J Mol Evol (2002) 54:715–724
DOI: 10.1007/s00239-001-0042-9

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 2002

The Evolution of Gymnosperms Redrawn by Phytochrome Genes: The Gnetatae Appear at the Base of the Gymnosperms

Marion Schmidt, Hansjörg A.W. Schneider-Poetsch

Botanisches Institut der Universität zu Köln, D-50931 Köln, Germany

Received: 18 June 2001 / Accepted: 2 October 2001

Abstract. Gymnosperms possess two to four phytochrome types which apparently are the result of successive gene duplications in the genomes of their common ancestors. Phytochromes are nuclear-encoded proteins whose genes, contrary to chloroplast, mitochondrion, and rRNA genes, have hitherto rarely been used to examine gymnosperm phylogenies. Since the individual phytochrome gene types implied phylogenies that were not completely congruent to one another, conflicting branching orders were sorted by the number of gene lineages present in a taxon. The Gnetatae (two gene types) branched at the base of all gymnosperms, a position supported by bootstrap sampling (distance and character state trees, maximum likelihood). The Gnetatae were followed by *Ginkgo*, Cycadatae, and Pinaceae (three gene types) and the remaining conifers (four gene types). Therefore, in phytochrome trees, the most ancient branch of the conifers (Pinatae) seems to be the Pinaceae. The next split appears to have separated Araucariaceae plus Podocarpaceae from the Taxaceae/Taxodiaceae/Cupressaceae group. Structural arrangements in the plastid genomes (Raubeson and Jansen 1992) corroborate the finding that there is no close connection between Pinaceae and Gnetatae as suggested by some publications. The analyses are based on 60 phytochrome genes (579 positions in an alignment of PCR fragments) from 28 species. According to rough divergence time estimates, the last common ancestor of gymnosperms and angiosperms is likely to have existed in the Carboniferous.

Correspondence to: Hansjörg Schneider-Poetsch; *email*: hansjoerg.schneider-poetsch@uni-koeln.de

Key words: Gymnosperm evolution — Gene lineage bias — Gnetatae — Phytochrome gene lineages — Third codon position

Introduction

Molecular data rapidly invade and increasingly dominate phylogenetic discourses. Reviews on the origin of and the relationships within seed plants of recent years, not even spanning a decennium, clearly reflect this development (see, e.g., Doyle 1994, 1998; Crane et al. 1995; Crepet 1998; Qiu et al. 1999; Donoghue and Doyle 2000). Not least, the development of computer-suitable algorithms handling large phylogenetic data sets have added to the power of today's phylogenetic studies.

Several molecular analyses of gymnosperm evolution have been published in recent years (see below). Discrepancies between the outcomes of these analyses and conclusions based on morphological and developmental characters quickly become evident. On the molecular side there is an increasing consensus among phylogenies, and on the classical side, as summarized in textbooks and monographies, there is an almost-complete lack of means to weight gymnosperm classes, orders, and even families by phylogenetic relationships that have a certain degree of probability beyond the fossil record (for examples see Strickberger 1996; Sitte et al. 1999; Kramer and Green 1990; Stewart and Rothwell 1993). According to the fossil record, various types of gymnosperm ancestors seem to have existed prior to the Carboniferous, so that in the Carboniferous the gymnosperms appear as a polyphyletic group.

In previous work on phytochrome gene function and evolution (Schneider-Poetsch et al. 1998; Clapham et al. 1999; Basu et al. 2000), we found that these genes could contribute to understanding gymnosperm phylogeny. There are up to four types of phytochrome genes in gymnosperms, N (which split into two sublineages), O, and P, which are, like phytochromes A through E in angiosperms, the result of gene splits before extant species emerged (Basu et al. 2000; this paper). We have now followed the evolution of these genes in gymnosperm species from different classes, orders, and families to address questions not yet settled in gymnosperm phylogeny.

Constructing phylogenetic trees with each of the phytochrome genes found, we expected trees that, if identical, would verify an inferred phylogeny and, if not, would warrant caution. Since phytochrome genes are nuclear-encoded protein genes, they were also assumed to compare or contrast excellently with chloroplast, mitochondrial, and rRNA genes, which were hitherto used in establishing gymnosperm phylogenies. Gene sequences may be subject to different evolutionary constraints, which may result in somewhat differing phylogenies. Reasons for biased phylogenies given by Qiu et al. (1999) are evolutionary rate heterogeneity among descendants, insufficient sampling of genes and taxa, and GC content bias. Another reason may be divergence points far in the past implicating saturation. A comparative study extensively analyzing the influence of rate heterogeneity of sites and partitions (including analysis of third codon positions) of two plastid genes (*psaA* and *psbB*) on gymnosperm phylogenies was done by Sanderson et al. (2000), although they drew no conclusion as to how the true tree may be extracted thereafter. That even the evolutionary rate in corresponding fragments of related genes within a species may differ considerably is shown in the present study. One may think of mutations altering the number of tolerable mutation in other parts of the gene. Here we show that, as a result of successive gene splits in ancestors, different numbers of genes in individual taxa may provide means to overcome some of the difficulties caused by differently biased genes.

