Zeitschrift: Archives des sciences et compte rendu des séances de la Société

Band: 49 (1996)

Heft: 2: Archives des Sciences

Artikel: A phylogenetic analysis of plants, using the chloroplast gene rps4 and the anataxis method

Autor: Bittar, Gabriel / Carter, Leigh / Nadot, Sophie

DOI: https://doi.org/10.5169/seals-740420

Archs Sci. Genève | Vol. 49 | Fasc. 2 | pp. 149-157 | Septembre 1996

Communication présentée à la séance du 8 février 1996

A PHYLOGENETIC ANALYSIS OF PLANTS, USING THE CHLOROPLAST GENE *rps*4 AND THE ANATAXIS METHOD

BY

Gabriel BITTAR*, Leigh CARTER, Sophie NADOT, Tatiana SOUZA-CHIES, Alexis EVRARD, Evelyne BESIN & Bernard LEJEUNE

ABSTRACT

A phylogenetic analysis of plants, using the chloroplast gene *rps*4 and the anâtaxis method. - Practical and theoretical advantages in using a protein-coding chloroplastic gene, and a new evolutionary tree reconstruction method, are discussed. Some results are presented for different taxonomic levels in the realm of plants. -

We have briefly seen in the preceding paper what are the main problems associated with the principal types of phylogenetic inference and tree reconstitution. We have also seen that numerical phenetic taxonomy methods are by far the quickest. Consequently, because of the inherent quality of quickness of distances methods, it was thought worth to develop and improve that kind of methods, in such a way that they would no more be affected by the systematic biases we have described, and thus be more accurate (Bittar & Carter, 1994; Bittar, 1995).

Thus, as an alternative to quick but biased distances-matrix phenetical approaches, a new kind of trees-compatibility distance method has been developed, that groups taxa while taking into account both the possibility of homoplasy and of different rates of evolution within different branches. The program in which this method has been implemented, named **Anâtaxis**, is the most general purpose program in the new Vivâras phylogenetic package; it can be applied to any kind of data obtained from evolving objects, with the sole conditions that one of these can be defined as outgroup to the others and that one can construct a symmetrical matrix of distances.

The objective of Anâtaxis is to allow a quick analysis of a huge set of data by using all possible information without having to analyse the possible evolutionary story of each site (as cladistic parsimony methods do), while nevertheless avoiding the phenetics pitfalls of widely differing evolutionary rates in the different lineages. It should be noted that the Anâtaxis trees indicate polychotomies when information is not judged sufficient for defining nodes, in order to avoid the unnecessary display of uncertain or un-

---

* G. Bittar, ch. du Pont-Noir 9A, CH-1226 Thônex.

supported phylogenetic relationships. The Anâtaxis method is quick and efficient, and the robustness of the trees produced by the program has been demonstrated on different sets of data.

We present here the phylogenetic results that we have obtained with Anâtaxis during the analysis of a short (*circa* 600 bases) chloroplastic gene, the ribosomal protein small subunit-4 (*rps*4). This gene presents the advantage of not seeming liable to much biochemical or natural environment pressures, thus, relatively to most other protein-coding genes, generally devoid of strong homoplasy phenomena. Molecular information is particularly useful for understanding the evolution of plants because of the high degree of morphological homoplasy within this realm, where individual cells display a profound plasticity of form and function and where continual organogenesis is the rule. In terms of developmental and genetic strategies, plants are very dynamic; genotypic and epigenetic changes are readily generated, and because (contrary to animals) they do not display a germ line stricto sensu, selection of cells with new alleles can occur within the soma and (in a rather Lamarckian way...) be transmitted to progeny. Another strong argument in favour of establishing plant molecular phylogenies is the limited help that plant (contrary to animal) palæontology brings, because of the paucity of plant fossils, especially flower fossils that are the most useful in taxonomic terms.

