Lagascalia 19(1-2): 331-344 (1997)

DNA AND PHYLOGENY IN PLANTS: HISTORY AND NEW PERSPECTIVES

PAOLO CAPUTO

Abstract

This paper briefly reviews the experimental techniques and the methods of phylogenetic inference used in plant molecular systematics. Special attention has been given to chloroplast DNA and nuclear ribosomal DNA, as well as to the results obtained by mapping or sequencing these molecules in seed plants. Limitations of the techniques are discussed, especially for what pertains to comparison with other sources of data. A possible future for molecular systematics is outlined.

Introduction

Understanding biological variation is one of the major intellectual endeavors that mankind has been undertaking for the last two thousand years. However, in the long history of this still unfinished quest, we can detect only few important breakthroughs, the majority of which only occurred in the last few centuries. I am obviously referring to Linnean hierarchies, evolutionary theory, Mendelian genetics, all of which were necessary prerequisites to the developments occurred in the second half of this century.

The past few decades have proven to be unusually momentous in modifying systematic thought, as the theoretical foundations of modern population biology, neodarwinism, phenetics, cladistics were laid at the same moment in which new techniques were being developed in microscopy, cytology, genetics and molecular biology. In particular, it has been an exceptionally felicitous coincidence that progress in molecular biology allowed systematists to make use of macromolecules at the same moment in which computer-aided methods were available to analyze large data sets, so opening the way for a novel and virtually inexhaustible source of evidence.

The origin of molecular systematics may be dated back to the publication of ZUCKERANDL & PAULING's seminal paper (1965), in which they suggested a potential phylogenetic usage of macromolecular sequence data, by indicating that mutation rate may be proportional to clock time, and that phylogeny may be deduced from mutations. Much ground was covered since then.

Possible substrates

The choice of the substrate DNA is perhaps the most crucial point in any molecular study; it heavily depends upon the evolutionary span of the problem at hands, and also upon a careful evaluation of the most cost-effective approach. Selection of the appropriate molecule is carried out by evaluating mutation rates and way of inheritance.

Ideally, the molecule or the gene(s) of choice should show enough variation as to allow systematic analysis, but not so much divergence as to hinder the understanding of phylogenetic relationships among the taxa in study. The chosen substrate should also be easy to obtain in large amounts.

Three different kinds of DNA molecules are normally found in a photosynthetic plant cell: nuclear DNA (nDNA), chloroplast DNA (cpDNA) and mitochondrial DNA (mtDNA). These molecules have been exploited to very different degrees by plant systematists, according to their suitability to phylogenetic research and to the easiness of their manipulation.

Chloroplast DNA. CpDNA is the most widely used nucleic acid in plant phylogeny, as it offers several advantages to the systematist. For example, its abundance in plant cells makes it easy to extract in quantitative amounts, so facilitating further analysis (PALMER, 1986). Extensive information on its sequence (OHYAMA & al., 1986; SHINOZAKI & al., 1986; HIRATSUKA & al., 1989) and structure (PALMER, 1985; LAVIN & al., 1990; DOWNIE & PALMER, 1992a and references therein) is available, so that investigation of changes in gene content, structural organization and mutation rates is easily accomplished. Finally, it shows a low rate of nucleotide substitution, at least in land plants (WOLFE & al., 1987; CLEGG & al., 1990), as well as infrequent modifications of structure and gene order (PALMER, 1985; PALMER & STEIN, 1986). This represents a great advantage both on a technical and a systematic standpoints: in fact, cpDNA conservativity allows the usage of the same molecular probes virtually across all embryophytes; moreover, and more importantly, its slow rate of nucleotide substitution makes cpDNA a most suitable tool for studies on distantly related taxa, in such a way relieving morphological analyses, often beset by parallelisms at that level. CpDNA is a circular, covalently closed molecule, typically ranging from 135 to 160 kilobases (kb) in angiosperms. It consists of two identical segments (approx. 25 kb) forming an inverted repeat (IR) which separates the rest of the molecule in two single copy regions (OLMSTEAD & PALMER, 1994 and references therein). Some exceptional genomes have been documented which greatly depart from this structure. Lack of a copy of the IR has been described for some legume taxa (PALMER & THOMPSON, 1982; LAVIN & al., 1990) and for conifers (RAUBESON & JANSEN, 1992); a triple-sized IR was discovered in a genus of *Geraniaceae* (PALMER, NUGENT & HERBON, 1987); loss of various genes and vast deletions have been demonstrated for members of *Orobanchaceae*, a non-photosynthetic and parasitic angiosperm family (DE PAMPHILIS & PALMER, 1990). Apart from these cases, rare major sequence rearrangements (usually in the form of large inversions) are present; they usually are very useful on a phylogenetic standpoint, as they precisely mark monophyletic groups. The split between *Barnadesioideae* and the rest of *Asteraceae*, for example, is revealed by a 22 kb inversion (BREMER & JANSEN, 1992 and references therein).

The most commonly employed chloroplast gene is *rbc*L (ribulose 1,5 diphosphate carboxylase, large subunit - the chief enzyme in carbon fixation), for which a large database is available (CHASE & al., 1993). Generous estimates (CLEGG, 1993) suggest that *rbc*L is phylogenetically informative in an interval of 400-100 million years before present, albeit more conservative opinions exist (ALBERT & al., 1994).

A promising gene is *rps*4 (encoding a chloroplast ribosomal protein), which has been used up to now to infer phylogeny in *Poaceae* (NADOT & al., 1994).

Other cpDNA genes which very recently entered the field of molecular systematics are: *mat*K (encoding a maturase), which has been used on *Saxifragaceae* (JOHNSON & SOLTIS, 1995); *ndh*F (which possibly encodes a dehydrogenase), used on *Scrophulariaceae* (OLMSTEAD & REEVES, 1995) and *Solanaceae* (OLMSTEAD & SWEERE, 1995); *atp*B, used on *Lardizabalaceae* (HOOT & al., 1995).

