Acta Bot. Neerl. 44(4), December 1995, p. 403–419

Progress and pitfalls in systematics: cladistics, DNA and morphology*

KONRAD BACHMANN

Hugo de Vries Laboratory, University of Amsterdam, Kruislaan 318, NL-1098 SM Amsterdam, The Netherlands

SUMMARY

The 'natural system' of organisms reflects their phylogenetic relationship. It is the result of an historical process and has to be inferred from the available evidence. In the morphological phenotype, historical traces are intermeshed with functional adaptations. Overall similarity, even if quantified, can be a misleading indicator of relatedness. Cladistics uses shared derived character states (synapomorphies) to identify groups of common ancestry. Synapomorphies are mostly inferred from their taxonomic context. If apparently equally valid characters suggest mutually exclusive groups, parsimony is invoked: a phylogenetic reconstruction requiring a minimum of evolutionary steps to describe the present character distribution is accepted as the most likely one. Cladistics sets very stringent requirements for informative characters, and a rigorous analysis of morphology is likely to yield very few reliable characters. The direct analysis of DNA sequences provides theoretically the optimal evidence for phylogenetic reconstruction. In practice, very little of this information is readily accessible. Occasionally major sequence rearrangements can be unequivocal synapomorphies. Many phylogenetic problems can be solved by comparative sequencing of an appropriate segment of DNA. Comparative sequencing of the chloroplast gene *rbc*L has become the model for such studies. Molecular data have confirmed much traditional taxonomy, elucidated doubtful cases and corrected misinterpretations. Molecular data also have clearly shown the limits of the cladistic approach by revealing both known and previously unsuspected modes of reticulate evolution. Molecular approaches separate phylogenetic reconstruction from biological evaluation and will never replace morphological analysis in Systematics. However, molecular methods also facilitate the direct investigation of morphological evolution by revealing the genetic basis of morphogenesis in model systems or by permitting the genetic analysis of diagnostic character changes by genetic mapping.

Key-words: cladistics, homology, molecular systematics, parsimony, QTL mapping, reticulate evolution.

---

*This paper is based on a lecture presented at a symposium on 'Plant and Evolution' held on 18 May 1995 in celebration of the 150th anniversary of the Royal Botanical Society of The Netherlands.

INTRODUCTION

When Darwin showed that the 'natural system' of organisms is the result of descent with modification, his explanation had surprisingly little impact on the practice of Systematics. With the benefit of hindsight, we see now that many of the present problems in Systematics could have been avoided if our predecessors had been less adept in assimilating the revolutionary aspects of evolutionary theory into a tradition that matured a century before Darwin. Darwin destroyed any lingering hope that the natural system might reflect some basic principle of order in nature and that it could be derived from first principles. If the pattern of biodiversity is the result of an historical process and therefore one of infinitely many equally probable ones, the only way to arrive at a 'natural' classification is the painstaking case-by-case reconstruction of phylogenetic relationships from the remaining evidence. With the exception of fossils, this evidence consists of the traces of history preserved in the structure of living organisms. Darwin also showed why much of this evidence is misleading. When selection and adaptation shape the appearance of organisms, morphological similarities due to common descent (homologies) become intermeshed with similarities due to common functional demands (convergencies, parallelisms). Identifying homologous characters and evaluating their similarities has been a major task of Systematics (Donoghue 1992). Even with the guidance of independent criteria (Remane 1952; Patterson 1988; see below for molecular data) this has regularly involved preconceived ideas about phylogenetic relationships and an iterative testing of the consistency of mutually dependent pieces of evidence. However, a plausible and internally consistent scenario is not necessarily the true story, and it is a tribute to the skill of experienced taxonomists that the classical taxonomy of plants and animals seems to be in general a good reflection of their true phylogenetic relationships.

We know this, because lately objective methods of taxonomical analysis have been developed and previously inaccessible data have become available. Together, these permit a statistical estimate of the consistency and especially of the stability of phylogenetic reconstructions and this can be used to evaluate alternative hypotheses with a measurable degree of confidence.

THE SEARCH FOR OBJECTIVE METHODS

An obvious approach towards an objective method in Systematics is the conversion of qualitative observation into quantitative measurement. However, quantitative determinations of overall phenotypic or genetic similarity (or distance) are fraught with problems to the extent that I shall not discuss them here further. There are circumstances under which trees based on a hierarchical ordering of some sort of distance measurement can be excellent reflections of the phylogenetic relationships of the groups in question. However, various applications of genetic distance determinations depend on assumptions about character evolution that are either unproven or demonstrably wrong in individual cases. This includes not only the assumption that morphological differences are the result of a continuous accumulation of numerous independent small gene effects and that they increase more or less in proportion to the time of divergence from a common ancestor, it concerns also the hope that with a large number of measurements similarities due to convergent or parallel evolution will cancel out

while similarities due to common descent will reinforce each other and increase the phylogenetic signal in the data.

A different approach, the much discussed cladistics, in its most basic form, disregards quantitative distances entirely. Cladistic analysis (Hennig 1966; Wiley 1981; Wiley *et al.* 1991; Forey *et al.* 1992) works with selected individual informative characters from which it reconstructs the phylogenetic history of extant groups as the most probable sequence of character changes that explains the present character state distribution. To be informative, a character should ideally have only two character states, an original one (a plesiomorphy) and one that is derived from it (an apomorphy). Also, the derived character state has to be present in at least two of the species under study. Such a shared derived character, a 'synapomorphy', identifies the two species as progeny from a common ancestor in which the derived character (state) arose and from which they have inherited it. All species derived from a common ancestor form a 'monophyletic' group.

The principle of this approach is straightforward. Implementation is another matter. The identification of derived character states is an important aspect of cladistics (Stevens 1980), but this is usually based on character state distributions rather than on independent conclusive evidence. As a rule, the nearest sister taxon to the group under study is included as an 'outgroup' in the analysis. Character states shared with the outgroup are considered plesiomorphic, and the branching point between 'ingroup' and 'outgroup' indicates the root of the ingroup phylogeny. Selecting the nearest sister taxon, however, usually involves an educated guess about the phylogeny to be analysed, and introduces one of the circular arguments that should be avoided. Iterative trial and error is not completely excluded, even from strict cladistic analysis.