Although the organellar (mitochondrial and, in plants, chloroplastic) genomes can exchange DNA with the nucleus of the host, this is seemingly always in the organellar to nucleus direction (according to the endosymbiont theory of organelles origins, the genomes of mitochondria and chloroplasts have progressively been depleted by transfer of organellar genes to the nucleus genome). Moreover, the organellar genomes of plants do not seem to be liable to infection by viruses and hence are not easily open to inter-specific DNA transfer (introgression), which could render the evolutionary diagrams reticulate (complicated networks rather than trees). Partially for these reasons, contrary to most nuclear (and plant mitochondrial) genomes, chloroplastic genomes do not contain many foreign sequences, and, contrary to most nuclear protein-coding genes (where homologous alignment, implicating both orthology and paralogy, can be fairly more complex), do not contain multigene families. Quite generally, the argument of proper homologous alignment is in favour of protein-coding genes, where ambiguity in the positioning of the nucleotide bases forming the characters is diminished thanks to the possibility of prior alignment of the polypeptide chains obtained from translation of the nucleotide chains. Finally, except for two cases of conversion of an ACG threonine codon to an AUG initiation codon (in *rpl*2 of maize and *psb*I of tobacco), and contrary to what happens on a rather large scale in plant mitochondria, not much mRNA editing has been observed between transcription and translation of chloroplastic genes.

For all these reasons, chloroplastic protein-coding genes are very useful instruments for establishing plant phylogenies: in particular, *rbc*L has legitimately become very popular, and *rps*4 has been demonstrated by us to be a particularly efficient tool for that purpose. We now have at our disposal the complete or nearly complete *rps*4 gene of 142 species (representing 133 distinct sequences because some species have identical rps4), and more are being sequenced. Except the sequences for *Marchantia*, *Pinus*, *Epifagus*, *Nicotiana*, *Spinacia*, *Zea* and *Oryza* (EMBL/GenBank accession numbers X04465, D17510, M81884, Z00044, M16878, X10608 and X15901), they have all been sequenced at the BMV-IBP laboratory of professor B. Lejeune (Université Paris-Sud, Orsay). The species of which the *rps*4 has been sequenced are mostly angiosperms, and particularly monocots (figure 3), with emphasis on the Iridaceae (fig. 4) and Poaceae (fig. 5) families (among dicots, an accent has been put on Oleaceae - fig. 2). It can be noted that multiple alignment of all 133 distinct sequences implies a 660 characters matrix, of which 193 characters are invariant.

We are clearly conscious of the necessity to carefully distinguish between classification (taxonomy) and phylogenesis (cladogenesis generally, though it may be more complicated than this). Thus our work is not meant for reorganising botanical classification, at least not at this modest level of taxa and genes sampling. A classification must be convenient and efficient for information storage and transmittal: it is mainly aimed at naturalists. A taxon is thus defined from both its inherent characteristics and its evolutionary origin, hence taking into account not only the synapomorphies and symplesiomorphies of its members, but also the autapomorphies that are important. Accordingly, a taxon cannot be polyphyletic because, in so being, its members would have been grouped together on the basis of independent apparitions of shared characters - through convergence (parallelism) or reversion -, which would not only be artificial but also biologically misleading. On the other hand a taxon may be monophyletic (Ernst Haeckel, 1868: "Natürliche Schöpfungsgeschichte") or paraphyletic, i.e. in the latter case it can if necessary be a grouping of symplesiomorphic clades that excludes (because of their peculiar autapomorphies) one or many parent clades which also originate from the direct ancestor to the taxon and the "outcasts". Hence, in case of practical necessity, paraphyletic taxa may be taxonomically acceptable, as long as classification is not confused with phylogeny: branching not being the whole of evolution, an analysis of evolutionary divergence (anagenesis) must also be done (this is the work of evolutionary systematists, or syncretistic taxonomists).

Bearing this in mind, we can point to some peculiar results of our analysis that illustrate these notions (there is no room here for a complete discussion), and that demonstrate how phylogenetic analysis bears on taxonomy.

Within the graminid (Aveneae, Poeae) clade (fig. 5), the Aveneae tribe forms an acceptable taxon, despite the Aveneae III and IV branches forming a clade with the Poeae tribe : the Aveneae form a paraphyletic tribe. Another interesting example within the Gramineae is the case of the *Brachypodium* genus: the two Brachypodium species cannot be members of the Triticeae tribe, because otherwise Bromus also should, but it could be a member of the Triticanae supertribe to which *Bromus* clearly belongs.

Within the Iridaceae (fig. 4), the *Patersonia*, *Nivenia* and *Aristea* genera, though clearly not forming a clade, can be considered as forming a paraphyletic subfamily (Nivenioideae), and *Isophysis* is clearly a peculiar member of this family. Among angiosperms, the Anâtaxis result shows a dichotomous separation between dicots (despite ambiguous *Laurus*) and monocots (despite ambiguous *Acorus*), seeming to confirm that the Magnoliopsida and Liliopsida do form two distinct sister clades (fig. 1).

