. For example, interspecific relationships have been investigated by chemotaxonomy using a variety of low molecular weight chemicals (e.g. Bergquist and Wells, 1983; Bergquist et al., 1986). Unfortunately, such chemicals are mostly of unknown genetic derivation and may also undergo environmental modification. The use of chemical characters therefore presents many of the same problems experienced when using the more conventionally employed taxonomic characters. However, electrophoresis has the advantage that the molecule is of known genetic derivation: normally the direct product of a single locus. Also, in almost all cases, the molecular structure is solely genetically determined and likely to be free from environmental modification.
2. *Allozymes are almost invariably codominant*: Heterozygotes have different phenotypes from the homozygotes. Identifying the hetero- and homozygotes makes the calculation of gene frequencies very simple: these frequencies can then be used for the comparison of populations. In contrast to much of the morphological, numerical and chemical systematics, molecular systematics is directly linked to the way genes evolve and their distribution in the species.
3. *Weighting*: The initial weighting of characters is not a problem in electrophoretic data. Each locus is accorded equal value. A posteriori weighting may be practiced if some loci appear to be of more value than others in elucidating systematic relationships (Avise, 1974).
4. *Large data sets*: Isozyme data often constitute the largest existing genetic data sets for many organisms, both within and between species (Park and Moran, 1995).
5. *Inexpensive method*: It is relatively inexpensive, particularly when compared to other molecular genetic methods.
6. *It allows for quick processing time*: A laboratory can assay many hundreds of samples per day for many different loci and this translates into low labor costs as well.

3.2. Theoretical disadvantages of the electrophoretical approach to systematics

1. *Taxonomic limits*: Electrophoresis is considered to be ineffective for comparing organisms that are, in evolutionary terms very distant. These organisms are likely to be different at all, or most, loci. When two proteins have the same electrophoretic mobility it is not possible for the number of different amino acids to be detected. This has meant that electrophoretic data cannot be used to discriminate among various species when these species differ at all, or nearly all, loci (Ayala and Kiger, 1984). Furthermore, studies of population structure, breeding biology and other intraspecific applications require sufficient levels of intraspecific variability. This is often not the case. Occasionally, allozyme analysis is unable to reveal species-diagnostic alleles because of low genetic variability and

in many species there are not enough allozyme markers available for examining large portions of the genome (Liu and Furnier, 1993).

2. *Sampling limitations*: The biggest drawback of electrophoretic data for evolutionary studies is that enzyme comparisons are limited to extant species. In addition, research budgets often limit the number of individuals that can be sampled together with the number of enzyme systems and ultimately the number of loci that can be studied. However, according to Gorman and Renzi (1979) many loci rather than many individuals should be sampled to obtain reliable information for population or phylogenetic studies. Electrophoretic methods sample (primarily) water-soluble proteins encoded by structural genes. Since we have little idea what percentage of genes encode this type of protein, we cannot evaluate the degree of bias in a sample of the genome.
3. *Chance identity in band mobility*: There is a finite number of distinguishable band mobilities on a gel. The larger the number of species that are compared, the more likely is it that some species will appear to share alleles when, in actuality, they do not.
4. *Non-detected protein differences*: Many nucleotide changes may occur without altering the amino acid sequence, and many amino acid changes may occur without altering the net charge of the polypeptides (Avise, 1974). This bias, together with chance mobility, will cause underestimates of protein differences among populations.
5. *Other sources of phenotypic variation of isozymes*: Epigenetic events taking place after translation of nucleotide sequences into primary protein structure may form sub-bands of different mobility on the gel. This may then be misinterpreted as additional loci or variants of a particular locus. Furthermore, protease degradation of some proteins that is associated with repeated freezing and thawing or long- and short-term aging of samples may also form sub-bands of different mobility. However, by taking sufficient care when interpreting gel banding patterns (Moritz and Hillis, 1990) and by following the guidelines proposed by Ferreira et al. (1984), the above-mentioned problems can be overcome. The guidelines are: (1) consider gene expression in closely related organisms; (2) if a locus is expressed in different tissue, the variant phenotypes should be parallel among tissues; (3) the sub-unit construction of a protein influences its electrophoretic gel banding pattern; (4) no unexpected phenotypes should appear on the gels; and (5) phenotypic proportions in a sample should fit phenotypic proportions expected with random mating or Hardy–Weinberg proportions. The scoring of gels can be difficult for inexperienced researchers, which makes allozyme analysis more difficult than DNA sequence analysis (i.e. where automated programs are used).