Mitochondrial DNA. MtDNA, which so much has been employed by zoologists, is virtually useless for systematic purposes in plants: its large size (ranging between 300 and 2000 Kb), its widespread intra- and intermolecular recombination, the speed of its rearrangements, and the slowness of the rate of point mutations (PALMER, 1992) make this molecule unsuitable for the restriction mapping approach which so much helped the understanding of molecular phylogeny in plants.

Nuclear DNA. Depending on the regions chosen, nDNA could be exceedingly useful for a multitude of systematic problems pertaining to very different hierarchical levels, but we are still far from exploiting it in its full power. Several of the difficulties met when dealing with nuclear DNA (e.g., obtaining the gene of interest for the first time), which, until few years ago, required the tedious and time-consuming procedure of genomic cloning have already been solved by the widespread use of PCR thermocyclers, oligonucleotide synthesizers and automated sequencing apparatuses.

Among the few regions of nuclear DNA which, up to now, have been closely scrutinized by plant systematist is ribosomal DNA (n-rDNA), i.e., the DNA which codes for ribosomal RNA (e.g., HAMBY & ZIMMER, 1992). Among the reasons of the widespread use of rDNA we may indicate both technical and fundamental issues: first, it is an abundant fraction of total DNA, as it is represented by a very high, yet highly variable number of copies per genome (see for example ROGERS & BENDICH, 1987); some of its regions are so well conserved that the same probes may be used across broad evolutionary spans; some others are so variable that can be used for microevolutionary studies. In facts, its structure consists of three highly conserved coding units, 18S, 5.8S and 26S, which have been used to provide data at the highest level of plant phylogeny (MISHLER & al., 1994 and references therein), separated by two quite variable internal transcribed spacers, ITS 1 and 2, normally used in infrageneric comparisons (SUH & al., 1993; KIM & JANSEN, 1994a and references therein; BALDWIN & al., 1995); each of these transcription units is separated by a highly variable intergenic spacer (IGS). Besides these, other ribosomal sequences have been used to infer phylogenetic hypotheses, and namely those of the 5S RNA, which are located in other chromosomal sites and code for a RNA molecule of different size. 5S sequences, because of their shortness and high substitution rates (STEELE & al., 1991), can be used only for lower taxonomic hierarchies.

Among the potentially useful nuclear genes which will probably be widely used in the next few years are heat shock genes (WATERS, 1995) and the phytocrome gene family (MATHEWS & al., 1995).

Data collection

The techniques used to analyze the DNA molecules listed above are manifold. For the sake of concision, only the two most popular will be described here: restriction analysis and sequencing.

Restriction endonucleases are bacterial enzymes which cleave DNA molecules only when they find a specific sequence, thus producing a set of discrete DNA fragments, the size of which varies according to the positions of restriction sites. By separating these fragments by gel electrophoresis, transferring them onto membranes and hybridizing them against specific DNA probes, it is possible to detect the presence or absence of a given DNA fragment in a given taxon and, with more effort, to determine the physical succession of restriction sites in a given DNA molecule. The presence or absence of these restriction sites can then be compared across the taxa in analysis.

By using this method, which is called restriction mapping analysis, not only the presence of sites, but also major physical rearrangements in DNA molecules of fairly large size can be easily detected.

The other technique in use, DNA sequencing, involves the detection of the order by which different nucleotides follow each other in a DNA helix. A burdensome technique until few years ago, it has been made much easier by the use of PCR and automated apparatuses. PCR technique simplifies isolation of the target sequence, which formerly required lengthy and expensive cloning, and automated apparatuses facilitate the actual collection of data.

Restriction analysis, being technically simpler, was historically the first technique employed for wide phylogenetic studies in plants. However, the generic expression "restriction analysis" encompasses very different methods (reviewed by BREMER, 1991). Mainly, either restriction fragment length polymorphisms (RFLP) or restriction sites may be used. Theoretical reasons would suggest the use of site data (i.e., restriction mapping) for phylogenetic inference (PALMER, 1987). However, mapping is not always straightforward, and, in the worst cases, may be so difficult to be unpractical. A partial defence of fragment data has been attempted by BREMER (1991) and MORETTI & al. (1993), and fragment data have helped to solve several low-level phylogenetic problems in different plant groups (e.g., STRAUSS & DOERKSEN, 1990; CAPUTO & al., 1991).

A major critique, however, has been raised also to restriction site data, when used to infer phylogenies at high hierarchical levels: gains and losses of sites, in fact, should be used only when gains and losses are proved to be originating from the same point mutation, and this is very difficult, if not impossible, to prove in most cases.

The only constructive way of using restriction data across broad evolutionary spans is mapping large DNA rearrangements (typically, this is done with cpDNA). By this method, for example, the phylogeny of the Asteridae sensu *Cronquist* (1981) was reassessed (DOWNIE & PALMER, 1992b).

Sequence data, which up to few years ago have been more difficult to obtain, are on the contrary of very simple and unequivocal interpretation. They provide more detailed information (but forcefully on smaller regions) than restriction data. Moreover, PCR-mediated techniques require very little amounts of leaf tissue and may use even badly degraded DNA, extracted from herbarium specimens or even from fossils

(GOLENBERG & al., 1990). This makes sequencing the method of choice over restriction analysis when starting material is a limiting factor. Finally, any additional taxon may be included in a pre-existing data matrix in a much easier way than with restriction data.

However, PCR technique is prone to several problems, among which misincorporation and, especially, danger of contamination with different DNAs. This may lead to inexplicable results originating from the casual amplification of the wrong molecule. OLMSTEAD & PALMER (1994), in a review on chloroplast DNA systematic methods, report some very interesting (and worrisome) stories on the matter.

Impact of DNA phylogenies on systematic thought

The exploitation of the above-mentioned techniques produced a host of data, which is readily available to the scientific community both as restriction maps or cladograms in papers and as raw sequences in computerized databases.