As a last example, we shall mention that the position of *Welwitschia* is apparently in favour of Chlamydospermae being parent to Angiospermae first, not to Gymnospermae (fig. 1); this has been the subject of much controversy, and clearly should not be considered as a firm conclusion from a single result from a single gene; other genes should be used in the comparison, and in the future the *rps*4 sequences of *Gnetum* and *Ephedra* may help us confirm this preliminary result.

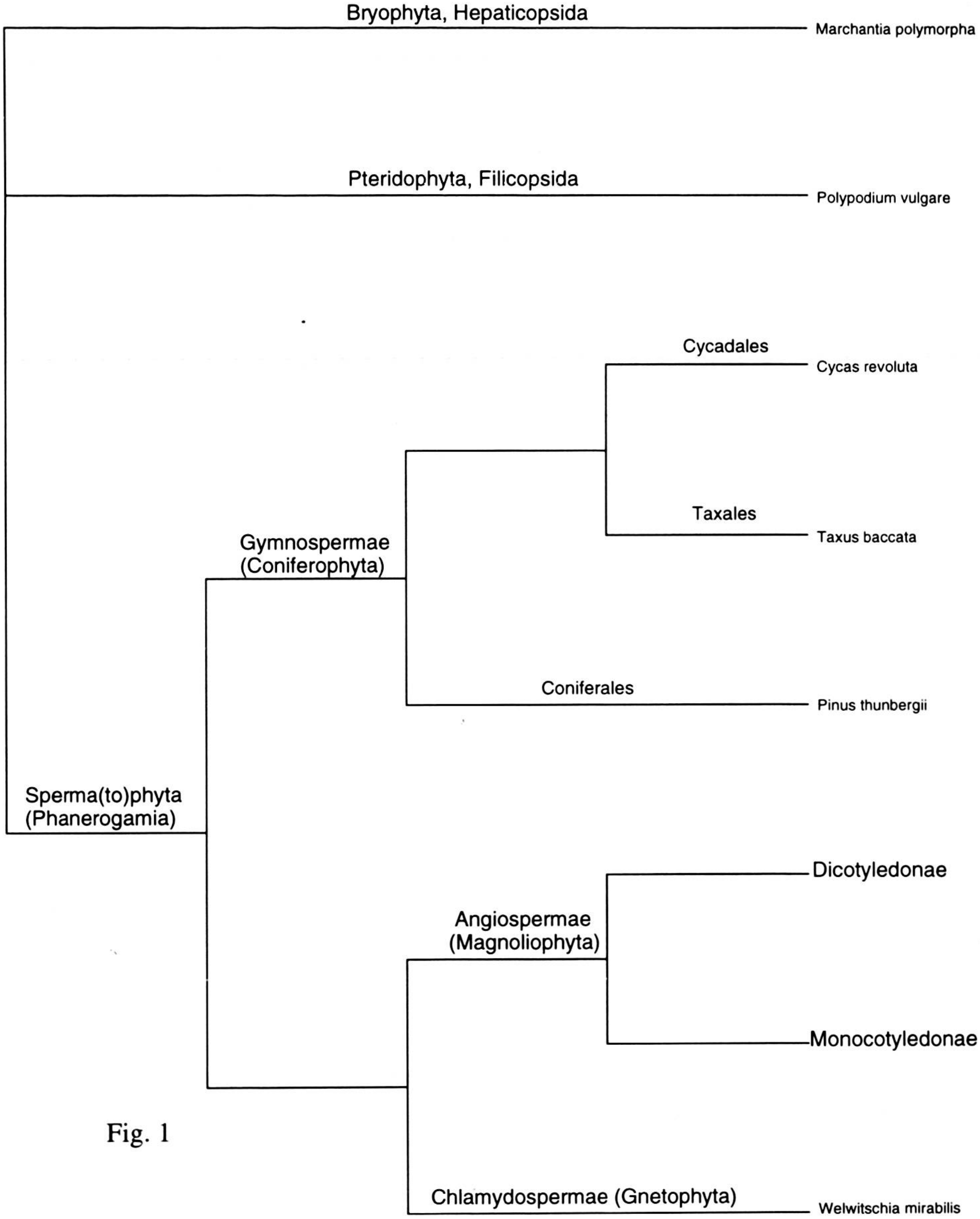
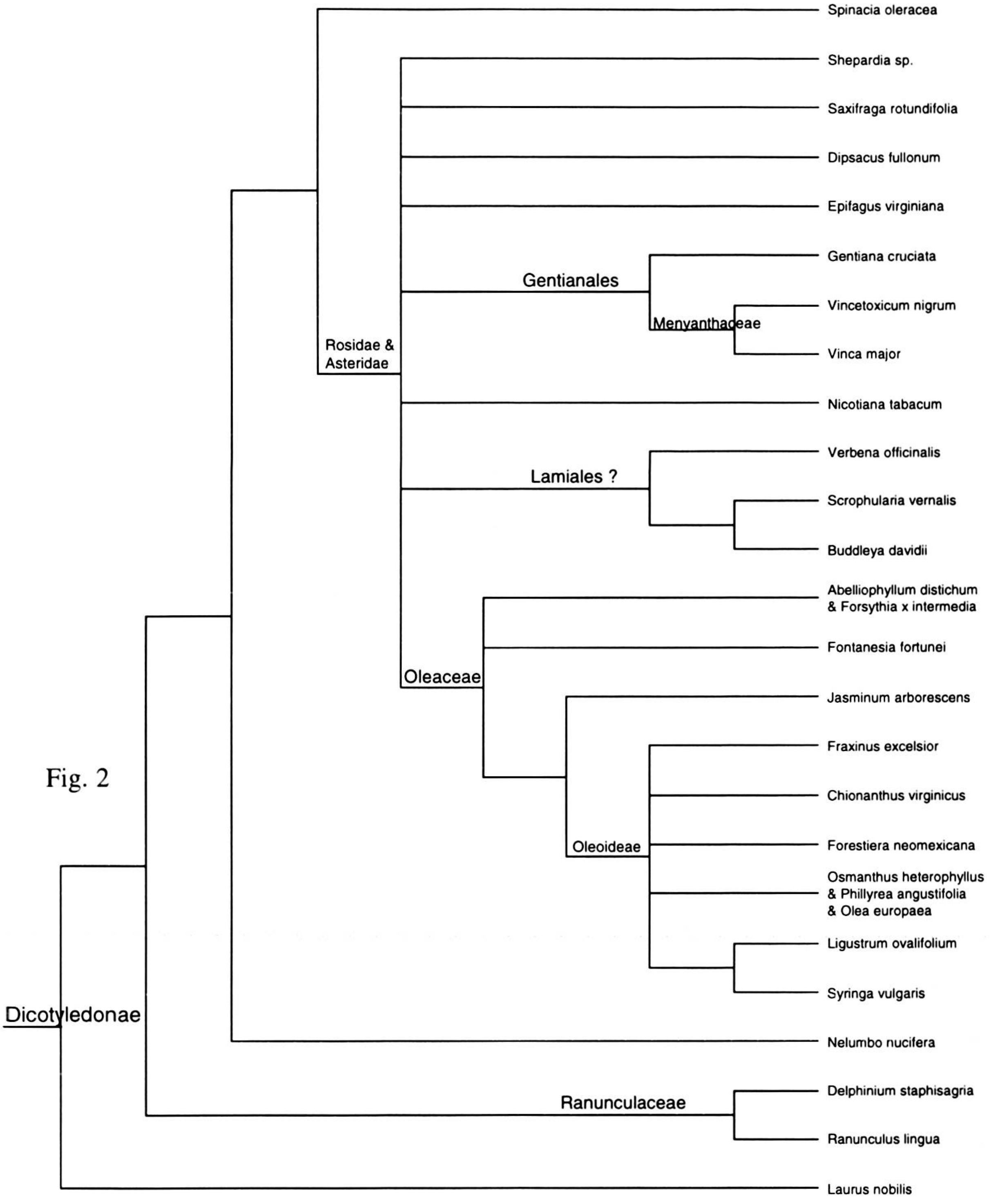