3.3. Taxonomic uses of electrophoretic data in botany

Despite the fact that enzyme electrophoresis has not been widely employed in taxonomic studies, the data have been useful in certain instances. The following

examples will demonstrate this point. Jefferies and Gottlieb (1982) studied two diploid species of *Salicornia* that are nearly indistinguishable morphologically, to determine whether they represent two distinct species or one variable population. These two taxa have different habitats in salt marshes, but can also be found intermixed. Jefferies and Gottlieb (1982) detected no allozyme variation among individuals within each of the species but found the two taxa to be monomorphic for alternative alleles at six of 30 genes. No evidence of hybridization was detected, and the allozyme data revealed powerful evidence for recognizing two species. Crawford and Wilson (1979) and Crawford (1979) examined allozyme variation within and among several closely related species of *Chenopodium* that are diploid, annual weeds. Taxonomic treatments of the complex had varied, and certain taxa seemingly intergrade (morphologically) when found growing together. The electrophoretic data indicated that particular species viewed as doubtfully distinct do indeed contain alternative alleles at several genes, with no evidence of hybridization based on isozyme loci. Allelic data were also useful for showing that plants previously recognized as *C. incognitum* in reality represent two biological entities, each of which is conspecific with another species in the complex. On the other hand, Walters (1988) questioned the relationship between isozymic and morphologic variation in *Chenopodium* species. Allozyme analysis indicated a high degree of genetic similarity among populations of *C. neomexicanum* and *C. palmeri* ($I = 0.974$), but numerical analysis of morphological characters demonstrated that these two species are relatively distinct (Walters, 1988). Based on allozyme data from only eight isozymic loci Walters (1988) concluded that *C. palmeri* should be reduced to a variety of *C. neomexicanum*.

In summary, the use of electrophoretic data as an aid to taxonomy and systematics is essentially limited to two main areas: (a) to distinguish or to confirm species (alpha systematics), and (b) to measure divergence between populations at any level from within and between species from related genera (beta systematics or phylogenetics). Data relating to levels of differences between gene pools can also be used to give an indication of whether these are likely to be conspecific. Another area where allozyme electrophoresis probably has a great role to play is to estimate divergence between congeneric species or confamilial genera. It was suggested by Moritz and Hillis (1990) that allozyme analysis can be useful to compare groups that have divergence times of up to approximately 50 million years ago. It should also be borne in mind that allozyme data cannot prove that two morphs, subspecies, etc., are conspecific; it can only be concluded that no significant differences could be found. It is always possible, even though unlikely in many cases, that small but genuine differences are concealed by sampling error or that differentiation may be present at loci which have not been examined (Thorpe and Solé-Cava, 1994). However, if fixed differences are found, then the evidence is conclusive even when small sample sizes are involved. See Thorpe and Solé-Cava (1994) for calculations of probabilities and a discussion of the exceptional power of electrophoretic data to differentiate between populations.

Two methods devised to estimate genetic differentiation between taxa is genetic identity and genetic distance. Thorpe and Solé-Cava (1994) state that the level of genetic differentiation between two species or populations over a range of enzyme

loci may be reduced to a single figure using any of several published measures of genetic similarity or identity (measures of similarity) or of genetic distance (measures of dissimilarity). Allelic frequency data are usually employed to calculate either the similarity or the differences among plant populations, infraspecific taxa, species, etc., although the genetic distance measurements of Hendrick (1971) are derived from genotype frequencies. The genetic distance indices of Rogers' (1972) and Nei (1972; 1978) have been used extensively, and Nei's measures are those, in general, used in plant systematics. Each of these measures of genetic distance has mathematical properties that may influence its selection, use and interpretation. Nei's genetic distance coefficients are said to measure biological properties (i.e. the mean number of electrophoretically detectable substitutions per locus that have accumulated since the two populations diverged from the common ancestor). This coefficient can be corrected for error due to small sample size (Nei, 1978) but is non-metric (Farris, 1981) and it does not satisfy inequality. However, Rogers' similarity coefficient has no biological properties. Although it has not been modified to correct for small population size, it is metric and this has fewer theoretical restrictions on its use in certain clustering algorithms (Buth, 1984). Nei et al. (1983) compared five genetic distance measures and three clustering methods and concluded that the best combination to use is Nei's standard (1972) distance coefficient with UPGMA.