PARSIMONY

Cladistics is explicit about the distinction between shared derived characters and homoplasies (parallel evolution, reversion to the ancestral state, similarity interpreted as identity). Still, it is unavoidable that virtually any data set contains a mixture of phylogenetically informative characters and false interpretations so that various characters suggest various mutually exclusive groupings of the species. To deal with this, there is an additional important concept in cladistic analysis: we assume that a phylogenetic tree that can accommodate all character states in the data with the least number of evolutionary steps contains the fewest *ad hoc* assumptions, and that this most efficient, or most parsimonious tree is the best available approximation to the real tree (Stewart 1993). A tree that can accommodate all characters with one single transition from the original to a permanently derived state per character would be the most satisfying solution to the evolutionary puzzle.

The application of the parsimony principle introduces a quantitative aspect: various mutually exclusive character state distributions are sorted so that the maximum possible agreement can be reached. To this degree the parsimony principle undoes the aim of the cladistic treatment of characters. In practice, a cladistic analysis using parsimony will figure somewhere in a continuum between an unequivocal reconstruction of the phylogenetic branching pattern using independently proven synapomorphies (and then a single one is enough to define a monophyletic group) and a quantitative counting of contradictory apparent synapomorphies that are individually equally probable. With many such characters, the number of synapomorphies supporting a monophyletic group

becomes not only a statistical measure of the reality of that group but (as the 'branch length' of the branch leading to that group) a measure of the evolutionary distance of that group from the last common ancestor with its sister group. This way, cladistics can incorporate features of a quantitative distance method. This can add information but it also introduces problems associated with evolutionary distance measurements. The most important of these, of course, is that for quantitative distance methods and parsimony methods alike the individual characters have to evolve independently. The evolution of selfing in *Eichhornia* (Barrett 1995) illustrates how the repeated coordinated evolution of a functional syndrome of characters can distort the phylogenetic analysis with parsimony.

Finding the most parsimonious tree permitted by the character state distributions for a given data set becomes a mechanical task that can be left to a computer (Swofford & Olsen 1990; Li & Graur 1991; Avise 1994; De Laet & Smets 1994). In fact, only a computer can even begin the enormous amount of calculation required in cladistic analysis. There seems to be no other way than to calculate all the possible relationships that the species in question can theoretically have, to determine for each and every character the minimal number of evolutionary steps that each of these trees requires in order to reach the character distribution among the terminal branches, and then to add them up for all characters per tree. With only 10 species, there are 34 459 452 possible trees connecting them to a common ancestral root. With an increasing number of species, searching for the most parsimonious tree soon becomes a task of astronomical proportions. Even if an exhaustive search can be completed, there may be many equally parsimonious trees and very many trees with only one or two additional steps so that no single solution is significantly more parsimonious than the others.

In order to see how these optimal and near-optimal arrangements of character state distributions are related to the true tree, model 'true trees' have been generated by computer, unordered discrete character state sets have been derived from them and used for cladistic analysis with maximum parsimony (Lamboy 1994). The result showed the following: when the taxa in the analysis have arisen by a series of lineage bifurcations, when they represent the full set of the result of these bifurcations without extinction or missing taxa, when there are only two character states per character, and when there are no parallelisms or reversals in the true tree, then maximum parsimony will only find one single tree and that will be the true tree. Unfortunately, these conditions are practically never met, and frequently none of them is met. With every relaxation from these conditions, chances of finding the true phylogeny decrease.

This does not mean that we should abandon cladistics. After all, we seem to have no better alternative. However, there are two lessons to be learned: we should be very conservative with the interpretation of the results and we should be aware of any potential bias that is introduced by the choice of the characters.

The nature of the calculation of maximum parsimony trees provides us with some internal controls for the reliability of the results (Felsenstein 1985; Farris 1989; Sanderson 1989). For instance, there will be groups of species that appear in all possible maximum parsimony trees and could only be broken up with an appreciable loss of parsimony. We can assume that such groups are very likely monophyletic groups. Conversely, there may be considerable disagreement even among the maximum parsimony trees about the grouping of some taxa. This would show that the available evidence does not allow a reliable phylogenetic reconstruction. Omitting weakly supported branches by collapsing them into unresolved multiple branch points and

accepting only very robust parts of the cladogram is certainly safer than pushing the data for each and every suggestive hint. Phylogenetic trees are used for the analysis of character evolution, biogeographic history or coevolution by plotting characters (Jansen *et al.* 1992; Morgan *et al.* 1994; Barrett 1995) or distribution areas against the tree or by comparing trees of interacting organisms (Herre *et al.*, in press). The conclusions in these cases depend on the assumption that the reconstructed tree represents the true phylogeny.

SYNAPOMORPHIES

A crucial feature of cladism that is frequently not fully appreciated is the very selective use of characters. The definition of a synapomorphy requires that a character has to remain identical for periods exceeding the life time of species or even higher taxonomic groups. This implies a rapid, saltatory evolution from one character state to a new, derived one followed by a long period of evolutionary stasis, similar to (and sometimes confused with) the 'punctuated equilibrium' model of evolution (Eldredge & Gould 1972), but at the level of individual characters rather than species. The success of cladistic analysis shows that such characters exist. However, this is empirical evidence and little is known about the nature or evolution of these characters. Practical experience indicates that there are often uncomfortably few morphological characters that satisfy these stringent conditions. Attempts to increase the data set by relaxing the conditions inevitably add noise and wrong information to the data, and some of the doubtful results of cladistic analysis clearly are due to insufficiently rigorous screening of the characters. A single proven synapomorphy is sufficient to define a monophyletic group of species, but there are hardly any morphological characters that can be recognized with certainty as derived characters of unique ancestry. The major limitation of cladistic analysis with morphological data is the very limited number of reliable characters.

MOLECULAR DATA

It is precisely this point where molecular data become important (Hillis & Moritz 1990; Li & Graur 1991; Soltis *et al.* 1992b; Avise 1994; Bakker *et al.* 1994). The nucleotide sequences of genomes are more than just another set of characters. These sequences encode all heritable information about phenotypic characters, i.e. precisely those aspects of the phenotype that are relevant for phylogenetic reconstruction. Moreover, they encode all of this information in the uniform language of a linear sequence of four different nucleotides, which is an ideal format for routine statistical analysis. In addition, there is sequence information in the genome that does not code for phenotypic characters. Theoretically, at least, a knowledge of the genomic DNA sequences alone comprises all the preserved phylogenetic information available from living organisms.

In practice, of course, we are far from being able to read and use this information. We may live to see one or two complete plant genomes sequenced, and this will provide invaluable information for molecular taxonomy. Typically, we have to choose a minute sample of the available information for comparison, either the complete nucleotide sequence of a limited homologous stretch of DNA or a random sampling of sequence variation at homologous sites across the genome. The advantages and limitations of

molecular information differ from those of morphological data. Here, I want to summarize only the most important features relevant for a comparison among the methods.