Of particular importance is the issue surrounding the position of the Gnetatae. Morphological and developmental patterns are such that they will never exclude angiosperm affinities. Informative recapitulations of the chronology of the arguments for and against Gnetatae–angiosperm relationships preferred by certain periods have been given by Doyle (1994) and Donoghue and Doyle (2000). Winter et al. (1999) and Hansen et al. (1999) suggested that morphological characters that are similar in Gnetatae and angiosperms and classified as homologous might be analogous. When eventually Gnetatae were more likely seen as less related to angiosperms based on molecular data (Goremykin et al. 1996; Chaw et al. 1997, 2000; Samigullin et al. 1999; Winter et al. 1999; Bowe et al. 2000; Sanderson et al. 2000), the question as to their next relatives and forefathers remained elusive, though recent studies imply a relationship between Gnetatae and Pinaceae. The most recent analyses (Sanderson et al. 2000) suggest two interpretations. First and second codon positions (of concatenated *psaA* and *psbB* genes) give rise to trees with the Gnetatae in a clade with *Pinus*, and third codon positions imply trees in which the Gnetatae are the sister group of all other seed plants. But the attempt to redraw gymnosperm phylogeny encounters even more questions. Regarding interrelationships between conifers, e.g., Page (1990) writes, "Evidence for interrelationships, if any, can be summarized as both speculative and extremely tenuous . . ." The same book, *Pteridophytes and Gymnosperms*, (Kramer and Green 1990) and *Strasburger* (Sitte et al. 1999) unite Cycadatae and Gnetatae in the subdivision Cycadophytina and Ginkgoatae and Pinatae (conifers) in the subdivision Coniferophytina, an order scarcely corroborated by molecular analyses. Thus, it appears that additional molecular data will be an important tool for studying gymnosperm relationships further.

Here we report results on phytochrome evolution largely agreeing with gymnosperm phylogenies based on molecular data sets very different from ours. They corroborate notions that the Gnetatae are more closely related to the rest of the gymnosperms than formerly expected and that Pinaceae are a sister group to the rest of the conifers. Bootstrap values, gene numbers, and gene affinities to angiosperms, however, support the view that the Gnetatae emerge at the base of the gymnosperms.

Materials and Methods

Plant Material

Plant material was kindly provided by the Botanical Gardens of the University of Bonn (Dr. Lobin) and the Botany Institute of the University of Cologne (Mr. Zimmer). Leaves or seedlings were harvested, shock-frozen in liquid nitrogen, and stored at −20°C. Table 1 lists the 27 species examined, phytochrome types, and GenBank accession numbers of 55 novel phytochrome sequences of gymnosperms which were phylogenetically analyzed together with 6 already known gymnosperm phytochromes.

Isolation of Genomic DNA, PCR, Cloning, and Sequencing

DNA was purified by a CTAB method (Doyle and Doyle 1990) from young leaves or whole seedlings, and phytochrome sequences were amplified by polymerase chain reaction (PCR) with degenerated primers (see Clapham et al. 1999) binding to two highly conserved sites flanking the chromophore binding region.

PCR reactions were routinely run with 200 ng of DNA, 50 pmol of each primer, 1 U of Taq-DNA-Polymerase (Promega, Heidelberg, Germany), and buffer as supplied and recommended by the manufacturer. Phytochrome sequences were amplified during 35 cycles, comprising hot start, denaturation (30 s, 94°C), annealing (30 s, 42–52°C, indi-

Table 1. Species examined, phytochromes sequenced, and GenBank accession numbers[a]

Taxon[b]	Type of PHY	GenBank accession no.
Ginkgoatae		
Ginkgo biloba (1)	*PHYN*	AJ286637
	PHYO	AJ286638
Pinatae		
Agathis dammara (2)	$PHYN_2$	AJ286620
	PHYO	AJ286619
	PHYP	AJ286621
Araucaria araucana (1)	$PHYN_2$	AJ286617
	PHYO	AJ286616
	PHYP	AJ286618
Cephalotaxus fortunei (2)	$PHYN_1$	AJ286629
	$PHYN_2$	AJ286630
Chameacyparis lawsoniana (1)	*PHYO*	AJ286632
	PHYP	AJ286631
Cryptomeria japonica (1)	$PHYN_1$	AJ286624
	PHYO	AJ286622
	PHYP	AJ286625
Cupressus sempervirens (1)	$PHYN_1$	AJ286626
	PHYO	AJ286627
	PHYP	AJ286625
Dacrydium franklinii (2)	$PHYN_2$	AJ286634
	PHYO	AJ286635
	PHYP	AJ286633
Juniperus phoenicea (1)	*PHYP*	AJ286639
Larix decidua (1)	*PHYN*	AJ286641
	PHYO	AJ286640
Phyllocladus trichomanoides (2)	$PHYN_1$	AJ286650
Pinus sylvestris (1)	*PHYN*	AJ286647
	PHYO	AJ286645
Podocarpus neriifolius (1)	$PHYN_1$	AJ286652
Pseudotsuga menziesii (1)	*PHYP*	AJ286651
Saxegothea conspicus (2)	$PHYN_1$	AJ286654
	$PHYN_2$	AJ286655
Sciadopitys verticillata (1)	$PHYN_1$	AJ286657
	PHYP	AJ286656
Sequoiadendron giganteum (1)	$PHYN_1$	AJ286653
Taxodium distichum (1)	$PHYN_1$	AJ286659
	PHYP	AJ286660
Taxus baccata (1)	*PHYP*	Y13796
Thuja plicata (1)	$PHYN_1$	AJ286662
	$PHYN_2$	AJ286663
	PHYO	AJ286661
Torreya nucifera (1)	$PHYN_1$	AJ286664
	$PHYN_2$	AJ286665
	PHYP	AJ286666
Cycadatae		
Cycas revoluta (1)	*PHYO*	Y07571
	PHYP	AJ286628
Macrozamia communis (1)	*PHYN*	AJ286642
	PHYO	AJ286644
	PHYP	AJ286643
Stangeria eriopus (1)	*PHYP*	AJ286658
Gnetatae		
Ephedra foeminea (1)	*PHYN/O*	AJ310934
	PHYP	AJ310935
Gnetum gnemon (1)	*PHYN/O*	AJ286636
	PHYP	X80295
Welwitschia mirabilis (1)	*PHYN/O*	AJ286667
	PHYP	Y13794