Fig. 1

Fig. 2


Spinacia oleracea
Shepardia sp.
Saxifraga rotundifolia
Dipsacus fullonum
Epifagus virginiana
Gentianales
Gentiana cruciata
Menyanthaceae
Vincetoxicum nigrum
Vinca major
Rosidae & Asteridae
Nicotiana tabacum
Lamiales ?
Verbena officinalis
Scrophularia vernalis
Buddleya davidii
Abelliophyllum distichum & Forsythia x intermedia
Fontanesia fortunei
Oleaceae
Jasminum arborescens
Fraxinus excelsior
Chionanthus virginicus
Oleoideae
Forestiera neomexicana
Osmanthus heterophyllus & Phillyrea angustifolia & Olea europaea
Ligustrum ovalifolium
Syringa vulgaris
Dicotyledonae
Nelumbo nucifera
Ranunculaceae
Delphinium staphisagria
Ranunculus lingua
Laurus nobilis

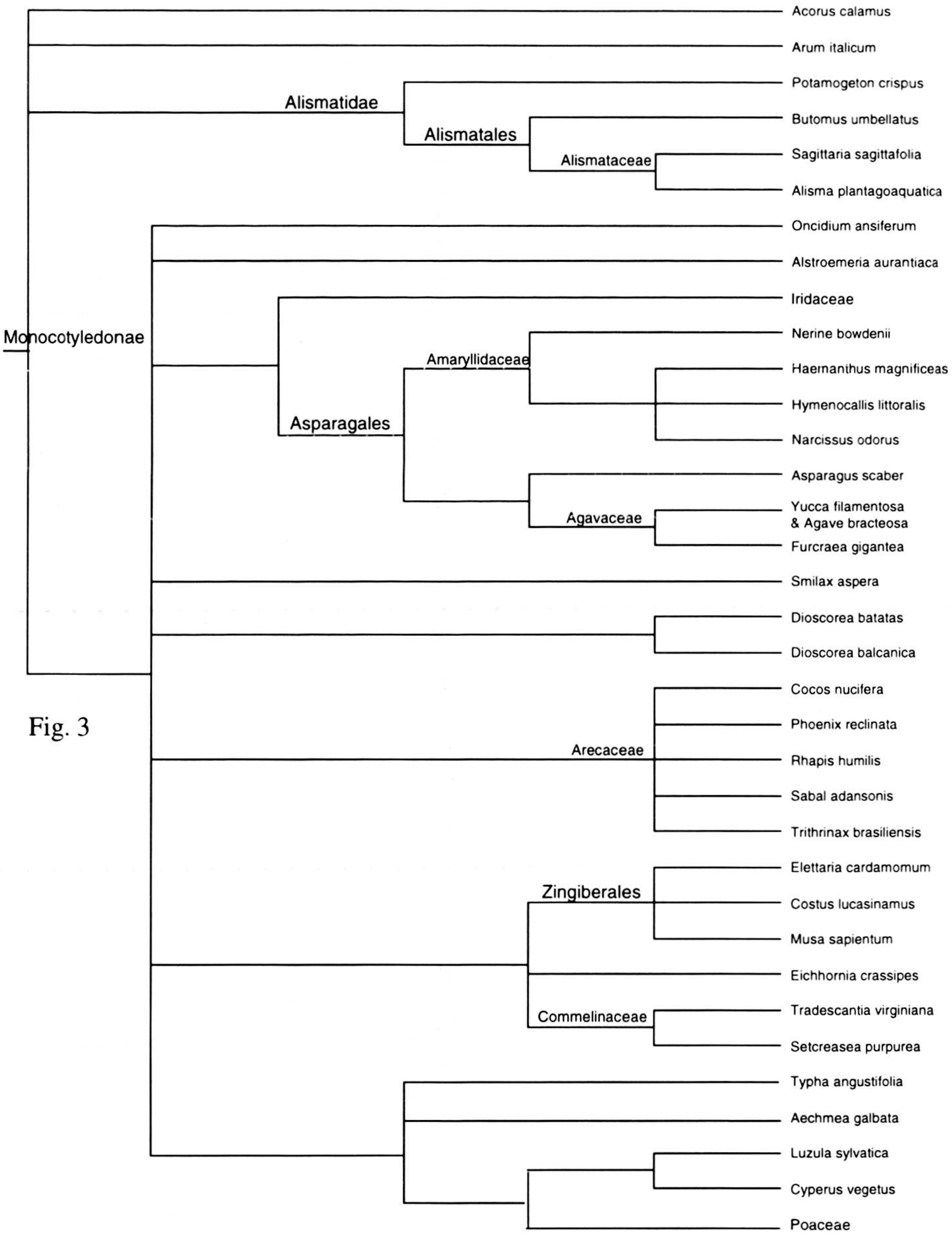
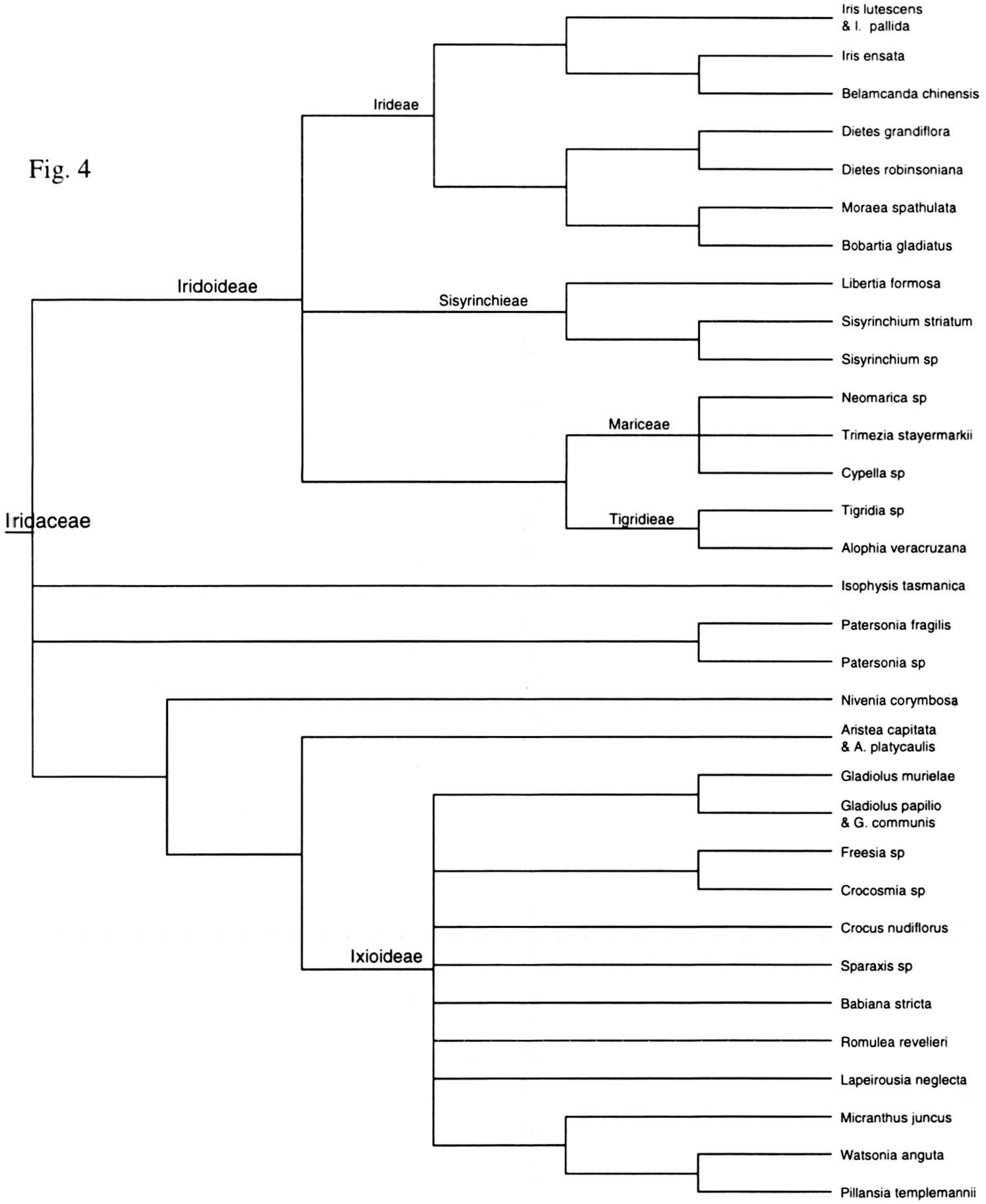