There should, within wide limits, be a general relationship between taxonomic separation and genetic divergence as estimated by Nei's or other measures because enzyme differentiation can be used to estimate genetic divergence between either populations or species (Thorpe and Solé-Cava, 1994). Several authors (Avise, 1974; Thorpe, 1979, 1982, 1983; Nei, 1987; Thorpe and Solé-Cava, 1994) have suggested a broad relationship between the two and it has been proposed that protein variation between populations could be used as a guide to their taxonomic relationships. Such data is useful when it is unclear whether two allopatric populations should be regarded as conspecific (Solé-Cava et al., 1991), or to help solve disputes over whether two species should be considered to be congeneric.

It was proposed by Thorpe (1979) that conspecific populations, congeneric species and species of different genera should, on average, have been isolated for different lengths of time and therefore have different probabilities of having given values of genetic identity (I). Different theoretical distribution curves were suggested. Using data from numerous publications on the frequency of different values of Nei's (1972) I for different levels of systematic divergence, Thorpe (1982) was able to plot empirical curves for probability against I values between confamilial genera, congeneric species and conspecific populations. Thorpe's results (Fig. 1a) cover data available for vertebrates, invertebrates and plants, but exclude birds, for which few studies were available. There is clearly a general relationship between taxonomic divergence and genetic distance. Thorpe (1982) concluded that genetic diversity within a species or genus appeared broadly comparable for most groups of vertebrates (except birds), invertebrates or plants. Predictably, there is considerable overlap of the range of I values between the lower part of the range for congeneric species and species of different genera, although the overlap between species and conspecific populations is surprisingly small.

Gottlieb (1977, 1981), Crawford (1983) and Giannasi and Crawford (1986) tabulated I values for conspecific populations, congeneric species and subspecific taxa of different plant taxa cited in the literature. These authors did not tabulate any data for confamilial genera since they anticipated that once higher taxa are being compared, a number of limitations become significant factors and that allelic data are valid primarily for comparing taxa within a genus or at best between closely related genera. Gottlieb (1981) calculated the mean I value for 21 pairs of congeneric plant species as 0.67 (± 0.04). Significantly, this represents a large decrease from the mean I between populations of a single species. For 13 self-fertilizing species, Gottlieb (1981) reported the mean $I=0.975$, whereas for outcrossers, the mean $I=0.956$. Similarly, Crawford (1983) reported high genetic similarities between conspecific populations with I values nearly always larger than 0.900. Crawford's tabulation of mean I values for 16 congeneric species, since Gottlieb's (1981) summary, showed the mean = 0.789, again supporting the view that populations of different species are considerably more differentiated genetically than are conspecific populations. These discrepancies lead to the present study, in which more recent data for plant taxa are included.

3.4. A review of genetic identities among plant taxa at three levels of divergence

I values for each of the three levels of taxonomic separation were broken down and classified into twenty 0.05 increments from 0 to 1 (Table 1). From these data, the frequencies of I values in each 0.05 increment of I were obtained and, hence, the probabilities associated with each increment. It is these probabilities which are plotted in Fig. 2. Total numbers of estimates of I used for Fig. 2c are 1572 for conspecific populations, 863 for congeneric species and 586 for confamilial genera. Publications used as sources of data are listed in Table 3.3 in Van der Bank (1999) and this table and the references can also be obtained from mvdb@na.rau.ac.za. Nei's (1978) I values were included because a recent survey of literature on electrophoresis indicated that authors prefer to use it. As can be seen in Fig. 2a, the critical value of I (Nei, 1972) necessary to distinguish between species and genera is around 0.60 ($I=0.35$; Thorpe, 1982; Thorpe and Solé-Cava, 1994). About 81.8% of I values between congeneric species exceed 0.60 (71.4% above 0.65 and 87.1% above 0.55), whereas all I values (100%) are below 0.60 between genera. At the upper end of the range about 84.7% of I values (between species) are below 0.90 while within species 68.1% exceed 0.90 ($I=97\%$ below 0.85 and 93% above 0.9, respectively; Thorpe, 1982; Thorpe and Solé-Cava, 1994). Approximately, 58% of conspecific I values are above 0.95 compared to 80% at this level reported by Thorpe (1982) and Thorpe and Solé-Cava (1994). There is a degree of overlap in the range of I values among all three of the different levels of systematic divergence (Table 2).