[a] Additional sequences used were *Arabidopsis thaliana PHYA* (X17341), *PHYB* (X17342), and *PHYC* (X17343); *Sorghum bicolor PHYA* (U56729), *PHYB* (U56730), and *PHYC* (U56731); *Ginkgo biloba PHYP* (X98698); *Metasequoia glyptostroboides PHYP* (X80297); *Picea abies PHYO* (U60264) and *PHYP* (X80298); *Pinus sylvestris PHYP* (X96738); *Pseudotsuga menziesii PHYN* (U22458); and *Marchantia polymorpha* phy 1 (X80296).

[b] Number in parentheses indicates the location of growth: (1) garden of the Botany Institute, University of Cologne (Germany); (2) Botanical Garden, Friedrich-Wilhelms-Universität Bonn (Germany).

vidually adjusted to the DNA applied), elongation (60 s, 72°C), and a final elongation step (10 min, 72°C).

PCR products were analyzed on agarose gels and amplificates of the expected length (about 600 nucleotides) were eluted from the gel with the QIAEX II Gel Extraction System (Qiagen, Hilden, Germany), ligated with pGEM-T easy (Promega, Mannheim, Germany), and multiplied in *E. coli* XL1-Blue (Promega). Different phytochrome types of a species were recognized by restriction fragment length polymorphism (RFLP) analyses using *Bam*HI, *Eco*RI, *Eco*RV, *Hae*II, and *Hin*dIII.

Sequencing reactions were run as recommended for the Genetic Analyzer ABI Prism 310 from Perkin Elmer (Weiterstadt, Germany) using standard primers and the ABI Prism Dye Terminator Cycle Sequencing Ready Reaction KIT SF (Perkin Elmer, Weiterstadt, Germany). Both strands of a sequence were analyzed and examined by sequencing a second clone.

Phylogenetic Analyses

Sequence data were processed by the Wisconsin Package Version 10.0, Genetics Computer Group (GCG) (Madison, WI), provided by the Regionale Rechenzentrum of the University of Cologne. The alignments were created by the program Pileup of the GCG package. Gaps were edited by hand according to the best fit of the respective amino acid sequences. The actual alignment may be requested from the authors. It comprises 579 positions.

Distance matrices were calculated with the two-parameter method of Kimura (1980) provided by PAUP* 4.03β (Swofford 1999). Other types of distance matrices (Jukes–Cantor, maximum likelihood) had little influence on the trees' branching orders. Phylogenetic trees were then constructed using the neighbor-joining method (Saitou and Nei 1987).

PAUP* 4.03β (Swofford 1999) was used for parsimony analyses. Because of the large data set, exhaustive branch-and-bound algorithms were not practicable. We used heuristic searches with the MULTREES option and the ACCTRAN optimization. The analyses included 100 replicates with stepwise random taxon addition and TBR branch-swapping.

To assess the strength of individual clades of distance as well as of parsimony trees, bootstrap analyses of 1000 replicates (Felsenstein 1985) were calculated. Further analyses used the maximum likelihood method corresponding to the HKY85 model using a heuristic search strategy (MULTREES on, TBR branch-swapping, one tree held at each step).

Results

The Evolution of Phytochrome Gene Lineages

Figure 1 shows the phylogenetic relationships of phytochrome genes. It is a simplified and synoptic tree uniting full-length amino acid sequences (about 1300 positions) of phytochromes from *Arabidopsis, Sorghum*, the gymnosperms *Pinus* and *Picea*, and lower plants (see Clapham et al. 1999). Its root is alga phytochrome. Phytochromes A, B, C, and E are found in angiosperms. (*PHYD* was omitted because it is a kind of *PHYB*.) These phytochrome types are the result of several gene splits, the first of which occurred very early in the evolution of land plants. *PHYE* is only present in dicotyledons. The gymnosperm phytochromes O and P are the descendants of the two most ancient phytochrome lineages, and as will be shown in an extended gymnosperm tree, there are additional phytochromes N_1 and N_2. Phytochromes of