Morphological characters are arbitrarily defined, isolated aspects of an integrated phenotype. Their definition requires insight and skill, and the recognition of homologous characters and identical character states is often less certain than we might want it to be. This is much less of a problem for molecular data. DNA separated into complementary single strands of nucleotides will re-anneal into the properly paired double-stranded molecule in solution, and this way homologous sequences with sufficient similarity still to allow base pairing can be detected experimentally with a pre-set statistical reliability (the 'criterion' of temperature and salt concentration).

Practically all presently used methods in molecular taxonomy will reveal polymorphisms (character state differences) as differences in banding patterns on gels when DNA fragments are separated according to length by gel electrophoresis. This common appearance of the raw data may even be confusing because the DNA fragments that produce the banding patterns may be generated in very different ways, and that will profoundly influence their interpretation. Depending on the methods, bands at equivalent positions in the gel may be anything from suggestive to certain indicators of character state identity. In general, though, both character homology and character state identity can be established rigorously by experimental means and scored objectively.

On the other hand, base pairs at homologous positions are not particularly good phylogenetic characters. There are only four of them, and in principle, and over long periods of evolution, each can replace every other one so that independent origins of identical base pairs cannot be excluded and differences in base pairs may not reflect single mutations. The cladistic analysis of DNA sequence variation often approaches the situation which I have characterized above as the quantitative sorting of many equally probable characters, where parsimony analysis approaches distance analysis, and it is not surprising that sequence data often are analysed by both methods or mixed strategies.

SEQUENCE COMPARISONS

Depending on the degree of stabilizing selection on the sequence, there is a window of informative sequence evolution between too few differences to resolve a phylogeny and so many differences that the phylogenetic signal disappears in statistical noise (Hillis 1991). It is fortunate that among the incredible mass of genetic information in any one plant, there is a nearly continuous scale of evolutionary rates from highly conserved sequences that are phylogenetically informative for higher taxa to sequences that vary among the individuals of a population (Wolfe *et al.* 1987). Among base pairs, sudden mutation and fixation followed by very long periods of persistence is a common mode of evolution. For any stretch of DNA the evolutionary rate (fixed base pair changes per site and million years) can be estimated empirically, and in some cases this rate is sufficiently constant to permit the use of overall similarity as a measure for evolutionary relatedness and an extrapolation of evolutionary timing (Olsen *et al.* 1994). However, here as anywhere else in evolution, we are dealing not with universal constants but with historical contingencies. The next case study may invalidate our generalizations, and any extrapolation beyond the proven empirical data is essentially guesswork.

One feature of nucleic acids that can greatly complicate comparative sequence analysis is the occasional loss or gain of one or more base pairs in a sequence. Formally, this will require introducing a gap in the shorter sequence in order to align homologous sites properly. Of course, nucleic acid sequences are continuous and losses or gains of basepairs leave no trace. They have to be inferred from the alignment of the surrounding bases (Swofford & Olsen 1990). A single deletion or insertion (an 'indel' if we do not know which is the derived condition) between two otherwise similar sequences is easy to recognize. When indels and base changes accumulate, placing gaps becomes increasingly arbitrary. The guiding principle, of course, is maximizing sequence similarity with a minimum of *ad hoc* gaps. Since this involves a relatively arbitrary weighing of improved alignment against number of gaps, the comparison of homologous sequences can become questionable or impossible when the independent evolution of the species under comparison has been long relative to the evolutionary rate of the sequence.

SEQUENCE REARRANGEMENTS

Occasionally, though, evolutionary restructuring of sequences may provide phylogenetic information that is so clear and reliable that all the cautionary statistics becomes superfluous. This can be the case with major sequence rearrangements such as the loss, the duplication, the translocation or inversion of a longer stretch of sequence. We are beginning to understand the mechanisms underlying such processes in nucleic acid sequences sufficiently (Hiratsuka *et al.* 1989; survey in Li & Graur 1991) to be able to identify with near certainty those that are unique historical accidents. All organisms sharing such a sequence rearrangement can be assumed to form a monophyletic group, however similar or different they may look. For example, the monophyly of the order Dasycladales in the Chlorophyta is supported by a short deletion in their genes for 18S rRNA that corresponds to a prominent point in the secondary structure of the RNA molecule (Olsen *et al.* 1994). A classical example in which a single sequence rearrangement defines a major monophyletic group in the angiosperms is an inversion of a stretch of 22·000 base pairs (22 kilobases, 22 kb) in the chloroplast DNA of all Compositae except the subtribe Barnadesiinae (Jansen & Palmer 1987). Many other examples for rearrangements in chloroplast DNA are listed in a recent review by Downie & Palmer (1992). Even such unique rearrangements, which are the results of very rare historical accidents, are not equally rare throughout all plant groups. Some groups, for instance conifers, Geraniaceae, Fabaceae, Campanulaceae and Lobeliaceae seem to be more prone to generate rearrangements in chloroplast DNA than others (Downie & Palmer 1992), and there is an easily accessible region in the chloroplast DNA (Taberlet *et al.* 1991) that is a hot spot for 'indels' that are informative at lower taxonomic levels (Böhle *et al.* 1994; Mes & 't Hart 1994; Van Ham *et al.* 1994).

Recently, Downie & Palmer (1994) have found a 6 kb sequence inversion in the chloroplasts of *Atriplex* and *Chaenopodium* (Chaenopodiaceae) that these species seem to share with *Pereskia* (Cactaceae). From all other evidence we can be reasonably sure that the two families are not closely related. The precise end-points of the inversions have not yet been determined. If they really are identical and the precisely same major inversion has occurred twice independently, the argument of the preceding paragraph would be severely weakened. This has implications for more than the taxonomic position of the cacti or the taxonomic significance of inversions. It illustrates that character evolution can be studied rigorously by making falsifiable predictions and by

experimental investigation. Character state changes of any one type, inversions in a DNA sequence or changes in the number of petals, can occur repeatedly and we may eventually even understand the underlying mechanisms sufficiently to induce such character state changes at will. This possibility provides a firm scientific basis for the selection of highly reliable character state matrices for phylogenetic reconstruction, and it integrates systematics with experimental botany. I shall return to this important point at the end of the paper.