When using Nei's (1978) measurements, the critical I value to distinguish between species and genera is higher than for Nei's (1972) index (± 0.75). About 63.7% of I values between congeneric species exceed 0.75 (52% above 0.80 and 72.8% above 0.70) whereas about 91.5% fall below 0.75 (93.1% below 0.80, and 89.5% below 0.70) between genera (Table 1). At the upper end of the range nearly 71.8% of I

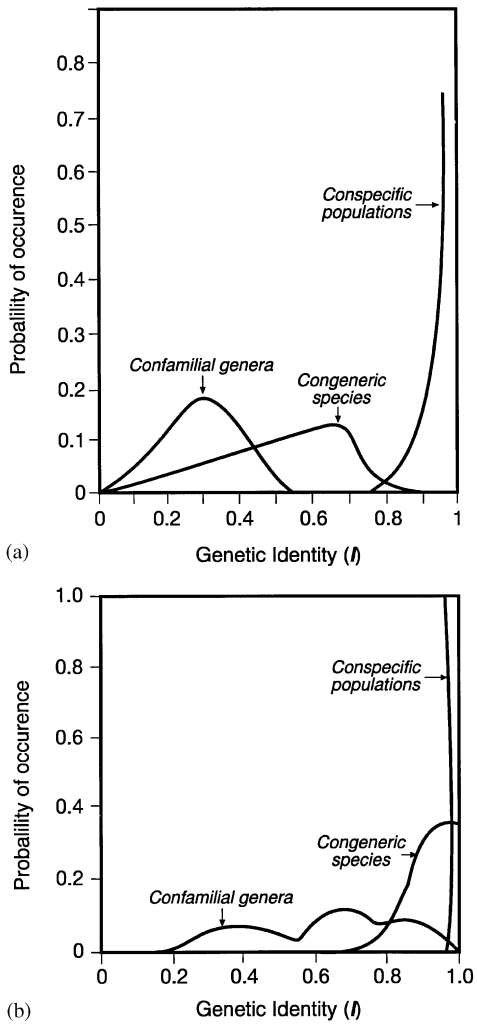

Fig. 1. (a) The distribution of genetic identities (Nei, 1972) between pairs of related taxa (after Thorpe, 1982, 1983) for all plants and animals excluding birds, (b) As for (a) but for bird data only.

values between species are below 0.90 compared to 86.9% that exceed 0.90 within species. About 65.5% of conspecific I values are above 0.95 (Fig. 2b).

Although different approaches are involved in the formulation of similarity coefficients they generally give rather similar summaries of the information contained in the allelic distribution (Avise, 1974). Avise (1974) compared 315 pair-wise populations by using similarity coefficients of Sokal and Sneath (1963), Stewart (in Rogers, 1972), Rogers (1972) and Cavalli-Sforza and Edwards (1967). A high correlation was found between all these different methods. In the current study no significant ($P<0.05$) differences were found between Nei's (1972) and (1978) I values

Table 1
Empirical frequencies of *I* values (Nei, 1972, 1978; and combined) determined at 0.05 increments from 0 (total dissimilarity) to 1 (total similarity) for genera from the same family, species from the same genus and conspecific populations. Percentage values per frequency are shown in brackets

Increment	Confamilial genera			Congeneric specific			Conspecific populations		
	Nei (1972)	Nei (1978)	Combined	Nei (1972)	Nei (1978)	Combined	Nei (1972)	Nei (1978)	Combined
0.1–0.15		5 (0.8)	5 (0.8)						
0.15–0.2		3 (0.5)	3 (0.5)		5 (0.8)	6 (0.6)			
0.2–0.25		12 (1.9)	12 (1.9)	1 (0.3)	1 (0.3)	4 (0.4)			
0.25–0.3	2 (20)	7 (1.0)	9 (1.4)	7 (2.1)	7 (1.1)	12 (1.2)			
0.3–0.35	1 (10)	30 (5.1)	31 (5.1)	8 (2.6)	5 (0.8)	13 (1.4)			
0.35–0.4	4 (40)	37 (6.3)	42 (7.1)	5 (1.5)	7 (1.2)	12 (1.3)			
0.4–0.45	2 (20)	57 (9.7)	59 (9.9)	2 (0.6)	9 (1.5)	11 (1.1)			
0.45–0.5	1 (10)	79 (13.6)	79 (13.4)	7 (2.3)	13 (2.2)	20 (2.1)			
0.5–0.55		57 (9.8)	57 (9.7)	3 (1.0)	17 (2.8)	20 (2.2)			
0.55–0.6		86 (14.7)	86 (14.5)	16 (5.3)	24 (4.1)	40 (4.6)		4 (0.4)	4 (0)
0.6–0.65		89 (15.3)	89 (15.0)	31 (10.4)	22 (3.8)	53 (6.0)	1 (0.1)	4 (0.4)	5 (0)
0.65–0.7		63 (10.8)	63 (15.0)	29 (9.8)	36 (6.3)	65 (7.5)	3 (0.3)	2 (0.2)	5 (0.1)
0.7–0.75		12 (2.0)	15 (2.4)	27 (9.0)	52 (9.1)	79 (9.1)	4 (0.4)	7 (0.7)	13 (0.4)
0.75–0.8		10 (1.6)	7 (1.1)	45 (15.3)	66 (11.7)	111 (12.9)	20 (3.0)	3 (0.3)	22 (1.2)
0.8–0.85		17 (2.6)	16 (2.6)	39 (13.1)	86 (15.2)	125 (14.5)	27 (4.0)	21 (2.1)	58 (3.4)
0.85–0.9		2 (0.3)	6 (0.9)	34 (11.4)	62 (10.9)	96 (11.2)	74 (11.8)	70 (5.6)	121 (7.9)
0.9–0.95		10 (1.5)	6 (0.9)	25 (8.4)	50 (8.8)	75 (8.6)	124 (20.1)	197 (21.4)	338 (22.1)
0.95–1.0		1 (0.1)	1 (0.1)	13 (4.4)	96 (17.1)	109 (12.7)	354 (58.0)	598 (65.5)	945 (62.2)