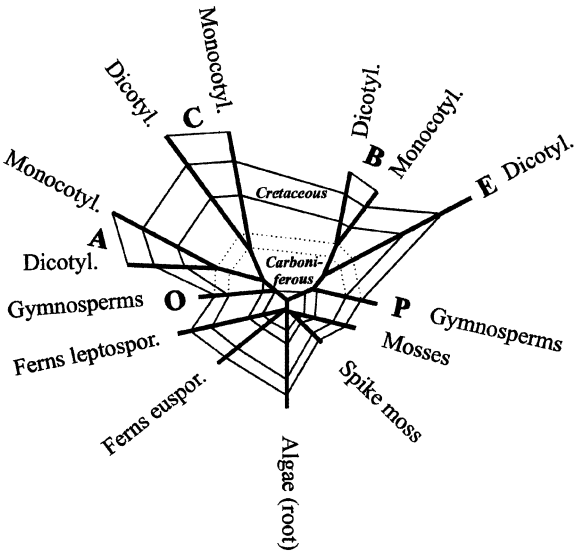

Fig. 1. Synoptic and simplified phylogenetic distance tree showing the evolution of phytochrome genes. A, B, C, and E refer to angiosperm, and O and P to gymnosperm, phytochrome gene lineages. Ferns may harbor more than the gene lineage depicted (see Schmidt 2000). The tree is based on full-length amino acid sequences of *Arabidopsis* A (X17341), B (X17342), C (X17343), and E (X176610); *Avena* A (M18822); *Oryza* B (X57563); *Sorghum* C (U56731); *Picea* O (U60264); *Pinus* P (X96738); *Psilotum* (X74930); *Adiantum* (D13519); *Physcomitrella* (X75025); *Selaginella* (X61458); and *Mougeotia* (X95550) (see Clapham et al. 1999). *Psilotum* clusters with eusporangiate ferns. Geological periods [Cretaceous (141–65 Myr b.p.) and Carboniferous (345–280 Myr b.p.)] are drawn assuming constant evolutionary rates along an individual branch. The split of mosses and the first vascular plants (spike moss) marks the origin of land plant evolution (450 Myr. b.p.). *Dotted lines* indicate the position of the Cretaceous assuming that up to this period the evolutionary rates in angiosperm and gymnosperm phytochrome genes were roughly similar.

mosses and first tracheophytes (*Selaginella*) branch off first, almost immediately followed by phytochromes of eu- and leptosporangiate ferns. Assuming rate constancy [which certainly does not hold in angiosperms (see Schneider-Poetsch et al. 1998)] within an individual phytochrome lineage, and calibrating the tree by the separating point of mosses and first tracheophytes [450 million years before present (Myr b.p.) (Shear 1991; Stewart and Rothwell 1993; Cai et al. 1996)], we have indicated two geological periods significant in land plant evolution, Carboniferous and Creteaceous. Since rate constancy apparently roughly holds within lower plants and gymnosperms, the separation of gymnosperm and angiosperm phytochromes can be located in the Carboniferous. The divergence time of mono- and dicotyledons has been discussed elsewhere (see Goremykin et al. 1997).

Phytochrome Lineages and the Evolution of Gymnosperms

Having clarified the relationships between angiosperm and gymnosperm phytochrome lineages, we began to screen a large variety of gymnosperm species for phytochromes. Finding two more gene lineages, in addition to PHYO and PHYP, surprised us. *PHYO* and the novel