CHLOROPLAST DNA, NUCLEAR DNA

It is not an accident that I have taken my examples mainly from chloroplast DNA. Both mitochondria and chloroplasts of plants contain their own small genomes. Of these, the chloroplast genomes have been particularly informative for evolutionary reconstruction (Palmer *et al.* 1988; Clegg & Zurawski 1992; Clegg 1993). The chloroplast genome is a continuous ring of DNA containing about 150·000 base pairs. Several representative chloroplast genomes have been completely sequenced (Shinozaki *et al.* 1986, *Nicotiana*; Ohyama *et al.* 1986, *Marchantia*; Hiratsuka *et al.* 1989, *Oryza*; Wolfe *et al.* 1992, *Epifagus*; Tsudzuki *et al.* 1992, *Pinus thunbergii*; Hallick *et al.* 1993, *Euglena*). This makes it relatively easy to identify precisely homologous sequences among various chloroplast genomes. There is also quite a bit of information on the relative rates and modes of evolution in different regions of the chloroplast DNA (Wolfe *et al.* 1987). Chloroplast genomes usually are inherited from one parent only, in angiosperms typically via the mother, in conifers through the pollen. There is usually only one type of chloroplast genome in a plant, there is no recombination, and the evolution of this DNA is therefore strictly clonal by the accumulation of fixed mutations. This provides ideal data for cladistic analysis at any and all taxonomic levels down to the individual. The overall rate of mutation in chloroplast DNA is relatively slow (Wolfe *et al.* 1987), and chloroplast DNA therefore is best suited for reconstructing phylogenies at the level genera and families.

The phylogenetic information contained in chloroplast genomes has been extensively exploited, and there are various technical approaches for a comparative analysis (Palmer *et al.* 1988; Clegg & Zurawski 1992). The coordinated effort of many laboratories in the United States to sequence the chloroplast encoded gene for the large subunit of ribulose 1,5 bisphosphate carboxylase (*rbc*L) in hundreds of representative higher plant species (Chase *et al.* 1993) is the best demonstration yet for the power of molecular phylogenetics (discussed in Baum 1994; Soltis & Soltis 1995).

The base substitutions or sequence rearrangements in chloroplast DNA that are used as phylogenetic characters have virtually no relation with the information expressed in morphological characters. They are truly additional independent information.

In fact, this is also true for most of the presently available data from nuclear DNA (Bachmann 1992). The sheer mass of DNA in nuclear genomes, the enormous variation in nuclear genome sizes among angiosperms (much of it due to sequences found only in one species or a limited taxonomic group), the fact that typically each nucleus contains two genomes with sequence polymorphisms between them, and the regular exchange of sequence variants by recombination in sexual reproduction are all complications even though it is here, of course, that the essential evolutionary processes take place. I shall deal with the genetic information for morphological characters below. Most of the nuclear DNA characters used for phylogenetic reconstruction are base changes in

selected sequences, especially the genes coding for ribosomal RNAs (Hamby & Zimmer 1992; Baldwin 1992) and are not responsible for morphological character differences.

MOLECULAR VERSUS MORPHOLOGICAL CHARACTERS

The immediate question, of course is the congruence of phylogenetic trees derived from DNA sequences and from morphological data. Above, I have suggested that molecular data frequently confirm the classic taxonomic groupings but that there are innumerable cases where molecular data demand a revision of details, and there are some instances where morphology seems to have been misleading in a major way. The comparison shows that there are some recurrent reasons for incongruence between data sets. Revealing case histories have been compiled by Sytsma (1990), Kadereit (1994) and Soltis & Soltis (1995).

In some cases apparently decisive morphological characters have alternatively supported two contradicting interpretations of the data, and the wrong one was chosen. Formally, these are extreme cases of mistaken character weighting to resolve a deadlock in parsimony. Character weighting of nucleotide sequence data (Albert & Mishler 1992) is based on knowledge of mechanisms and consequences of base pair mutations and can enhance the signal in molecular data sets (Bakker *et al.* 1994). I have mentioned the overriding weight of independently confirmed synapomorphies. However, choosing one phenotypic character over another to resolve an impasse in phylogenetic analysis has occasionally led to wrong conclusions.

Often one or more subgroups within a group are easily recognized by striking morphological characters or even character combinations. This can be the case for species groups within a genus. It is tempting to recognize these as separate genera and to remove them from the remaining species. The classical case for this is the genus *Heterogaura* which differs from its nearest relatives in the genus *Clarkia* mainly by the possession of four stamens and four staminodes as opposed to eight stamens in *Clarkia*, by its unlobed stigma while *Clarkia* has four-lobed stigmas, and especially by its fruit. The fruit of *Heterogaura* is a one- or two-seeded nutlet, that of *Clarkia* a many-seeded dehiscent capsule. In spite of this syndrome of character differences, molecular evidence (Sytsma & Gottlieb 1986) unambiguously aligned the monotypic *Heterogaura heterandra* with *Clarkia dudleyana*, an association for which there seems to be no hint among the morphological characters. Less dramatic examples for this effect abound. They are typical for large genera, but occur also at higher taxonomic levels. The family Cupressaceae, for instance, is not a sister group but part of the family Taxodiaceae without the genus *Sciadopitys* (Brunsfeld *et al.* 1994).

There are probably more urgent tasks than straightening up the taxonomic bookkeeping in all of these cases, especially since the existing taxonomy often is more practical, widely used and very familiar. However, as long as we strive for a taxonomy that reflects the natural hierarchy of phylogenetic relationships, all the 'paraphyletic' rest groups (groups sharing their common ancestor with species excluded from the group) are, in fact, false (De Queiros & Gauthier 1994). How to combine the on-going improvement of phylogenetic analysis and the changes in classification and naming that it necessitates with the urgent demand for a stable and user-friendly taxonomy of plants is a question that becomes ever more urgent for the future of systematics. Nowhere else does the scientific progress of a field that is essential for its long-term validity generate so much short-term irritation.

I think it is fair to say that molecular taxonomy, in the very short time that it has been applied on a larger scale, has shown that it is a very powerful technique that can supply a nearly unlimited stream of data and will allow us to resolve taxonomic relationships to the theoretical limit. Many phylogenetic trees based only on a single molecule have already provided data that resolve classical problem groups. Where there are differences between molecular and morphological data, the morphological data sometimes seem convincing (Ronse Decraene & Smets 1995), more often the molecular data reveal unavoidable misinterpretations of the morphological evidence.

LIMITS TO CLADISTICS

The application of molecular data has also forced us to think about the limits of the cladistic approach and has revealed some effects that could not be detected with morphological characters.