The recent history of plant molecular systematics can be easily followed by looking at the tables of contents of a few botanical journals, among which American Journal of Botany, Plant Systematics and Evolution, Annals of Missouri Botanical Garden, Systematic Botany.

At the beginning, on the contrary, molecular systematic papers were more frequently published in molecular journals, and, of course, attention was drawn mainly by selected cultivated plants. Among others, *Lycopersicon* (PALMER & ZAMIR, 1982), *Cucumis* (PALMER, 1982; PERL-TREVES & GALUN, 1985), *Brassica* (PALMER & al., 1983), various legumes (PALMER & THOMPSON, 1982; PALMER & al., 1985; PALMER & al., 1987), *Triticum* (BOWMAN & al., 1981, 1983), *Hordeum* (POULSEN, 1983), *Zea* (TIMOTHY & al., 1979). This possibly not so much for their applicative implications, but because they were very well studied systems, from which DNA was easily extracted. Later on, problems of general taxonomic interest started being faced by molecular techniques (and molecular papers started shifting from molecular journals to botanical ones); nowadays molecular systematics is part of many thorough botanical studies. Among seed plants, and regardless to method of molecular analysis, studies are available for the following groups (only the most recent and relevant studies above tribal level are included): *Acanthaceae* (HEDRÉN & al., 1995), *Asteraceae* (JANSEN & al., 1992; KIM & al., 1992; MICHAELS & al., 1993), *Asteridae* (DOWNIE & PALMER, 1992b; OLMSTEAD & al., 1993), *Arecaceae* (UHL & al., 1995), *Berberidaceae* (KIM & JANSEN, 1994b), the major clades of Bromeliiflorae-Commeliniflorae-Zingiberiflorae (CLARK & al., 1993), Caryophyllales (RETTIG & al., 1992), Caryophyllidae (GIANNASI & al., 1992), *Cornaceae* p.p. (XIANG & al., 1993), *Cupressaceae* and *Taxodiaceae* (GADEK & QUINN, 1993; BRUNSFELD & al., 1994), *Dipsacales* (DONOGHUE & al., 1992), *Droseraceae* (WILLIAMS & al., 1994), *Ericales* (KRON & CHASE, 1993), *Fabaceae* (LAVIN & al., 1990), *Geraniaceae-Geraniales* (PRICE & PALMER, 1993), *Hydrangeaceae* (SOLTIS & al., 1995), *Juncaceae-Cyperaceae* (PLUNKETT & al., 1995), *Lardizabalaceae* (HOOT & al., 1995), *Liliaceae* p.p. (SHINWARI & al., 1994), Magnoliidae (QIU & al., 1993), *Loasaceae-Hydrostachyaceae* (HEMPEL & al., 1995),

Nepetoideae-Lamiaceae (WAGSTAFF & al., 1995), *Onagraceae* (CONTI & al., 1993), *Orchidaceae*-Orchidales (DRESSLER & CHASE, 1995), *Poaceae* (DOEBLEY & al., 1990; NADOT & al., 1994), *Ranunculaceae* (JOHANNSON & JANSEN, 1993), *Rosaceae* (MORGAN & al., 1994), *Rubieae-Rubiaceae* (MANEN & al., 1994), *Scrophulariaceae* (OLMSTEAD & REEVES, 1995), *Saxifragaceae* (SOLTIS & al., 1993; JOHNSON & SOLTIS, 1995), *Solanaceae* (OLMSTEAD & PALMER, 1992; OLMSTEAD & SWEERE, 1995), *Winteraceae* (SUH & al., 1993), *Zamiaceae* p.p., (CAPUTO & al., 1991), *Zingiberales* (SMITH & al., 1993). Furthermore, various wider studies exist on monocotyledons (e.g, DUVALL & al., 1993), angiosperms (e.g, NICKRENT & SOLTIS, 1995 and references therein), "gymnosperms" (e.g, HASEBE & al., 1992), seed plants as a whole (e.g., CHASE & al., 1993; DOYLE & al., 1994; MISHLER & al., 1994).

DNA characters presently provide an indispensable complement to traditional systematic analyses; in many cases they help to choose among different hypotheses of phylogenetic relationships not discriminable by traditional methods, or provide new notions on the location of misplaced taxa.

DNA opened new insights in the phylogeny of several major groups of flowering plants. One of the most studied groups, at all levels, is Asteridae sensu CRONQUIST (1981). The phylogeny of this group and its intersection with Rosidae can be counted among the most fascinating intricacies in angiosperm phylogeny. Classical techniques alone have been of little avail in this case, as it appears from the fact that CRONQUIST's (1981) and TAKHTAJAN's (1987) systems, which are usually so similar, greatly differ in the circumscription of some orders and in the hypotheses of mutual affinities. What is more, some small families (e.g., *Menyanthaceae, Oleaceae, Callitrichaceae, Loasaceae*) were of uncertain placement within Asteridae. Two independent molecular analyses respectively based on restriction mapping of the cpDNA inverted repeat and on rbcL sequencing (DOWNIE & PALMER, 1992b; OLMSTEAD & al., 1992) showed that Asteridae sensu Cronquist may be divided at least in two major lineages, and that, in order to be monophyletic, they should include also Cornales, Apiales and Ericales, as well as *Hydrangeaceae, Loasaceae* and *Fouquieraceae*, all included by Cronquist into Dilleniidae and Rosidae.

Even when molecular studies do not completely allow the choice of one hypothesis against all others, they may still be of great help in excluding some of them. One of the most debated issues in flowering plant phylogeny is the identification of the most archaic living groups. At present, it is still unclear whether magnolioid taxa (CRONQUIST, 1981), calycanthoid taxa (LOCONTE & STEVENSON, 1991), monocotyledonous or chloranthoid taxa (BURGER, 1977, 1981), "paleoherbs" (TAYLOR & HICKEY, 1992) or *Ceratophyllum* (CHASE & al., 1993) are at the base. In spite of the fact that no conclusive hypothesis has been published up to now, and that molecular trees are overall at least as different from one another as they are from morphological trees, molecular hypotheses constantly rule out the presence of monocotyledonous and chloranthoid taxa at the base, so reducing the number of competing hypotheses (DONOGHUE, 1994).