between the data for conspecific populations and congeneric species using the null hypothesis test for two proportional values. *I* values for confamilial genera is not as well separated for Nei's (1978) graph (Fig. 2b) compared to Nei's (1972) values (Fig. 2a). This is because Nei's (1978) *I* values are usually larger for a given pair of taxa compared to Nei's (1972) estimate for the same pair.

The combined data set is depicted in Fig. 2c. The critical *I* value that is necessary to distinguish between species and genera is approximately 0.75, and 0.90 between species and populations. About 59.9% of *I* values between congeneric species exceed 0.75 and almost 92.3% fall below 0.75 between genera, and 84.3% exceed 0.90 within species. Differences were observed when comparing this data on plants with those obtained by Thorpe (1982, 1983) and Thorpe and Solé-Cava (1994). For any given taxonomic rank, plant species appear to be, in general, genetically less distinct than predicted by Thorpe (1982, 1983) and Thorpe and Solé-Cava (1994). It is important to note that the data used by Thorpe was heavily biased towards vertebrates and invertebrates and also towards organisms native to North America. Only approximately 7% of the publications used by Thorpe (1982) to construct his graph (Fig. 1) were based on plants. In the present study 3021 pair-wise comparisons for only plant species taken from the literature were used to construct Fig. 2. The number of estimates of *I* used to construct Fig. 1 is 7000 for conspecific populations,