Cladistic analysis implies that the present species have arisen by repeated lineage splitting from common ancestors. Fusion of lineages resulting in the formation of allopolyploids is a very frequent event in plant evolution (Gottschalk 1976; Lewis 1979), and the involvement of diploid hybridization in plant speciation needs to be investigated further (Anderson 1949; Arnold 1992; Rieseberg *et al.* 1995). We have to realize that the inclusion of hybrids and allopolyploids in a cladistic analysis will inevitably result in false results (Funk 1985; McDade 1990, 1992). I have mentioned that cladistic analysis of chloroplast DNA effectively circumvents this problem, since chloroplasts evolve clonally. At the same time it highlights an important feature of molecular taxonomy: using chloroplast sequences will reveal the phylogeny of chloroplasts, not necessarily that of the organisms that harbour these chloroplasts. In the case of allopolyploid angiosperms, this may even be an advantage because it will reveal the maternal contribution to the original hybrid and may allow us to recognize multiple independent origins of an allopolyploid combination (Soltis *et al.* 1992a). In the *rbc*L (chloroplast) phylogeny of the Rosaceae (Morgan *et al.* 1994), the Maloideae with $x=17$ chromosomes form a monophyletic group within the Spiraeoideae ($x=9$). Whatever ancient polyploidization event gave rise to the ancestor of the Maloideae, at least the maternal parent was a spiraeoid species. In general, though, phylogenetic trees based on the DNA sequences of single genes are 'gene trees', and their relationship with 'organism trees' will have to be investigated independently. An obvious approach, the comparison of many gene trees of the same organisms for their congruence, is straightforward but will raise the costs of analysis appreciably. The comparison of chloroplast trees with organismal (nuclear) trees has shown that the problem may be more pervasive than we thought, because there are surprisingly many discrepancies at the diploid level that suggest cytoplasmic introgression without visible traces of nuclear hybridization (Rieseberg & Brunsfeld 1992; Roelofs & Bachmann, in press).

The significance of these observations for systematics needs to be investigated. Many allopolyploid species seem to be terminal species in trees. Polyploids can easily be detected by their chromosome numbers, they can be excluded from the cladistic analysis and added later by indicating the diploid taxa involved in their formation. However, there is no reason why polyploids should always be evolutionary dead ends and appear as terminal taxa. Where the parental contributions of the two nuclear genomes can be recognized, they can be separately included in the phylogenetic analysis, but there is always the possibility that the two parental genomes do not evolve independently

in a polyploid (Van Houten *et al.* 1993) and that a polyploid species is the progenitor of new polyploid lineages as in the Maloideae cited above. In these cases biparental analysis becomes very complex. The entire problem of 'reticulate' evolution versus lineage splitting needs to be examined thoroughly, especially in groups of presumably 'paleopolyploids'.

THE FUTURE OF MORPHOLOGICAL DATA

I have tried to show that molecular data have not only contributed to cladistic analysis but have also revealed limits to cladistic analysis and that, together with cytogenetics, they provide the best clues to detect and examine these limits. This strong endorsement of molecular methods must not be taken as an indication that molecular taxonomy should or even could replace the use of morphological characters. We must not forget that virtually all molecular studies up to now have made full use of previous morphological analyses to select interesting material and to identify groups requiring further examination. The molecular data have been used to test existing hypotheses based on morphology. Once the molecular methods are more routine, the preliminary morphological sorting can possibly be simplified, but it will never be abandoned.

Moreover, reconstructing phylogenetic relationships is only one task of systematics. The taxonomic units also have to be described so that they can be recognized in the field, their abundance, geographical and ecological distribution have to be mapped, and their potential uses have to be determined. When we use morphological and other phenotypic characters, classification and the biological characterization of species go hand in hand, while molecular taxonomy as we do it now reveals nothing besides the phylogenetic relationship of the units that are classified. Their essential biological features either are known already or they have to be determined independently by additional methods. Eventually, the relative roles of molecular methods and phenotypic (including morphological, ecological, biochemical) analysis will depend on the efficiency with which they contribute to providing the necessary answers.

USING MOLECULAR CHARACTERS TO UNDERSTAND MORPHOLOGY

Molecular methods can contribute very much more to phylogenetic research than providing convenient additional characters for an improved resolution of phylogenetic relationships. In the preceding discussion, I have alluded to the fact that the usefulness of a character depends on our understanding of how the character evolves. In the absence of independent evidence, character evolution will have to be inferred from character state distributions, and one of the aims of cladistics is to break out of the circular reasoning that this implies. For the evolution of molecular characters, we can draw on a large store of information on mechanisms and rates of mutation, recombination and sequence rearrangements, on the effects of base pair replacements on the stability of nucleic acid molecules and on the proteins that may be coded by the nucleic acids. All this is important independent evidence that will affect our interpretation of molecular characters (Li & Graur 1991). In contrast, we know very little about the evolutionary mechanisms involved in morphological character changes, especially of the characters that are diagnostic for species or synapomorphies for higher groups. Their very stability suggests that stabilizing selection from the environment alone may be

insufficient to explain their stasis, and that internal, genetic and developmental canalizing mechanisms may play a crucial role. If we knew these mechanisms and the genes that are involved, it would greatly help in defining homologous morphological characters and recognizing identical character states. We could also identify and study the molecular events involved in morphological character changes and have an integrated view of plant evolution rather than a reconstruction from essentially neutral characters that happen to evolve along with the plants.

This integration of molecular and morphological analysis of evolution is now becoming feasible, and the possibility of finding homologous DNA sequences in different species by DNA/DNA hybridization opens the entire range of investigations in plant molecular biology for comparative evolutionary studies. Two approaches deserve the special attention of systematists: the genetic analysis of developmental processes in model systems such as *Arabidopsis thaliana* and the use of Mendelian genetics with molecular markers to provide maps of the nuclear genome in which genes for morphological characters can be localized (Doebley 1993). A few selected examples can illustrate the potential of these approaches for systematics.

EXTRAPOLATING FROM MODEL SYSTEMS

A systematic search for genes involved in flower development in *Arabidopsis thaliana* has revealed several genes with decisive influence on floral meristem identity. Mutants that destroy or reduce the function of these genes show striking and specific phenotypes. In a double mutant of the two genes, *APETALA1* and *CAULIFLOWER*, cells that would normally constitute a floral meristem instead behave as an inflorescence meristem and give rise to additional meristems in a spiral phyllotactic pattern (Bowman *et al.* 1993). The plants produce an extensive orderly proliferation of meristems at each position that in a non-mutant *Arabidopsis* would give rise to a single flower. The phenotype gets its name from its resemblance to cultivated cauliflower, *Brassica oleraceae* var. *botrytis*.