The pessimism of reason

The constant improvement and the power of molecular techniques have convinced many systematists that all evolutionary problems can be solved with molecular data; regarding molecular methods as the panacea, however, would be a gross misconception.

Among the various problems met when using molecular data, the choice of the method of phylogenetic inference is one of the most vexing. In fact, reliable phylogenetic hypotheses can be derived from raw data if, and only if, the data fit (or do not depart too much from) the assumptions of the model used to analyze them.

The crucial issue is that we still do not entirely know which are the selective constraints acting upon DNA sequences and how they relate to those acting on the corresponding proteins (see ALBERT & al., 1994 for an example in which *rbc*L sequences yield a topology different from that obtained with the aminoacid translation of the same data). We also ignore whether neutralistic models (KIMURA, 1983 and references therein) can always be assumed in case of silent nucleotide substitutions.

Methods of phylogenetic inference. The standard methods of analysis until few years ago, have been Wagner's (FARRIS, 1970) and FITCH's (1971) parsimony, both of which find cladograms minimizing the number of transitions between character states required to justify the results, with no constraints on the direction of change. However, the straight use of cladistics implies that any change (e.g., gain vs. loss of restriction sites or transitions vs. transversions) is equally likely as another. This is not true, both for point mutations and, especially, for restriction site mutations (for which, theoretically, the gain/loss ratio is 1/18); for the latter, the usage of Dollo parsimony (which does not allow parallel gains of characters) was deemed more suitable (LE QUESNE, 1974; FARRIS, 1977). Recently, however, ALBERT & al. (1992) demonstrated that Dollo parsimony is inappropriate for restriction site data, by indicating that the proper weight for character states in restriction analyses should be between 1 and 2. They conclude that a parsimony model allowing, but biasing against, less probable events should instead be used.

Still, transitions are overall more likely than transversions (LAKE, 1987), and any synonymous change in a codon more likely than a non synonymous one (KIMURA, 1983). For this reason, generalized parsimony methods (SWOFFORD & OLSEN, 1990 and references therein), which assign a cost for any kind of character state transformations, have been devised, but they are of still cumbersome implementation. What is more, generalized parsimony moves the problem just a step forward: in fact, even if some quite technical procedures have been devised (e.g., ALBERT & al., 1992; ALBERT & al., 1993), it is often difficult to objectively define the costs, the choice of which seems somehow arbitrary (or, however, implying a model of uncertain application).

The choice of different methods of phylogenetic inference is questionable to the same or higher extent. In fact, techniques which may seem virtually assumption-free (e.g., cluster analysis), actually require at least that data are ultrametric (i.e., the distance between any pair of taxa equals the sum of the length of the branches joining them and all taxa should be equidistant from the root); unfortunately, ultrametricity is an assumption which can hardly be demonstrated for the majority of molecular data (SWOFFORD & OLSEN, 1990).

Integration with other sources of data. Another, even more severe problem with molecular analyses is the difficulty of reconciling results with data produced by using morphology. This is true on computational grounds, but even truer because there is no acceptable way of determining which relative weight to attribute to the different subsets of characters: one cannot be sure, in fact, that sequence mutations, or even major DNA rearrangements in a very short region of the genome, are comparable with the gain of a complex morphological trait.

Moreover, in some cases, the topology of a gene tree can be positively misleading when used to infer organismal phylogeny. This may happen especially when using chloroplast DNA (which is usually uniparentally inherited) in the study of hybridogenous/introgressive groups.

Various techniques have been devised to compare trees derived from molecular and morphological data. A rather extreme and provocative view is that maintained by DOYLE (1992), by which molecular data should be inserted in non-molecular studies as a single, multistate, ordered character. The other extreme consists in simply combining data matrices, regardless to the fact that the usually greater length of molecular matrices may cause (in case of comparable homoplasy) morphological data to virtually disappear. Apparently, however, in spite of its counter-intuitiveness (numbers of characters, character weighting, nature of the sampling seem just to be too different) combining matrices has some theoretical merit (DONOGHUE & SANDERSON, 1992).

More moderate views (PAGE, 1993; CAPUTO & al., 1992) consist in combining in different ways the cladograms obtained by separate analyses.

In spite of all difficulties, there is a deeply felt need of combining all possible sources of evidence, when inferring phylogenies. Arguments in favor of this may be found in DONOGHUE & SANDERSON (1992) and in DONOGHUE (1994). Recent studies using this combined approach (ALBERT & al., 1994; DOYLE & al., 1994; MISHLER & al., 1994) have in fact shown that the combination of the different data sets may produce more resolved results, and may cause the appearance of clades not supported by either single analysis.

A glimpse of the future

To foresee the future development of molecular systematics is an arduous task, given the continuous, and often abrupt, technical progress in the field. Certainly, widening the choice of sequence substrates to include many more genes (especially nuclear ones) is a deeply felt need. Large databases exist in fact only for *rbc*L and ribosomal sequences. The study of different sequences may allow to understand better the mechanisms underlying selective value (or potential neutrality) of DNA mutations. On the other hand, widening the existing databases is a stringent necessity: many of the discrepancies between molecular and morphological hypotheses may find their reasons in paucity of sampling for molecular analyses; the latter in fact may cause the presence of long diverging branches attracting one another by chance (FELSENSTEIN, 1978; ALBERT & al., 1994).

In terms of data analysis, new softwares, which can handle very large data matrices are needed, as well as algorithms whose assumptions fit better the tempo and mode of DNA evolution. This in turn will hopefully produce more consistent and objective methods of merging different sources of data, in such way to allow a better integration in phylogenetic studies.

A parallel effort should be made by classical systematists to increase the number of phylogenetic analyses available, as well as to undertake developmental studies, which may still provide many valuable characters (DONOGHUE, 1994).