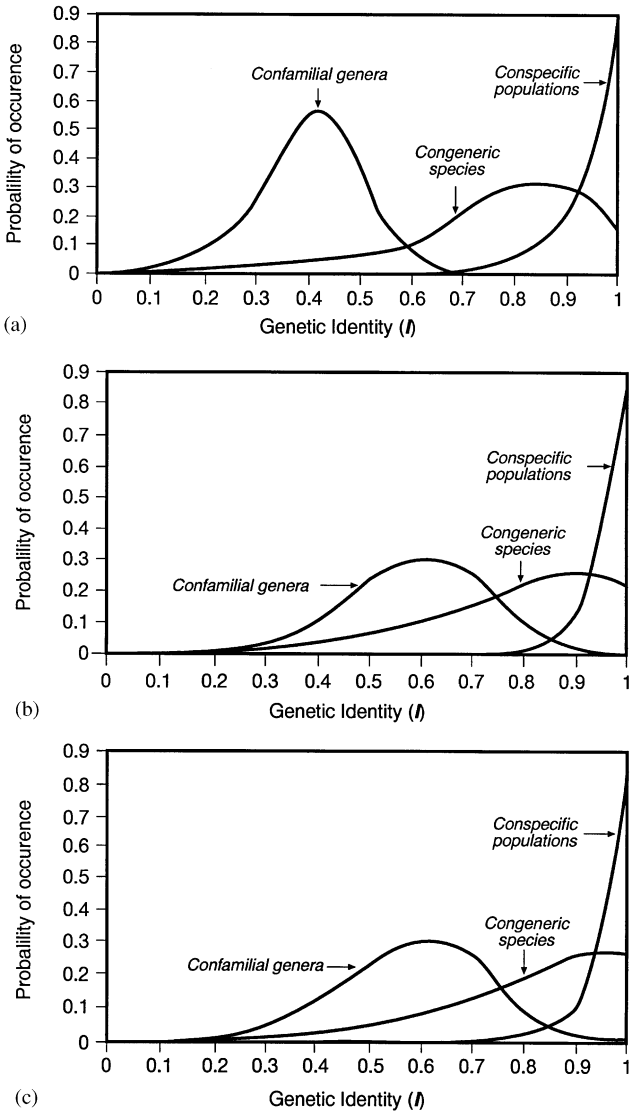

Fig. 2. The results of this study, based on calculations using data from the literature. The distribution of: (a) Nei's (1972), (b) Nei's (1978), and (c) combined data from Nei's (1972, 1978) *I* values (3021 data points). Note the large degree of overlap in the *I* values of genera, species and populations compared to the more discrete values reported by Thorpe (1982, 1983) — as shown in Fig. 1a.

900 for congeneric species and 160 for confamilial mammal, bird, reptile, amphibian, fish, invertebrate and plant genera (Thorpe, 1983). The data obtained for plants show that a relative high percentage of *I* values between species fall within the range typical of conspecific populations in other taxa. It would appear that in plants

Table 2
Descriptive statistics of genetic identities (Nei, 1972, 1978; and combined)

	Confamilial genera			Congeneric species			Conspecific populations		
	Nei (1972)	Nei (1978)	Combined	Nei (1972)	Nei (1978)	Combined	Nei (1972)	Nei (1978)	Combined
N	10	576	586	280	583	863	625	947	1 572
Sum	3.665	313.779	317.444	208.143	463.260	671.403	590.253	902.543	1 492.796
Min.	0.273	0.111	0.111	0.182	0.167	0.167	0.642	0.576	0.576
Max.	0.455	0.980	0.980	0.993	1.000	1.000	1.000	1.000	1.000
Mean	0.366	0.544	0.541	0.743	0.794	0.778	0.944	0.953	0.950
Standard deviation	0.061	0.146	0.146	0.175	0.179	0.180	0.062	0.056	0.059
Standard error	0.019	0.006	0.006	0.105	0.007	0.006	0.002	0.001	0.002

speciation (or at least enzyme evolution) occurs at drastically faster rates than in other taxa (quantum speciation in the concept of Grant, 1981) or alternatively that the genetic distance molecular clocks run at a much reduced rate. It was surprising that there is no clear distinction between the data sets for confamilial genera and the other taxonomic levels (Figs. 2a–c). This is probably a result of the low taxonomic resolution of the technique at this (high) level. It should also be noted that there are less allozyme studies on this level. For example, only 10 *I* (Nei, 1972) values could be found (Table 2). This result is important because all previous comparisons of plant data were done with a graph (Fig. 1) that is clearly not representative of data for plants in general.

In conclusion, it is evident from the present study that plants, like birds (Fig. 1b), differ substantially from other groups. The molecular systematist should always keep two points in mind regarding the use of distance coefficients in systematic studies. Firstly, it is essential that calculation of a genetic distance must be based upon a random sample of individuals and loci, without prior knowledge of the percentage of polymorphic loci or the degree of differentiation between the taxa. Secondly, it is important to remember that genetic distance is not an abstract and idealized measure of ''differentness''. It is an estimate of a parameter of the model that is thought to have generated the differences observed. As an estimate, it has statistical properties that can be assessed and helps the systematist (as clearly indicated in this review) to reconstruct phylogenies (Skelton, 1993), and the biochemical criteria should only be employed where alternative methods that have proven unsatisfactory (Thorpe, 1983).

Appendix

Allopatric populations: Geographically separated populations.

Allozymes: Different forms of enzymes encoded by different alleles at the same locus.

Epigenetic effects: Non-genetic causes of a phenotype.

Genetic identity (*I*): Many computational methods have been devised to construct molecular phylogenies based on various similarity coefficients. Values of *I* range from 0.0 (total dissimilarity) to 1.0 (identity).

Genetic distance (*D*): Values of (*D*) range from 0.0 (no genetic distance or identical) to infinity. The useful range of *D* is only from 0.0 to about 1.5; beyond this range *D* is compressed because of the inability of electrophoretic methods to detect multiple mutation events.

Null hypothesis test: The null hypothesis under test assumes that there is no real difference between the true value of *p* (probability) in the population from which we sampled and the hypothesized value of $p = 0.5$.

Similarity coefficients: See genetic identity.