The *CAULIFLOWER* gene of *Arabidopsis* has been isolated and characterized. This then has been used to identify the homologous gene in *Brassica oleracea*, and it turns out that this gene is mutated and non-functional in cauliflower (Kempin *et al.* 1995). The two similar phenotypes in *Arabidopsis* and *Brassica* therefore are strictly homologous characters. However, since the mutations inactivating the genes in either species are different, the identical morphological phenotypes are not synapomorphies but results of parallel events. This case illustrates clearly why shared derived morphological phenotypes that look, and in fact are, identical, still can be homoplasies rather than synapomorphies. It also shows that eventually, where this becomes crucial, we can look directly for the primary genetic mutation and determine character state identity at this ultimate level.

The observation that the inactivation of one or two genes can have such profound and specific phenotypic effects as the conversion of cabbage into cauliflower is significant. I have pointed out above that the introduction of cladistics has implicitly replaced our impression that morphological characters evolve by small quantitative steps with a concept of character evolution by single crucial mutations with large phenotypic effects. The significance of this type of character evolution, especially for taxonomically relevant characters, has been explicitly suggested by several authors in the early 1980s (Hilu 1983; Bachmann 1983; Gottlieb 1984). Since then, molecular methods in genetic analysis have

greatly facilitated the experimental investigation of character evolution and shown that single mutations with large effects do indeed play a role in plant evolution (Doebley 1993).

MARKER ASSISTED GENETIC ANALYSIS

Often, plants from two species differing in diagnostic characters can be crossed and produce at least partially fertile offspring. In these cases, a direct Mendelian analysis of character differences is possible, but the very complex phenotypic segregation in interspecies hybrids combined with effects such as heterosis and phenotypic plasticity, has traditionally complicated the analysis of phenotypic segregations.

An example is the reduction from 4 to 2 microsporangia on the anthers of three species of *Microseris*, which is a synapomorphy for these species (Battjes *et al.* 1994). Plants with 4 and with 2 microsporangia can be crossed, and the first-generation hybrid has mostly 4 microsporangia. Upon selfing, this hybrid produces a segregating population in which many plants have variable numbers of microsporangia with averages from 2 to 4. In fact, most of this continuous quantitative variation is due to the 3:1 segregation of one single gene with a dominant allele determining 4 microsporangia. Multiple modifier genes with minor effects that stabilize the phenotypic expression in the parents segregate randomly in the hybrid and destabilize the phenotype more often than they stabilize it (Battjes *et al.* 1994). A very similar system of one main gene and several modifiers has been found for the constant number of 5 pappus parts in the annual species of *Microseris* (Vlot *et al.* 1992). In homogyous recessives for a major dominant gene keeping the number canalized to 5, pappus part number can vary down to zero, and the effect of multiple modifiers becomes evident (Vlot *et al.* 1992; Bachmann & Hombergen, in press). Such systems are crucial models for the genetic basis of character evolution. However, phenotypic analysis alone cannot go much farther than illustrating the basic structure of the system.

This has changed with the introduction of molecular markers, specifically polymorphisms in the DNA sequence distributed all over the genome that can be scored unambiguously and mapped genetically. The segregation of such polymorphisms (which need not have any known functional significance in themselves) can be compared with that of quantitative traits, and they can be used to detect regions of the genome that contain genes influencing these traits (quantitative trait loci, QTLs; Beckmann & Soller 1986; Lander & Botstein 1989; Paterson *et al.* 1988). The method is widely used in agriculture and forestry (Phillips & Vasil 1994). Doebley *et al.* (1990) have used a cross between maize and its progenitor, teosinte, to detect and map loci involved in the evolution of maize. Recently, simple and relatively inexpensive methods for the detection of markers have been developed, especially 'Random Amplified Polymorphic DNA' (RAPD) markers (Welsh & McClelland 1990; Williams *et al.* 1990). With these, marker-based QTL mapping becomes available for research in evolutionary genetics of 'wild' species. Using this method, the major gene and at least one of the modifier genes determining deviations from 5 pappus parts in *Microseris* have been marked and mapped (Bachmann & Hombergen, in press).

The possibility of using molecular markers not only as very convenient indicators of evolutionary relationship, but also to use them to evaluate homology and evolution of morphological characters guarantees that molecular marker systematics will not become a sterile routine exercise separate from evolutionary biology. It will revitalize the

analysis of morphological characters for phylogenetic reconstruction and link systematics to other areas in the forefront of biology. The common methodology will greatly facilitate communication among biologists and reveal basic biological questions and the contribution of the various fields to their solution.

REFERENCES

Albert, V.A. & Mishler, B.D. (1992): On the rationale and utility of weighting nucleotide sequence data. *Cladistics* **8:** 73–83.

Anderson, E. (1949): *Introgressive Hybridization.* Wiley, New York.

Arnold, M.L. (1992): Natural hybridization as an evolutionary process. *Ann. Rev. Ecol. Syst.* **23:** 237–261.

Avise, J.C. (1994): *Molecular Markers, Natural History and Evolution.* Chapman & Hall, London.

Bachmann, K. (1983): Evolutionary genetics and the genetic control of morphogenesis in flowering plants. *Evol. Biol.* **16:** 157–208.

Bachmann, K. (1992): Nuclear DNA markers in angiosperm taxonomy. *Acta Bot. Neerl.* **41:** 369–384.

Bachmann, K. & Hombergen, E.J. (in press): Mapping genes for phenotypic variation in *Microseris* (Lactuceae) with molecular markers. In: Hind, L. (ed.) *Compositae: Systematics, Biology, Utilization.* Kew Gardens, Richmond, UK.

Bakker, F.T., Olsen, J.L., Stam, W.T. & Van den Hook, C. (1994): The *Cladophora* complex (Chlorophyta): new views based on 18S rRNA gene sequences. *Mol. Phylog. Evol.* **3:** 365–382.

Baldwin, B.G. (1992): Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. *Mol. Phyl. Evol.* **1:** 3–16.

Barrett, S.C.H. (1995): Mating-system evolution in flowering plants: micro- and macroevolutionary approaches. *Acta Bot. Neerl.* **44:** 385–402.

Battjes, J., Chambers, K.L. & Bachmann, K. (1994): Evolution of microsporangium numbers in *Microseris* (Asteraceae: Lactuceae). *Am. J. Bot.* **81:** 641–647.

Baum, D. (1994): *rbc* L and seed plant phylogeny. *Trends in Ecol. Evol.* **9:** 39–41.

Beckmann, J.S. & Soller, M. (1986): Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica* **35:** 111–124.