Finally, regulatory sequences should enter the database available to phylogenetic inference, and greater attention should be paid to floral developmental sequences: as traditional systematics is mainly founded on reproductive structures, the study of floral homeotic genes will provide the still missing link between DNA and complex morphological features.

References

ALBERT, V. A., BACKLUND, A., BREMER, K., CHASE, M. W., MANHART, J. R., MISHLER, B. D. & K. C. NIXON (1994). Functional constraints and rbcL evidence for land plant phylogeny. *Ann. Missouri Bot. Gard.* **81**: 534-567.

——, CHASE, M. W. & B. D. MISHLER (1993). Character state weighting for cladistic analysis of protein-coding DNA sequences. *Ann. Missouri Bot. Gard.* **80**: 753-766.

——, MISHLER, B. D. & M. W. CHASE (1992). Character state weighting for restriction site data in phylogenetic reconstruction, with an example from chloroplast DNA. in: SOLTIS, P., SOLTIS, D. E. & J. J. DOYLE (eds.), *Molecular systematics of plants*. New York.

BALDWIN, B. G., SANDERSON, M. J., PORTER, M. J., WOJCIECHOWSKI, M. F. CAMPBELL, C. S. & M. J. DONOGHUE (1995). The ITS region of nuclear ribosomal DNA: a valuable source of evidence in angiosperm phylogeny. *Ann. Missouri Bot. Gard.* **82**: 247-277.

BOWMAN, C. M., BONNARD, G. & T. A. DYER (1983). Chloroplast DNA variation between species of Triticum and Aegilops: location of the variation on the chloroplast genome and its relevance to the inheritance and classification of the cytoplasm. *Theor. Appl. Genet.* **65**: 247-262.

——, KOLLER, B., DELIUS, H. & T. A. DYER (1981). A physical map of wheat chloroplast DNA showing the location of the structural genes for the ribosomal RNAs and the large subunit of ribulose 1,5 biphosphate carboxylase. *Mol. Gen. Genet.* **183**: 93-101.

BREMER, B. (1991). Restriction data from chloroplast DNA for phylogenetic reconstruction: Is there only one accurate way of scoring? *Pl. Syst. Evol.* **175**: 39-54.

—— & R. K. JANSEN (1992). A new subfamily of the Asteraceae. *Ann. Missouri Bot. Gard.* **79**: 414-415.

BRUNSFELD, S. J., SOLTIS, P. S., SOLTIS, D. E., GADEK, P. A., QUINN, C. J., STRENGE, D. D. & T. A. RANKER (1994). Phylogenetic relationships among genera of the conifer family Taxodiaceae and Cupressaceae: evidence from rbcL sequences. *Syst. Bot.* **19**: 253-262.

BURGER, W. (1977). The Piperales and the monocots: alternative hypotheses for the origin of monocotyledonous flowers. *Bot. Rev.* **43**: 345-393.

—— (1981). Heresy revived: the monocot theory of angiosperm origin. *Evol. Theory* **5**: 189-225.

CAPUTO, P., STEVENSON, D. W. & A. MORETTI (1992). Sharpening Occam's razor: an attempt to apply parsimony to competing hypotheses. *Brittonia* **44**: 376-382.

CAPUTO, P. , STEVENSON, D. W. & E. T. WURTZEL (1991). A phylogenetic analysis of American Zamiaceae (Cycadales) using chloroplast DNA restriction fragment length polymorphisms. *Brittonia* **73**: 135-145.

CHASE, M. W. & 41 others (1993). Phylogenetic analysis of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Missouri Bot. Gard.* **80**: 528-580.

CLARK, W. D., GAUT, B. S., DUVALL, M. R. & M. T. CLEGG (1993). Phylogenetic relationships of the Bromeliiflorae-Commeliniflorae-Zingiberiflorae complex of monocots based on rbcL sequence comparisons. *Ann. Missouri Bot. Gard.* **80**: 987-998.

CLEGG, M. T. (1993). Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 363-367.

——, LEARN, G. H. & E. M. GOLENBERG. (1990). Molecular evolution of chloroplast DNA in: R. K. SELANDER, A. G. CLARK & T. S. WHITTAM (eds.), *Evolution at the molecular level*. Sunderland (MA).

CONTI, E., FISHBACK, A. & K. J. SYTSMA (1993). Tribal relationships in Onagraceae: implications from rbcL data. *Ann. Missouri Bot. Gard.* **80**: 672-685.

CRONQUIST, A. (1981). *An integrated system of classification of flowering plants*. New York.

DEPAMPHILIS, C. W. & J. D. PALMER (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348**: 337-339.

DOEBLEY, J., DURBIN, H., GOLENBERG, E. D., CLEGG, M. T. & D. P. MA (1990). Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequences among the grasses (Gramineae). *Evolution* **44**: 1097-1108.

DONOGHUE, M. J. (1994). Progress and prospects in reconstructing plant phylogeny. *Ann. Missouri Bot. Gard.* **81**: 405-418.

——, OLMSTEAD, R. G., SMITH, J. F., & J. D. PALMER (1992). Phylogenetic relationships of Dipsacales based on rbcL sequences. *Ann. Missouri Bot. Gard.* **79**: 333-345.

—— & M. J. SANDERSON (1992). The suitability of molecular and morphological evidence in reconstructing plant phylogeny in: P. SOLTIS, D. E. SOLTIS & J. J. DOYLE (eds.), *Molecular systematics of plants*. New York.

DOWNIE, S. R. & J. D. PALMER (1992a). Use of chloroplast DNA rearrangements in reconstructing plant phylogeny in: P. S. SOLTIS, D. E. SOLTIS & J. J. DOYLE, (eds.): *Molecular systematics of plants*. London.

—— & J. D. PALMER (1992b). Restriction site mapping of the chloroplast DNA inverted repeat: a molecular phylogeny of the Asteridae. *Ann. Missouri Bot. Gard.* **79**: 266-283.