Acorus calamus
Arum italicum
Potamogeton crispus
Alismatidae
Butomus umbellatus
Alismatales
Alismataceae
Sagittaria sagittafolia
Alisma plantagoaquatica
Oncidium ansiferum
Alstroemeria aurantiaca
Iridaceae
Monocotyledonae
Nerine bowdenii
Amaryllidaceae
Haemanthus magnificeas
Hymenocallis littoralis
Asparagales
Narcissus odorus
Asparagus scaber
Yucca filamentosa
& Agave bracteosa
Agavaceae
Furcraea gigantea
Smilax aspera
Dioscorea batatas
Dioscorea balcanica
Cocos nucifera
Phoenix reclinata
Arecaceae
Rhapis humilis
Sabal adansonis
Trithrinax brasiliensis
Elettaria cardamomum
Zingiberales
Costus lucasinamus
Musa sapientum
Eichhornia crassipes
Tradescantia virginiana
Commelinaceae
Setcreasea purpurea
Typha angustifolia
Aechmea galbata
Luzula sylvatica
Cyperus vegetus
Poaceae

Fig. 3

Fig. 4


Iris lutescens & I. pallida
Iris ensata
Belamcanda chinensis
Irideae
Dietes grandiflora
Dietes robinsoniana
Moraea spathulata
Bobartia gladiatus
Iridoideae
Libertia formosa
Sisyrinchieae
Sisyrinchium striatum
Sisyrinchium sp
Neomarica sp
Mariceae
Trimezia stayermarkii
Cypella sp
Iridaceae
Tigridia sp
Tigridieae
Alophia veracruzana
Isophysis tasmanica
Patersonia fragilis
Patersonia sp
Nivenia corymbosa
Aristea capitata & A. platycaulis
Gladiolus murielae
Gladiolus papilio & G. communis
Freesia sp
Crocosmia sp
Crocus nudiflorus
Ixioideae
Sparaxis sp
Babiana stricta
Romulea revelieri
Lapeirousia neglecta
Micranthus juncus
Watsonia anguta
Pillansia templemannii