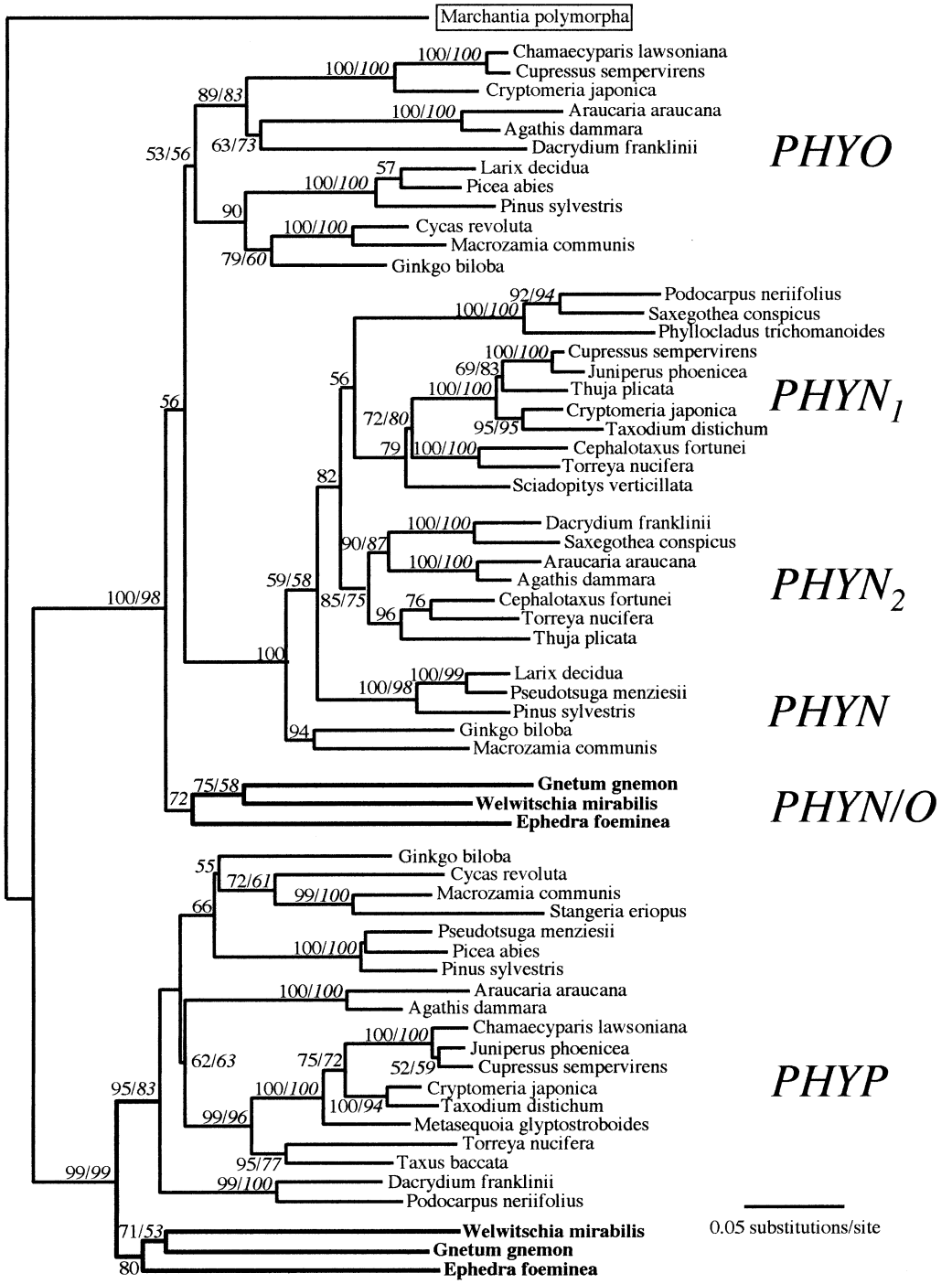

Fig. 2. Distance tree (Kimura matrix, neighbor-joining) uniting 60 phytochrome genes (fragments of about 600 nucleotide positions) of 28 gymnosperm species. The phytochrome gene lineages N_1, N_2, O, and P are indicated. N_1 and N_2 are offspring of an N lineage emerging from an ancestor N/O lineage. Bootstrap values (percentage of 1000 replicates) are given at the nodes. *Italics* indicate respective values from a maximum parsimony tree of almost-identical branching order. Values below 50 have been omitted. The bootstrap values at the node joining *PHYN*s and *PHYO*s and at the base of *PHYO*s increase considerably (96 and 100) if these phytochromes are analyzed separately with the (N/O)–Gnetatae.

gene lineage *PHYN* appear to emerge from a common lineage (N/O) before *PHYN* gave rise to the gene lineages N_1 and N_2.

These relationships are summarized in the distance tree in Fig. 2. It is based on an alignment comprising 579 positions of 60 phytochrome sequences from 28 gymnosperm species. Any lower plant outgroup other than *Marchantia* gave rise to a very similar tree. Thus the conclusions drawn are independent of the outgroup chosen. Bootstrap values, however, often improve if the different gene lineages are analyzed individually.

Our hopes that each of these phytochrome lineages

would give rise to phylogenetic trees with coinciding branching orders were met only in part. Distortions, to all appearance caused by discontinuous mutation rates within the individual phytochrome gene lineages and individual clusters, blur phylogenetic relationships in some regions. The split of Araucariaceae and Podocarpaceae, e.g., occurs much farther from the base of the tree within the N_2 lineage than within the O lineage; and within the P lineage the relationships between these two orders are blurred. In addition, there are considerable differences as to the bootstrap values at forks linking phytochromes of the same taxa in different gene lineages (for example, see the nodes linking *Ginkgo* and Cycadatae). Further uncertainties may be introduced by the possibility that in some of the analyzed taxa the one or the other phytochrome gene lineage evaded analysis because the PCR-based search is not a warranty for the complete possible set of phytochromes. An example of an unsuccessful search is the lack of Araucariaceae phytochromes in N_1, although three species of the Podocarpaceae are present.[1]

However, if a gene lineage is uniformly found to be lacking in related taxa, we get a measure to assess the succession of taxa by gene splits (or losses). So the Gnetatae were found to be lacking two, and *Ginkgo*, Cycadatae, and Pinaceae one, of four phytochrome lineages. Considering that in many taxa four phytochrome types were easily detectable, the probability that our primers specifically failed to detect certain phytochrome types in these other taxa is low. Every tested member was devoid of them and the search for more gene lineages in Gnetatae, e.g., was extensive.

Consequently, ordering gymnosperm taxa according to the number of gene lineages found in them will help to overcome problems caused by inconsistencies within gene lineages and give us the information that the implied phylogenetic succession is Gnetatae (two lineages), Cycadatae, *Ginkgo*, Pinaceae (three lineages), the rest of the conifers (four lineages). The lacking fourth gene in Araucariaceae could order them also to the three gene taxa, but neither branching order is in favor of this interpretation (see Footnote 1). The bootstrap values backing the position of the Gnetatae at the base of the gymnosperms are considerably high. The poor separation of *PHYO*s and *PHYN*s from the gnetalean phytochromes on the N/O branch is due to the fact that both lineages are apparently equally related to the gnetalean phytochromes. High bootstrap values (98 and 100) appear at the base of the O and N lineage if the two lineages are analyzed separately.