DOYLE, J. J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. *Syst. Bot.* **17**: 144-163.

——, DONOGHUE, M. J. & E. Z. ZIMMER (1994). Integration of morphological and ribosomal RNA data on the origin of Angiosperms. *Ann. Missouri Bot. Gard.* **81**: 419-450.

DRESSLER, R. L. & M. W. CHASE (1995). Whence the Orchids? in: P. J. RUDALL, P. J. CRIBB, D. F: CUTLER & C. J. HUMPHRIES (eds.), *Monocotyledons: systematics and evolution*. Kew (UK).

DUVALL, M. R., CLEGG, M. T., CHASE, M. W., CLARK, W. D., KRESS, W. J., HILLS, H. G., EGUIARTE, L. E., SMITH, J. F., GAUT, B. S., ZIMMER, E. A. & G. H. LEARN (1993). Phylogenetic hypothesis for the monocotyledons constructed from the rbcL sequence data. *Ann. Missouri Bot. Gard.* **80**: 607-619.

FARRIS, J. S. (1970). Methods for computing Wagner trees. *Syst. Zool.* **19**: 83-92.

—— (1977). Phylogenetic analysis under Dollo's law. *Syst. Zool.* **26**: 77-88.

FELSENSTEIN, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**: 401-410.

Fitch, W. M. (1971). Towards defining the course of evolution: minimal changes for a specific tree topology. *Syst. Zool.* **20**: 406-416.

GADEK, P. A. & C. J. QUINN (1993). An analysis of relationships within Cupressaceae sensu stricto based on rbcL sequences. *Ann. Missouri Bot. Gard* **80**: 581-586.

GIANNASI, D. E., ZURAWSKI, G., LEARN, G. H. & M. T. CLEGG (1992). Evolutionary relationships of the Caryophyllidae based on comparative rbcL sequences. *Syst. Bot.* **17**: 1-5.

GOLENBERG, E. M., GIANNASI, D. E., CLEGG, M. T., SMILEY, C. J., DURBIN, M., HENDERSON, D. & G. ZURAWSKI (1990). Chloroplast DNA sequence from a Miocene Magnolia species. *Nature* **344**: 656-658.

HAMBY, R. K. & E. A. ZIMMER (1992). Ribosomal RNA as a phylogenetic tool in plant systematics. in: P. SOLTIS, D. E. SOLTIS & J. J. DOYLE (eds.), *Molecular systematics of plants*. New York.

HASEBE, M., KOFUJI, K., ITO, M., KATO, M., IWATSUKI, K. & K. UEDA (1992). Phylogeny of gymnosperms inferred from rbcL gene sequences. *Bot. Mag.* (Tokyo) **105**: 673-679.

HEDRÉN, M., CHASE, M. W. & R. G. OLMSTEAD (1995). Relationships in the Acanthaceae and related families as suggested by cladistic analysis of rbcL nucleotide sequences. *Pl. Syst. Evol.* **194**: 93-109.

HEMPEL, A. L., REEVES, P. A., OLMSTEAD, R. G. & R. K. JANSEN (1995). Implications of rbcL sequence data for higher order relationships of the Loasaceae and the anomalous aquatic plant Hydrostachys (Hydrostachyaceae). *Pl. Syst. Evol.* **194**: 25-37.

HIRATSUKA, J., SHIMADA, H., WHITTIER, R., ISHIBASHI, T., SAKAMOTO, M., MORI, M., KONDO, C., HONJI, Y., SUN, C.-R., MENG, B.-Y., LI, Y.-Q., KANO, A., NISHIZAWA, Y., HIRAI, A., SHINOZAKI, K. & M. SUGIURA (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.

HOOT, S. B., CULHAM, A. & P. R. CRANE (1995). The utility of atpB gene sequences in resolving phylogenetic relationships: comparison with rbcL and 18S ribosomal DNA sequences in Lardizabalaceae. *Ann. Missouri Bot. Gard.* **82**: 194-207.

JANSEN, R. K., MICHAELS, H. J., WALLACE, R. S., KIM, K. -J., KEELEY, S., WATSON, L. & J. D. PALMER (1992). Chloroplast DNA variation in the Asteraceae: phylogenetic and evolutionary implications. in: P. SOLTIS, D. E. SOLTIS & J. J. DOYLE (eds.), *Molecular systematics of plants*. New York.

JOHANNSON, J. T. & R. K. JANSEN (1993). Chloroplast DNA variation and phylogeny of the Ranunculaceae. *Pl. Syst. Evol.* **187**: 29-49.

JOHNSON, L. A. & D. E. SOLTIS (1995). Phylogenetic inference on in Saxifragaceae sensu stricto and Gilia (Polemoniaceae) using matK sequences. *Ann. Missouri Bot. Gard.* **82**: 149-175.

KIM, Y.-D. & R. K. JANSEN (1994a). Comparison of phylogenetic hypotheses among different data sets in dwarf dandelions (Krigia, Asteraceae): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. *Pl. Syst. Evol.* **190**: 157-185.

—— & R. K. JANSEN (1994b). Characterization and phylogenetic distribution of a chloroplast DNA arrangement in the Berberidaceae. *Pl. Syst. Evol.* **193**: 107-114.

——, JANSEN, R. K., WALLACE, R. S., MICHAELS, H. J. & J. D. PALMER (1992). Phylogenetic implications of rbcL sequence variation in the Asteraceae. *Ann. Missouri Bot. Gard.* **79**: 428-445.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge.

KRON, K. A. & M. W. CHASE (1993). Systematics of the Ericaceae, Empetraceae Epacridaceae and related taxa based upon rbcL sequence data. *Ann. Missouri Bot. Gard.* **80**: 735-741.

LAKE, J. A. (1987). A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol. Biol. Evol.* **4**: 167-191.

LAVIN, M., DOYLE., J. J. & J. D. PALMER (1990). Evolutionary significance of the loss of chloroplast inverted repeat in the Leguminosae subfamily Papilionoideae. *Evolution* **44**: 390-402.