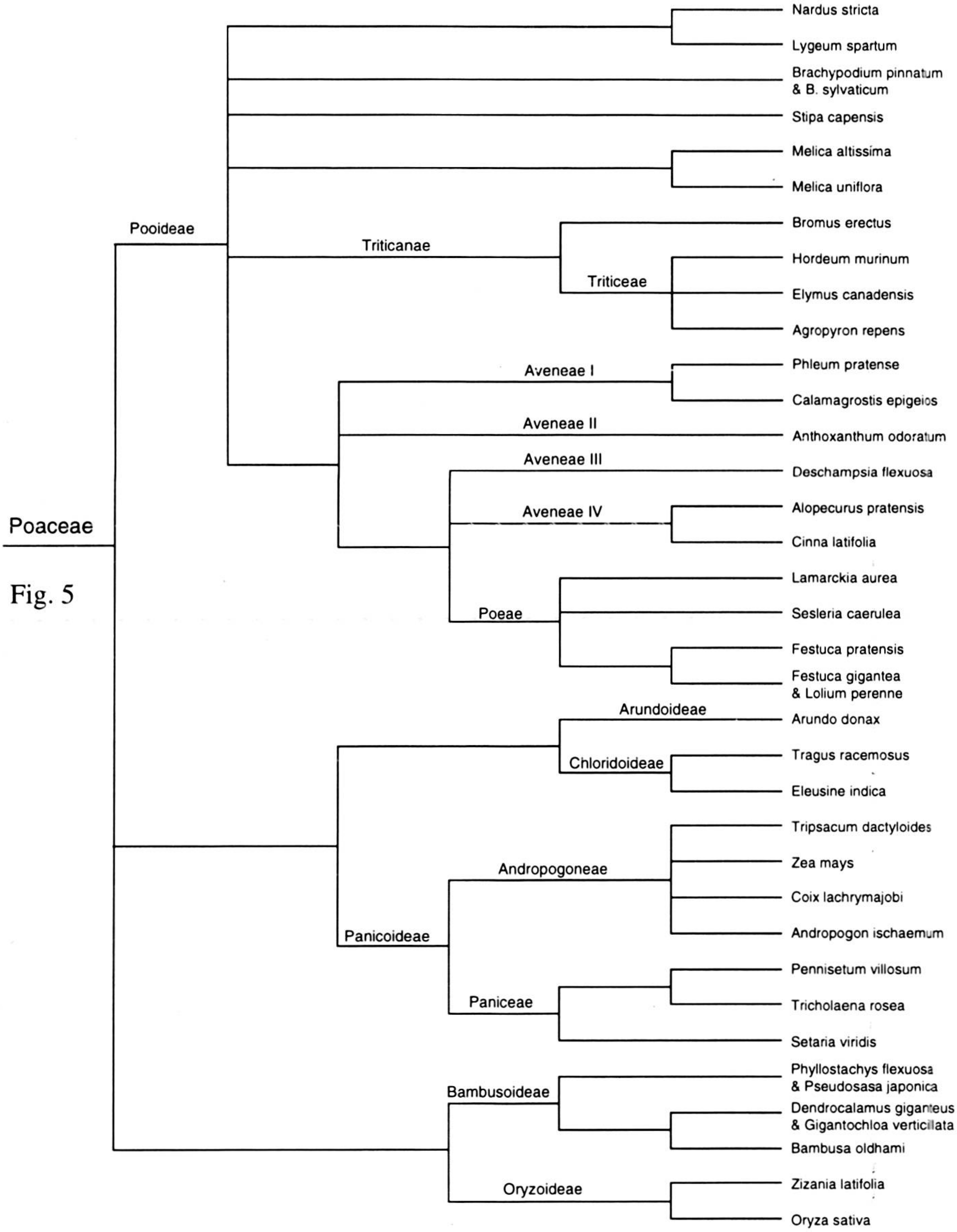

Nardus stricta
Lygeum spartum
Brachypodium pinnatum & B. sylvaticum
Stipa capensis
Melica altissima
Melica uniflora
Pooideae
Bromus erectus
Triticanae
Hordeum murinum
Triticeae
Elymus canadensis
Agropyron repens
Aveneae I
Phleum pratense
Calamagrostis epigeios
Aveneae II
Anthoxanthum odoratum
Aveneae III
Deschampsia flexuosa
Aveneae IV
Alopecurus pratensis
Poaceae
Cinna latifolia
Lamarckia aurea
Poeae
Sesleria caerulea
Festuca pratensis
Festuca gigantea & Lolium perenne
Arundoideae
Arundo donax
Chloridoideae
Tragus racemosus
Eleusine indica
Tripsacum dactyloides
Andropogoneae
Zea mays
Coix lachrymajobi
Andropogon ischaemum
Panicoideae
Pennisetum villosum
Paniceae
Tricholaena rosea
Setaria viridis
Bambusoideae
Phyllostachys flexuosa & Pseudosasa japonica
Dendrocalamus giganteus & Gigantochloa verticillata
Bambusa oldhami
Oryzoideae
Zizania latifolia
Oryza sativa

Fig. 5

RÉSUMÉ

Les avantages pratiques et théoriques de l'utilisation d'un gène chloroplastique codant pour une petite protéine, ainsi que d'une nouvelle méthode de reconstitution d'arbre évolutif, sont discutés. Des résultats concernant les plantes sont présentés, à différents niveaux taxonomiques.

SHORT BIBLIOGRAPHY

Bittar G. (1995). "Phylogénie: histoire, concepts, principes et présentation de nouvelles méthodes d'analyse numérique de l'évolution moléculaire". PhD in Sciences - Doctorate Dissertation n°2789, Université de Genève (Switzerland), 480 pp.

Bittar G. & Carter L. (1994). "New probabilistic and numerical phenetics methods for inferring natural groupings and phylogenesis". *In:* Proceedings of the International Meeting "Ecology and statistical methods", Niort (France), pp. 145-150.

Nadot S. (1994). "Utilisation de l'outil moléculaire pour la reconstruction d'une phylogénie des plantes monocotylédones en général et de la famille des Poaceae (Graminées) en particulier". PhD in Sciences - Doctorate Dissertation, Université de Paris XI, Orsay (France), 165 pp.

Nadot S., Bajon R. & Lejeune B. (1994). "The chloroplast gene *rps*4 as a tool for the study of Poaceae phylogeny". *Plant Systematics and Evolution* 191: 27-38.

Nadot S., Bittar G., Carter L., Lacroix R. & Lejeune B. (1995). "A phylogenetic analysis of monocotyledons based on the chloroplast gene *rps*4, using parsimony and a new numerical phenetics method". *Molecular Phylogenetics and Evolution* 4 (3): 257-282.

Souza-Chies T., Bittar G., Nadot S., Carter L., Besin E. & Lejeune B. (1996). "Phylogenetic analysis of Iridaceae with parsimony and distances methods, using the plastid gene *rps*4". *Plant Systematics and Evolution*. (in print).