UPGMA: Unweighted pair group method with arithmetic mean. One of several computational methods have been devised to construct molecular phylogenies from comparative data.

References

Avise, J.C., 1974. The systematic value of electrophoretic data. Syst. Zool. 23, 465–481.

Avise, J.C., Lansman, R.A., 1983. Polymorphism of mitochondrial DNA in populations of higher animals. In: Nei, M., Koehn, R.K. (Eds.), Evolution of Genes and Proteins. Sinauer, Sunderland, MA, pp. 147–164.

Ayala, F.A., Kiger, J.A., 1984. Modern Genetics,, 2nd Edition.. Benjamin/Cummings, California.

Baum, B.R., 1989. Numerical taxonomic analyses of the *Gramineae*. In: Soderstrom, T., Hilu, K.W., Campbell, C., Barkworth, M.E. (Eds.), Proceedings of the International Symposium on Grass Syst. Evol. Smithsonian Inst. Press, Washington, DC.

Beckner, M., 1959. The Biological Way Of Thought. Columbia University Press, New York.

Bergquist, P.B., Wells, R.J., 1983. Chemotaxonomy of the *Porifera*: the developments and current status of the field. In: Scheuer, P.J. (Ed.), Marine Natural Products. Academic Press, London, pp. 1–50.

Bergquist, P.B., Lavis, A., Cambie, R.C., 1986. Sterol composition and the classification of *Porifera*. Biochem. Syst. Ecol. 14, 105–112.

Buth, D.G., 1984. The application of electrophoretic data in systematic studies. Ann. Rev. Ecol. Syst. 15, 501–522.

Cavalli-Sforza, L.L., Edwards, A.W.F., 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21, 550–570.

Crawford, D.J., 1979. Allozyme variation in *Chenopodium incanum*: intraspecific variation and comparison with *Chenopodium fremontii*. Bull. Torrey Bot. Club. 196, 257–261.

Crawford, D.J., 1983. Phylogenetic and systematic inferences from electrophoretic studies. Part A. In: Tanksley, O.S., Orton, T.J. (Eds.), Isozymes in Plant Genetics and Breeding. Elsevier, Amsterdam, pp. 257–287.

Crawford, D.L., Wilson, H.D., 1979. Allozyme variation in several closely related diploid species of *Chenopodium* of the western United States. Am. J. Bot. 66, 237–244.

Dowling, T.E., Smith, G.R., Brown, W.M., 1990. Reproductive isolation and introgression between *Notropis cornutus* and *Notropis chrysocephalus* (family Cyprinidae): comparison of morphology, allozymes and mitochondrial DNA. Evolution 43, 620–634.

Farris, J.S., 1981. Distance data in phylogenetic analysis. In: Funk, V.A., Brooks, D.R. (Eds.), Advances in Cladistics. New York Botanical Gardens, New York, pp. 3–23.

Ferreira, J.T, Grant, W.S., Avtalion, R.R., 1984. Workshop on Fish Genetics. CSIR Special Publication, CSIR, Pretoria, South Africa, pp. 1–72.

Ferguson, G.A., 1976. Statistical Analysis in Psychology and Education. McGraw-Hill, London, pp. 175–201.

Giannasi, D.E., Crawford, D.J., 1986. Biochemical systematics 2. A reprise. Evol. Biol. 20, 25–248.

Gorman, G.C., Renzi, J., 1979. Genetic distance and heterozygosity estimates in electrophoretic studies: effects of sample size. Copeia 2, 242–249.

Gottlieb, L.D., 1977. Evidence for duplication and divergence of the structural gene for phosphoglucoisomerase in diploid species of *Clarkia*. Genetics 86, 289–307.

Gottlieb, L.D., 1981. Electrophoretic evidence and plant populations. Progr. Phytochem. 7, 1–46.

Grant, V., 1981. Plant Speciation, 2nd Edition.. Columbia University Press, New York.

Hendrick, P.W., 1971. A new approach to measuring genetic similarity. Evolution 25, 276–280.

Hoelzel, A.R., 1992. Molecular Genetic Analysis of Populations. Oxford University Press, Oxford.

Hull, D.L., 1965. The effect of essentialism on taxonomy — two thousand years of stasis 1,2. Brit. J. Phil. Sci. 15, 314–326; 16, 1–18.

Jefferies, R.L., Gottlieb, L.D., 1982. Genetic differentiation of the microspecies *Salicornia europea* L. (*sensu stricto*) and *S. ramosissima* J. Woods. New Phytol. 92, 123–129.