As for the conifers, few doubts remain about the relationships and the succession of taxa such as *Taxus, Torreya, Cephalotaxus*, Taxodiaceae, and Cupressaceae. They form a single group in each of the phytochrome gene lineages. It appears that the Taxaceae *sensu lato* (including *Torreya* and *Cephalotaxus*) were separated before the split of Cupressaceae and Taxodiaceae occurred. The single-species family *Sciadopitys* (Two gene lineages only? only one is shown) may be found prior to the Taxaceae.

Phytochromes of Pinaceae and *Ginkgo* plus Cycadatae cluster together in the lineages O and P. Bootstrap values are high for O (90) and low for P (66). In N the Pinaceae are separated from *Ginkgo* and the Cycadatae and appear as the sister group of the conifers, showing that phytochrome genes are differently biased in different lineages. However, the position of the Pinaceae on the N lineage probably reflects true relationships. Since all conifers are characterized by the lack of a common inverted repeat in the chloroplast genome, and the Pinaceae share this character state (Raubeson and Jansen 1992), the Pinaceae belong to the conifers. Their proximity to *Ginkgo* and the Cycadatae by the number of phytochrome genes (three) and their obvious gene similarity imply that the Pinaceae were separated very early from the lineage giving rise to the conifers with four phytochrome lineages. It is almost unnecessary to mention that the conifers are also separated from *Ginkgo* and the Cycadatae by their mode of fertilization (nonmotile sperm in conifers and motile sperm in *Ginkgo* and Cycadatae, although convergence in the latter two cannot be excluded).

Accepting this succession of events the next split creating new taxa appears to have separated the Araucariaceae plus Podocarpaceae lineages from the lineage leading to the Taxaceae, Taxodiaceae, and Cupressaceae. Phytochrome genes of the N lineages and O back such a branching order which, however, is blurred in the lineage P, where the Podocarpaceae form a sister group to all gymnosperms except the Gnetatae.

Taking the available information together, the data indicate a tree where *Ginkgo* and Cycadatae form a sister branch to a branch that is split in the Pinaceae and the Taxaceae/Cupressaceae/Taxodiaceae branch. Araucariaceae and Podocarpaceae emerge together at the base of the latter branch but are separated soon thereafter. The lowest branch is formed by the Gnetatae. Such a tree is tentatively drawn in Fig. 4.

What Concatenated Phytochrome Sequences Say

Trees concatenating N, O, and P (1737 characters) genes of gymnosperm species and A, C, and B genes of angiosperm species (Fig. 3A) support the branching order just outlined. There is no difference between these trees that would contradict the conclusions drawn. In each tree the Gnetatae appear with a high bootstrap significance between angiosperms and gymnosperms. The same holds for more extended trees where genes not available from

[1]In addition to the *N_2*, *O,* and *P* genes (GenBank accession Nos. AJ420750 through AJ420752), recent analyses have detected the *N_1* gene (AJ420749) in *Araucaria heterophylla.*