LE QUESNE, W. J. (1974). The uniquely evolved character concept and its cladistic application. *Syst. Zool.* **23**: 513-517.

LOCONTE, H. & D. W. STEVENSON (1991). Cladistics of the Magnoliidae. *Cladistics* **7**: 267-296.

MANEN, J.-F., NATALI, A. & F. EHRENDORFER (1994). Phylogeny of Rubiaceae-Rubieae inferred from the sequence of a cpDNA intergene region. *Pl. Syst. Evol.* **190**: 195-211.

MATHEWS, S., LAVIN, M. & R. A. SHARROCK (1995) Evolution of the phytocrome gene family and its utility for phylogenetic analyses of angiosperms. *Ann. Missouri Bot. Gard.* **82**: 296-321.

MICHAELS, H. J., SCOTT, K. M., OLMSTEAD, R. G., SZARO, T., JANSEN, R. K. & J. D. PALMER (1993). Interfamilial relationships of the Asteraceae: insights from rbcL sequence variation. *Ann. Missouri Bot. Gard.* **80**: 742-751.

MISHLER, B. D., LEWIS, L. A., BUCHHEIM, M. A., RENZAGLIA, K. S., GARBARY, D. J., DELWICHE, C. F., ZECHMAN, F. W., KANTZ, T. S. & R. L. CHAPMAN (1994). Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. Missouri Bot. Gard.* **81**: 451-483.

MORETTI, A., CAPUTO, P., COZZOLINO, S., DE LUCA, P., GAUDIO, L., SINISCALCO GIGLIANO, G. & D. W. STEVENSON (1993). A phylogenetic analysis of Dioon (Zamiaceae). *Amer. J. Bot.* **80**: 204-214.

MORGAN, D. R., SOLTIS, D. E. & K. R. ROBERTSON (1994). Systematic and evolutionary implications of rbcL variation in Rosaceae. *Amer. J. Bot.* **81**: 890-903.

NADOT, S., BAJON, R. & B. LEJEUNE (1994). The chloroplast gene rps4 as a tool for the study of Poaceae phylogeny. *Pl. Syst. Evol.* **191**: 27-38.

NICKRENT, D. L. & D. E. SOLTIS (1995). A comparison of angiosperm phylogenies from nuclear 18S rDNA and rbcL sequences. *Ann. Missouri Bot. Gard.* **82**: 208-234.

OHYAMA, K., FUKUZAWA, H., KOHCHI, T., SHIRAI, H., SANO, T., SANO, S., UMESONO, K., SHIKI, Y., TAKEUCHI, M., CHANG, Z., AOTA, S., INOKUCHI, H. & H. OZEKI (1986). Chloroplast gene organization deduced from the complete sequence of liverwort Marchantia polymorpha chloroplast DNA. *Nature* **322**: 572-574.

OLMSTEAD, R. G., BREMER, B., SCOTT, K. M. & J. D. PALMER (1993). A parsimony analysis of the Asteridae sensu lato based on rbcL sequences. *Ann. Missouri Bot. Gard.* **80**: 700-722.

——, MICHAELS, H. J., SCOTT, K. M. & J. D. PALMER (1992). Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of rbcL. *Ann. Missouri Bot. Gard.* **79**: 249-265.

—— & J. D. PALMER (1992). A chloroplast DNA phylogeny of Solanaceae: subfamilial relationships and character evolution. *Ann. Missouri Bot. Gard.* **79**: 346-360.

—— & J. D. PALMER (1994). Chloroplast DNA systematics: a review of methods and data analysis. *Amer. J. Bot.* **81**: 1205-1224.

—— & P. A. REEVES Evidence for polyphyly of the Scrophulariaceae based on chloroplast rbcL and ndhF sequences. *Ann. Missouri Bot. Gard.* **82**: 176-193.

—— & J. A. SWEERE Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43**: 467-481.

PAGE, R. D. M. (1993). Genes, organisms, and areas: the problem of multiple lineages. *Syst. Biol.* **42**: 77-84.

PALMER, J. D. (1982). Physical and gene mapping of chloroplast DNA from Atriplex triangularis and Cucumis sativa. *Nucleic Acids Res.* **10**: 1593-1605.

—— (1985). Comparative organization of chloroplast genomes. *Annual Rev. Genet.* **19**: 325-354.

—— (1986). Isolation and structural analysis of chloroplast DNA. *Meth. Enzymol.* **118**: 167-186.

PALMER, J. D. (1987). Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Amer. Naturalist* **130**: S6-S29.

—— (1992). Mitochondrial DNA in plant systematics: applications and limitations in: P. S. SOLTIS,, D. E. SOLTIS & J. J. DOYLE (eds.), *Molecular systematics of plants*. London.

——, JORGENSEN, R. A. & W. F. THOMPSON (1985). Chloroplast DNA variation and evolution in Pisum: patterns of change and phylogenetic analysis. *Genetics* **109**: 195-213.

——, NUGENT, J. M. & L. A. HERBON (1987). Unusual structure of geranium chloroplast DNA: a triple-sized inverted repeat, extensive gene duplications, multiple inversions and two repeat families. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 769-773.

——, OSORIO, B., ALDRICH, J. & W. F. THOMPSON (1987). Chloroplast DNA evolution among legumes: loss of a large inverted repeat occurred prior to other sequence rearrangements. *Curr. Genet.* **11**: 275-286.

——, SHIELDS, C. R., COHEN, D. B. & T. J. ORTON (1983). Chloroplast DNA evolution and the origin of amphidiploid Brassica species. *Theor. Appl. Genet.* **71**: 417-429.

—— & D. B. STEIN (1986). Conservation of chloroplast genome structure among vascular plants. *Curr. Genet.* **10**: 823-833.

—— & W. F. THOMPSON (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **29**: 537-550.

—— & D. ZAMIR (1982). Chloroplast DNA evolution and phylogenetic relationships in Lycopersicon. *Proc. Natl. Acad. U.S.A.* **79**: 5006-5010.