Böhle, U.-R., Hilger, H., Cerff, R. & Martin, W.F. (1994): Non-coding chloroplast DNA for plant molecular systematics at the infrageneric level. In: Schierwater, B., Streit, B., Wagner, G.P. & DeSalle, R. (eds) *Molecular Ecology and Evolution: Approaches and Applications.* Birkhäuser, Basel.

Bowman, J.L., Alvarez, J., Weigel, D., Meyerowitz, E.M. & Smyth, D.R. (1993): Control of flower development in *Arabidopsis thaliana* by *APETALA1* and interacting genes. *Development* **119:** 721–743.

Brunsfeld, S.J., Soltis, P.S., Soltis, D.E., Gadek, P.A., Quinn, C.J., Strenge, D.D. & Ranker, T.A. (1994): Phylogenetic relationships among the genera of Taxodiaceae and Cupressaceae: evidence from *rbc* L sequences. *Systematic Botany* **19:** 253–262.

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qui, Y.-L., Kron, K.A., Rettig, J.H., Conti, E., Palmer, J.D., Manhart, J.R., Sytsma, K.J., Michaels, H.J., Kress, W.J., Karol, K.G., Clark, W.D., Hedrén, M., Gaut, B.S., Jansen, R.K., Kim, K.-J., Wimpee, C.F., Smith, J.F., Furnier, G.R., Strauss, S.H., Xiang, Q.-Y., Plunkett, G.M., Soltis, P.S., Swensen, S.M., Wiliams, S.E., Gadek, P.A., Quinn, C.J., Eguiarte, L.E., Golenberg, E., Learn, G.H.Jr., Graham, S.W., Barrett, S.C.H., Dayanandan, S. & Albert, V.A. (1993): Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals Miss. Bot. Garden* **80:** 528–580.

Clegg, M. (1993): Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA* **90:** 363–367.

Clegg, M.T. & Zurawski, G. (1992): Chloroplast DNA and the study of plant phylogeny: present status and future prospects. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants.* 1–13. Chapman and Hall, New York.

De Laet, J. & Smets, E. (1994): Inleiding tot het cladisme. *Belg. Journ. Bot.* **127:** 207–229.

De Queiros, K. & Gauthier, J. (1994): Toward a phylogenetic system of biological nomenclature. *Trends Ecol. Evol.* **9:** 27–30.

Doebley, J. (1993): Genetics, development and plant evolution. *Curr. Op. Genet. Dev.* **3:** 865–872.

Doebley, J., Stec, A., Wendel, J. & Edwards, M. (1990): Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proc. Natl. Acad. Sci. USA* **87:** 9888–9892.

Donoghue, M.J. (1992): Homology. In: Keller, E.F. & Lloyd, E.A. (eds) *Keywords in Evolutionary Biology.* 170–179. Harvard University Press, Cambridge.

Downie, S.R. & Palmer, J. D. (1992): Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants*. 14–35. Chapman and Hall, New York.

Downie, S.R. & Palmer, J.D. (1994): A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeat restriction site variation. *Systematic Botany* **19:** 236–252.

Eldredge, N. & Gould, S.J. (1972): Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T. J. (ed.) *Models in Paleobiology*. 82–115. Freeman, San Francisco.

Farris, J.S. (1989): The retention index and the rescaled consistency index. *Cladistics* **5:** 417–419.

Felsenstein, J. (1985): Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783–791.

Forey, P.L., Humphries, C.J., Kitching, I.J., Scotland, R.W., Siebert, D.L. & Williams, D.M. (1992): *Cladistics. A Practical Course in Systematics*. Clarendon Press, Oxford.

Funk, V.A. (1985): Phylogenetic patterns and hybridization. *Ann. Mo. Bot. Gard.* **72:** 681–715.

Gottlieb, L.D. (1984): Genetics and morphological evolution in plants. *Am. Nat.* **123:** 681–709.

Gottschalk, W. (1976): *Die Bedeutung der Polyploidie für die Evolution der Pflanzen*. Gustav Fischer, Stuttgart.

Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielman, A. & Stutz, E. (1993): Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucl. Acids Res.* **21:** 3537–3544.

Hamby, R.K. & Zimmer, E.A. (1992): Ribosomal RNA as phylogenetic tool in plant systematics. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants*. 50–91. Chapman and Hall, New York.

Hennig, W. (1966): *Phylogenetic Systematics*. University of Illinois Press, Urbana, Illinois.

Herre, E.A., Machado, C.A., Bermingham, E., Nason, J.D., Windsor, D.M., MacCafferty, S.S., Van Houten, W. & Bachmann, K. (in press): Molecular phylogenies of figs and their pollinating wasps. *J. Biogeography*

Hillis, D.M. (1991): Discriminating between phylogenetic signal and random noise in DNA sequences. In: Miyamato, M.M. & Cracraft, J. (eds) *Phylogenetic Analysis of DNA Sequences*. 278–294. Oxford University Press, Oxford.

Hillis, D.M. & Moritz, C. (1990): *Molecular Systematics*. Sinauer, Sunderland, MA.

Hiratsuka, J., Shimada, H., Whittier, R., Ishibshi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989): The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* **217:** 185–194.

Hilu, K.W. (1983): The role of single-gene mutations in the evolution of flowering plants. *Evol. Biol.* **26:** 97–128.

Jansen, R.K., Michaels, H.J., Wallace, R.S., Kim, K.-J., Keeley, S.C., Watson, L.E. & Palmer, J.D. (1992): Chloroplast DNA variation in the Asteraceae: phylogenetic and evolutionary implications. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants*. 252–279. Chapman and Hall, New York.

Jansen, R.K. & Palmer, J.D. (1987): A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). *Proc. Natl. Acad. Sci. USA* **84:** 5818–5822.

Kadereit, J.W. (1994): Molecules and morphology, phylogenetics and genetics. *Bot. Acta* **107:** 369–373.

Kempin, S.A., Savidge, B. & Yanofsky, M.F. (1995): Molecular basis of the cauliflower phenotype in *Arabidopsis*. *Science* **267:** 522–525.

Lander, E.S. & Botstein, D. (1989): Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121:** 185–199.

Lamboy, W.F. (1994): The accuracy of the maximum parsimony method for phylogeny reconstruction with morphological characters. *Systematic Botany* **19:** 489–505.

Lewis, W.H. (ed.) (1979): *Polyploidy: Biological Relevance*. Plenum Press, New York.

Li, W.-H. & Graur, D. (1991): *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, MA.

McDade, L. (1990): Hybrids and phylogenetic systematics. I. Patterns of character expression in hybrids and their implications for cladistic analysis. *Evolution* **44:** 1685–1700.