Lewontin, R.C., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press, New York.
Liu, Z., Furnier, G.R., 1993. Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor. Appl. Genet. 87, 97–105.
Moritz, C., Hillis, D.M., 1990. Molecular systematics: context and controversies. In: Hillis, D.M., Moritz, C. (Eds.), Molecular Systematics. Sinauer Associates. Sunderland, MA, pp. 1–10.
Nei, M., 1972. Genetic distance between populations. Am. Nat. 106, 283–292.
Nei, M., 1978. Estimation of average heterozygosity and genetic distance from small number of individuals. Genetics 89, 583–590.
Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.
Nei, M., Tajima, F., Tateno, Y., 1983. Accuracy of estimated phylogenetic trees from molecular data. J. Molec. Evol. 19, 153–170.
Park, L.K., Moran, P., 1995. Developments in molecular genetic techniques in fisheries. In: Carvalho, G.R., Pitcher, T.J. (Eds.), Molecular Genetics in Fisheries. Chapman & Hall, London, pp. 1–27.
Rogers, J.S., 1972. Measure of genetic similarities and genetic distance. Studies in Genetics VII. University of Texas Publication. 7213, 145–153.
Skelton, P., 1993. Evolution, a Biological and Paleontological Approach. Addison-Wesley, Wokingham, England.
Sokal, R.R., 1974. Classification: purpose, principles, progress, prospects. Science 185, 1115–1123.
Sokal, R.R., Sneath, P.H.A., 1963. Principles of Numerical Taxonomy. Freeman, San Francisco.
Solé-Cava, A.M., Thorpe, J.P., Manconi, R., 1991. A new species of Mediterranean *Axinella* (Porifera-Demospongiae) detected by isozyme electrophoresis. In: Reitner, J., Keupp, H. (Eds.), Fossil and Recent Sponges. Springer, Berlin, pp. 313–321.
Stebbins, G.L., 1950. Variation and Evolution in Plants. Columbia University Press, New York.
Stebbins, G.L., 1980. Botany and the synthetic theory and evolution. In: Mayr, E., Province, W.B. (Eds.), The Evolutionary Synthesis, Perspectives on the Unification of Biology. Harvard University Press, Cambridge, MA, pp. 139–152.
Thorpe, J.P., 1979. Enzyme variation and taxonomy: the estimation of sampling errors in measurements of interspecific genetic similarity. J. Linn. Soc. (Biol.) 11, 369–386.
Thorpe, J.P., 1982. The molecular clock hypothesis: biochemical evaluation, genetic differentiation and systematics. Ann. Rev. Ecol. Syst. 13, 139–168.
Thorpe, J.P., 1983. Enzyme variation, genetic distance and evolutionary divergence in relation to levels of taxonomic separation. In: Oxford, G.S., Rollinson, D. (Eds.), Protein Polymorphism: Adaptive and Taxonomic Significance. Academic Press, London, pp. 131–152.
Thorpe, J.P., Solé-Cava, A.M., 1994. The use of allozyme electrophoresis in invertebrate systematics. Zool. Scripta 23, 3–18.
Van der Bank, F.H., 1996. Inter- and intraspecific allozyme comparisons of mormyrids (Pisces, Mormyridae) from South Africa and Namibia, with reference to an undescribed species. Water S. A. 22, 285–290.
Van der Bank, M., 1999. Genetic variation and phylogenetic relationships in the Podalyrieae and related tribes (Fabaceae). Ph. D. Thesis, Rand Afrikaans University, South Africa.
Van der Bank, F.H., Kramer, B., 1996. Phylogenetic relationships between eight African species of mormyriform fish (Teleostei, Osteichthyes): resolution of a cryptic species, and reinstatement of *Cyphomyrus*. Biochem. Syst. Ecol. 24, 275–290.
Viljoen, A.M., Van Wyk, B.-E., Van der Bank, F.H., Smith, G.F., Van der Bank, M., 1996. A biochemical and Chemical evaluation of the species concept of *Aloe candelabrum* (Aloaceae). Taxon 45, 461–471.
Walters, T.W., 1988. Relationships between isozymic and morphologic variation in the diploids *Chenopodium fremontii*, *C. neomexicanum*, *C. palmeri*, and *C. watsonii*. Am. J. Bot. 75, 97–105.
Wittgenstein, L., 1958. Preliminary Studies for the "Philosophical Investigations", generally known as The Blue and Brown Books. Oxford, Blackwell.