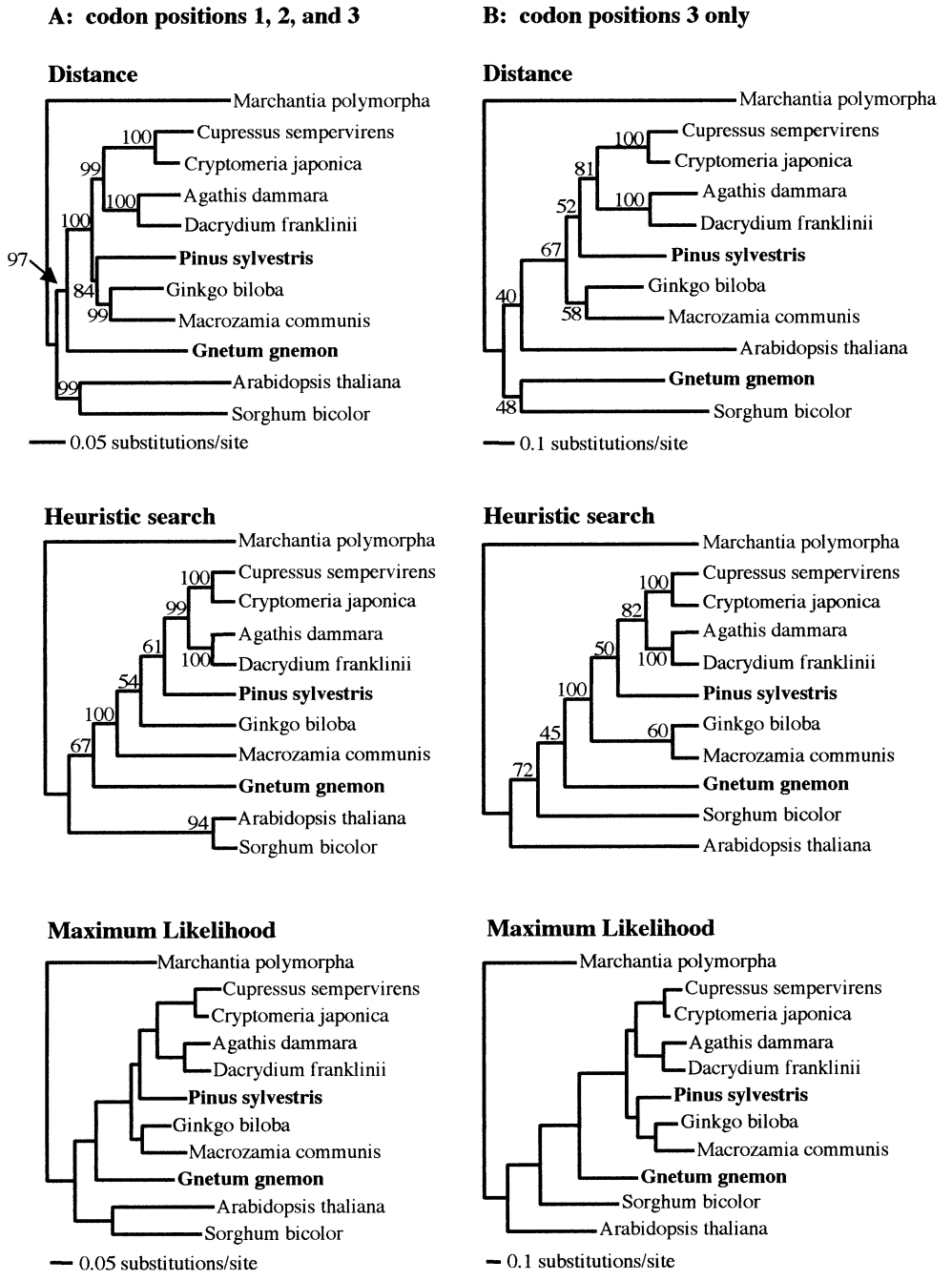

Fig. 3. **A** Phylogenetic trees based on concatenated sequences (1737 characters) including angiosperm phytochrome sequences from *Arabidopsis* and *Sorghum*. The concatenating order was N, O, and P for gymnosperms and A, C, and B for angiosperms. Results did not change much if the positions of C and A were mutually substituted (it cannot be decided which of these genes is the orthologue of N and O). For *Gnetum*, having only one N/O gene, this gene was taken twice; the phytochrome gene of *Marchantia* (outgroup) was taken three times. Bootstrap values (percentage of 1000 replicates) are given at the nodes. Parsimony heuristic search (stepwise random taxon addition) gave rise to three trees, each of 2455 steps. The one with *Gnetum* as a sister group of the gymnosperms is shown. The second one showed *Gnetum* as a sister group of the angiosperms, and in the third tree *Gnetum* was the sister group of all seed plants. **B** Trees examining the influence of codon positions using the same sequence alignment as in A, but omitting codon positions 1 and 2 (579 characters of codon position 3).

individual species were substituted by genes from the same family (not shown), But parsimony trees show (legend to Fig. 3) that the position of *Gnetum* may change. On trees of the same length, *Gnetum* also appeared as the sister group of all seed plants, or the angiosperms.

According to the theory, third codon positions are completely saturated after about 100 million years, but Sanderson et al. (2000) have recently shown that, despite that, third codon positions were not without information. So we tested the influence of this positions. The resulting trees were not drastically different from those using positions 1 through 3 (Fig. 3B). However, in distance trees

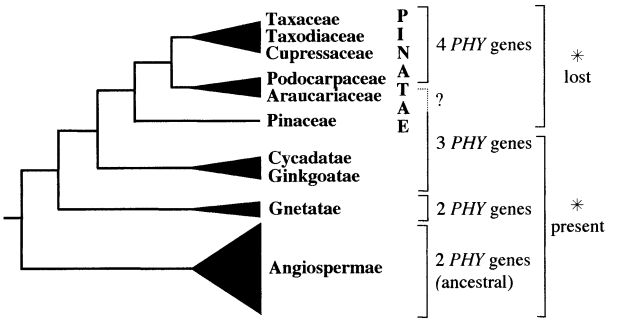

Fig. 4. Synoptic tree tentatively summarizing the main results of the present approach. The *asterisks* indicate a cpDNA inverted repeat which is lost or present (see Raubeson and Jansen 1992). No time scale is given (see Footnote 1 on p. 720).

Gnetum came close to the angiosperms. The average statistical significance is moderate, but as *Gnetum* is never close to *Pinus* and most molecular data are not in favor of the Gnetatae being connected with angiosperms (see above quotations and Discussion), even these trees are not adverse to the previous conclusions.

Distance trees based on codon positions 1 and 2 largely agree with third-position trees, but *Pinus* is in a clade with *Ginkgo* and *Macrozamia.* In the first- and second-position maximum likelihood tree, *Gnetum* joins *Pinus*, however, together with the angiosperms, and the shortest heuristic search first- and second-position tree obtained (893 steps) resembles the third-position tree except that *Gnetum* clusters with the angiosperms. Thus, it appears that relevant differences refer mainly to angiosperm–*Gnetum* affinities.