McDade, L. (1992): Hybrids and phylogenetic systematics. II. The impact of hybrids on cladistic analysis. *Evolution* **464:** 1329–1346.

Mes, T.H.M. & 't Hart, H. (1994): *Sedum surculosum* and *S. jaccardianum* (Crassulaceae) share a unique 70 bp deletion in the chloroplast DNA trnL (UAA)-trnF(GAA) intergenic spacer. *Plant Syst. Evol.* **194:** 213–221.

Morgan, D.R., Soltis, D.E. & Robertson, K.R. (1994): Systematic and evolutionary implications of *rbc* L sequence variation in the Rosaceae. *Amer. J. Bot.* **81:** 890–903.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-I., Inokuchi, H. & Ozeki, H. (1986): Complete nucleotide sequence

of liverwort *Marchantia polymorpha* chloroplast DNA. *Plant Molecular Biology Reporter* **4:** 149–175.

Olsen, J.L., Stam, W.T., Berger, S. & Menzel, D. (1994): 18S rDNA and evolution in the Dasycladales (Chlorophyta): modern living fossils. *J. Phycol.* **30:** 729–744.

Palmer, J.D., Jansen, R.K., Michaels, H.J., Chase, M.W. & Manhart, J.R. (1988): Chloroplast DNA variation and plant phylogeny. *Ann. Missouri Bot. Garden* **75:** 1180–1206.

Paterson, A.H., Lander, E.S., Hewitt, J.D., Peterson, S., Lincoln, S.E. & Tanksley, S.D. (1988): Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction length polymorphisms. *Nature* **335:** 721–726.

Patterson, C. (1988): Homology in classical and molecular biology. *Mol. Biol. Evol.* **5:** 603–625.

Phillips, R.L. & Vasil, I.K. (eds) (1994): *DNA-based Markers in Plants.* Kluwer, Dordrecht.

Remane, A. (1952): *Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der Phylogenetik.* Akademische Verlagsgesellschaft, Leipzig.

Rieseberg, L.H. & Brunsfeld, S.J. (1992): Molecular evidence and plant introgression. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants.* 151–176. Chapman and Hall, New York.

Rieseberg, L.H., Van Fossen, C. & Desrochers, A.M. (1995): Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* **375:** 313–316.

Roelofs, T.F.M. & Bachmann, K. (in press): Chloroplast and nuclear variation among homozygous plants in a population of the autogamous annual *Microseris douglasii* (Asteraceae, Lactuceae). *Plant Syst. Evol.*

Ronse Decraene, L.P. & Smets, E.F. (1995): The floral development of *Neurada procumbens* L. (Neuradaceae). *Acta Bot. Neerl.* **44:** 439–451.

Sanderson, M.J. (1989): Confidence limits on phylogenies: The bootstrap revisited. *Cladistics* **5:** 113–129.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayshida, N., Matsubayasha, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kata, A., Tohdoh, N., Shimada, H. & Sugiura, M. (1986): The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Molecular Biology Reporter* **4:** 111–147.

Soltis, P.S., Doyle, J.J. & Soltis, D.E. (1992a): Molecular data and polyploid evolution in plants. In: Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) *Molecular Systematics of Plants.* 177–201. Chapman and Hall, New York.

Soltis, P.S., Soltis, D.E. & Doyle, J.J. (eds) (1992b): *Molecular Systematics of Plants.* Chapman and Hall, New York.

Soltis, P.S. & Soltis, D.E. (1995): Plant molecular systematics. *Evol. Biol.* **28:** 139–94.

Stevens, P.F. (1980): Evolutionary polarity of character states. *Ann. Rev. Ecol. Syst.* **11:** 333–358.

Stewart, C.-B. (1993): The powers and pitfalls of parsimony. *Nature* **361:** 603–607.

Swofford, D.L. & Olsen, G.J. (1990): Phylogeny reconstruction. In: Hillis, D.M. & Moritz, C. (eds) *Molecular Systematics.* 411–501. Sinauer, Sunderland, MA.

Sytsma, K.J. (1990): DNA and morphology: inference of plant phylogeny. *Trends Ecol. Evol.* **5:** 104–110.

Sytsma, K.J. & Gottlieb, L.D. (1986): Chloroplast DNA evidence for the origin of the genus *Heterogaura* from a species of *Clarkia* (Onagraceae). *Proc. Natl. Acad. Sci. USA* **83:** 5554–5557.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991): Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* **17:** 1105–1109.

Tsudzuki, J., Nakashima, K., Tsudzuki, T., Hiratsuka, J., Shibata, M., Wakasugi, T. & Sugiura, M. (1992): Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequence of trn Q, trn K, psb A, trn I, and trn H and the absence of rps 16. *Mol. Gen. Genet.* **232:** 206–214.

Van Ham, R.C.H.J., 't Hart, H., Mes, T.H.M. & Sandbrink, J. (1994): Molecular evolution of noncoding regions of the chloroplast genome in the Crassulaceae and related species. *Curr. Genet.* **25:** 558–566.

Van Houten, W.H.J., Scarlett, N. & Bachmann, K. (1993): Nuclear DNA markers of the Australian tetraploid *Microseris scapigera* and its North American diploid relatives. *Theor. Appl. Genet.* **87:** 498–505.

Vlot, E.C., Van Houten, W.H.J., Mauthe, S. & Bachmann, K. (1992): Genetic and non-genetic factors influencing deviations from five pappus parts in a hybrid between *Microserus douglasii* and *M. bigelovii* (Asteraceae, Lactuceae). *Int. J. Plant Sci.* **153:** 89–97.

Wallace, R.S. & Jansen, R.K. (1990): Systematic implications of chloroplast DNA variation in the genus *Microseris* (Asteraceae: Lactuceae). *Systematic Botany* **15:** 606–616.

Welsh, J. & McClelland, M. (1990): Fingerprinting genomes using PCR with arbitrary primers. *Nucl. Acids Res.* **18:** 7213–7218.

Wiley, E.O. (1981): *Phylogenetics: The Theory and Practice of Phylogenetic Systematics.* Wiley Interscience, New York.

Wiley, E.O., Siegel-Causey, D.J., Brooks, D.R. & Funk, V.A. (1991): *The Compleat Cladist: A Primer of Phylogenetic Procedures.* Museum of Natural History, Lawrence, Kansas.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, J.W. (1990): DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* **18:** 6531–6535.

Wolfe, K.H., Li, W.-H. & Sharp, P. (1987): Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84:** 9054–9058.

Wolfe, K.H., Morden, C.W., Ems, S.C. & Palmer, J.D. (1992): Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. *J. Mol. Evol.* **35:** 304–317.