PERL-TREVES, R. & E. GALUN (1985). The Cucumis plastome: physical map, intrageneric variation and phylogenetic relationships. *Theor. Appl. Genet.* **71**: 417-429.

PLUNKETT, G. M., SOLTIS, D. E., SOLTIS, P. S. & R. E. BROOKS (1995). Phylogenetic relationships between Juncaceae and Cyperaceae: insights from rbcL sequence data. *Amer. J. Bot.* **82**: 520-525.

POULSEN, C. (1983). The barley chloroplast genome: physical structure and transcriptional activity in vivo. *Carlsberg Res. Comm.* **48**: 57-80.

PRICE, R. A. & J. D. PALMER (1993). Phylogenetic relationships of the Geraniaceae and Geraniales from rbcL sequence comparison. *Ann. Missouri Bot. Gard.* **80**: 661-671.

QIU, Y. -L., CHASE, M. W., LES, D. H. & C. R. PARKS (1993). Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene rbcL. *Ann. Missouri Bot. Gard.* **80**: 587-606.

RAUBESON, L. A. & R. K. JANSEN (1992). A rare chloroplast-DNA mutation is shared by all conifers. *Biochem. Syst. Ecol.* **20**: 17-24.

RETTIG, J. H., WILSON, H. D. & H. D. MANHART (1992). Phylogeny of Caryophyllales - gene sequence data. *Taxon* **41**: 201-209.

ROGERS, S. O. & A. J. BENDICH (1987). Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Mol. Biol.* **9**: 509-520.

SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIDA, N., MATSUBUYASHI, 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., KATO, A., TOHDOH, N., SHIMODA, H. & M. SUGIURA (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5**: 2043-2049.

SHINWARI, Z. K., KATO, H., TERAUCHI, R. & S. KAWANO (1994). Phylogenetic relationships among genera in the Liliaceae-Asparagoideae-Polygonatae s.l. inferred from rbcL gene sequence data. *Pl. Syst. Evol.* **192**: 263-277.

SMITH, J. K., KRESS, W. J. & E. A. ZIMMER (1993). Phylogenetic analysis of the Zingiberales based on rbcL sequences. *Ann. Missouri Bot. Gard.* **80**: 620-630.

SOLTIS, D. E., MORGAN, D. R., GRABLE, A., SOLTIS, P. S. & R. KUZOFF (1993). Molecular systematics of Saxifragaceae sensu stricto. *Amer. J. Bot.* **80**: 1056-1081.

——, XIANG, Q. -Y & L. HUFFORD (1995). Relationships of the Hydrangeaceae based on rbcL sequence data. *Amer. J. Bot.* **82**: 504-514.

STEELE, K. P., HOLSINGER, K., JANSEN, R. & D. TAYLOR (1991). Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. *Molec. Biol. Evol.* **8**: 240-248.

STRAUSS, S. H. & A. H. DOERKSEN (1990). Restriction fragment analysis of pine phylogeny. *Evolution* **44**: 1081-1096.

SUH, Y., THIEN, H. E. & E. A. ZIMMER (1993). Molecular evolution and phylogenetic implications of internal transcribed sequences of ribosomal DNA in Winteraceae. *Amer. J. Bot.* **80**: 1042-1055.

SWOFFORD, D. L. & G. J. OLSEN (1990). Phylogeny reconstruction in: D. M. HILLIS & C. MORITZ (eds.), Molecular systematics. Sunderland (Ma).

TAKHTAJAN, A. (1987). *Systema Magnoliophytorum*. Leningrad (in Russian).

TAYLOR, D. W. & L. J. HICKEY (1992). Phylogenetic evidence for the herbaceous origin of angiosperms. *Pl. Syst. Evol.* **180**: 137-156.

TIMOTHY, D. H., LEVINGS, C. S., PRING, D. R., CONDE, M. F. & J. L. KERMICLE (1979). Organelle DNA variation and systematic relationships in the genus Zea: Teosinte. *Proc. Natl. Acad. Sci. U.S.A.* **76**: 4220-4224.

UHL, N. W., DRANSFIELD, J., DAVIS, J. I., LUCKOW, M. A., HANSEN, K. S. & J. J. DOYLE (1995). Phylogenetic relationships among palms: cladistic analyses of morphological and chloroplast DNA restriction site variation. in: P. J. RUDALL, P. J. CRIBB, D. F. CUTLER & C. J. HUMPHRIES (eds.), *Monocotyledons: systematics and evolution*. Kew (UK).

WAGSTAFF, S. J., OLMSTEAD, R. G. & P. D. CANTINO (1995). Parsimony analysis of cpDNA restriction site variation in subfamily Nepetoideae (Labiatae). *Amer. J. Bot.* **82**: 886-892.

WATERS, E. R. (1995) An evaluation of the small heat shock genes for phylogenetic analysis in plants. *Ann. Missouri Bot. Gard.* **82**: 278-295.

WILLIAMS, S. E., ALBERT, V. A. & M. W. CHASE (1994). Relationships of Droseraceae: a cladistic analysis of rbcL sequence and morphological data. *Amer. J. Bot.* **81**: 1027-1037.

WOLFE, K. H., LI, W.-H. & P. M. SHARP (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 9054-9058.

XIANG, Q. -Y., SOLTIS, D. E., MORGAN, D. R. & P. S. SOLTIS 1993. Phylogenetic relationships of Cornus L. sensu lato and putative relatives inferred from rbcL data. *Ann. Missouri Bot. Gard.* **80**: 723-734.

ZUCKERANDL, E. & L. PAULING (1965). Evolutionary divergence and convergence in proteins in: V. BRYSON & H. J. VOGEL (eds.), *Evolving genes and Proteins*. New York.

Address of the author:

Prof. P. Caputo, Dipartimento di Biologia vegetale, Università degli Studi di Napoli "Federico II", Via Foria, 223. 80139 Napoli, Italy.