If there were no good molecular evidence for gymnosperm phylogenies differing in some detail from those just suggested (see Discussion), one could assume that the phytochrome gene-based trees have largely answered the pending questions on gymnosperm phylogeny.

Discussion

The phylogenetic tree in Fig. 4 is a summary of the essential outcomes of the present analyses. Conflicting branching orders resulting from trees uniting 60 phytochrome genes from 28 gymnosperm species (Fig. 2) were sorted by the number of genes found in individual taxa. Trees based on concatenated phytochrome genes show very similar branching orders (Fig. 3A). None of them is really contradictory to our arguments. The topology of these trees is strikingly compatible with the most parsimonious trees based on partial 28 S rRNA sequences published by Stefanovic et al. (1998). Using *Ginkgo* and cycads as the outgroup, (*Sciadopitys*), Taxaceae (including Cephalotaxaceae), Taxodiaceae, and Cupressaceae form a monophyletic group with Araucariaceae and Podocarpaceae as their sister group. The Pinaceae appear as the sister group of all these conifers together. Chaw et al. (1997) obtained a similar topology comparing 18 S rRNA genes in neighbor-joining and most parsimonious analyses.

The latter authors (Chaw et al. 1997) also address the position of the Gnetatae. The bootstrap value (75) that joins the Gnetatae with the gymnosperms is not assumed to be conclusive, but the Gnetatae sequences analyzed are clearly separated from those of angiosperms by specific indels (insertion and deletions), so it appears unlikely that the Gnetatae are a sister group of the angiosperms. That the Gnetatae are more closely related to gymnosperms than formerly assumed is also inferred by the results of other authors. Goremykin et al. (1996), analyzing chloroplastic 4.5S rRNA flanked by two adjacent intergenic transcribed spacer regions and the *rbc*L gene, found no support for a sister-group relationship between the Gnetatae and angiosperms, and MADS-box genes of *Gnetum* are always found clustering with gymnosperm and not with angiosperm MADS-box genes (Winter et al. 1999).

So far, the results obtained by different molecular data sets are largely in agreement. However, there are differences as to the position of the Gnetatae within the gymnosperms. The finding that there were probably two phytochrome gene duplications in the course of gymnosperm evolution implies that the Gnetatae, with only two genes, form the lowest branch and herewith a sister group to the rest of the gymnosperms. In our tree (Fig. 2), both primeval phytochrome lineages show the Gnetatae in this position. Using lower plant phytochromes as the outgroup, the basic position of the Gnetatae is supported on the *PHYN/O* as well as on the *PHYP* branch (both neighbor-joining and parsimonious trees) by bootstrap values of 99/98 and 98/98. If *PHYN*s and *PHYO* are analyzed separately, high bootstrap values separate also the non-Gnetatae gymnosperms from the Gnetatae.

Trees concatenating *PHYN*$_1$/*N*$_2$, *PHYO*, and *PHYP* sequences and *PHYA, PHYC*, and *PHYB* sequences from angiosperms (Fig. 3A; 1737 characters) show the Gnetatae in the same position. There are other phylogenetic studies which imply a similar order (Goremykin et al. 1996; Samigullin et al. 1999), and a study on seed proteins (Fischer et al. 1996), although including only the gymnosperms *Pseudotsuga, Ginkgo*, and *Welwitschia*, orders these species primarily in the same way that phytochrome genes do. But the majority of published trees shows the Gnetatae as a sister group of *Pinus* or the Pinaceae (Malek et al. 1996; Goremykin et al. 1996; Samigullin et al. 1999; Qiu et al. 1999; Bowe et al. 2000; Chaw et al. 2000; Sanderson et al. 2000). Sometimes, as a consequence of poorly resolved branching points and the tree construction methods applied, there are also trees showing the Gnetatae positioned differently. But conflicting branching orders may also come with a high bootstrap significance as Sanderson et al. (2000), examining the influence of individual codon positions of the

plastid *psaA* and *psbB* genes, report. Trees in which the Gnetatae appear as a sister group of all conifers (Pinatae) (Chaw et al. 1997, 2000) or of all seed plants (Chaw et al. 2000; Sanderson et al. 2000) have been published, and a sister-group relationship with angiosperms was even found (Stefanovic et al. 1998). Thus, it appears that, with respect to the Gnetatae, most topologies based on genes other than phytochrome genes are different from the present ones and prefer the *Pinus*–(Pinaceae)–Gnetatae connection. It should, however, be added that all Pinatae are separated from the Gnetatae and other seed plants by the loss of an inverted repeat in chloroplast DNA (Raubeson and Jansen 1992). That a structural character lost in a common ancestor was regained in successor genomes is highly improbable.

Although the sorting of taxa according to phytochrome gene numbers is persuasive, we do not have means to exclude the possibility that the Gnetatae have lost genes. As monocotyledons show, phytochrome gene losses may happen (Mathews and Sharrock 1996). Consequently, if there were a gene loss in the Gnetatae, we would have lost an argument corroborating the significance of the bootstrap values supporting the position of the Gnetatae at the base of the gymnosperms. But we should keep in mind that there are four phytochrome lineages which show the gnetalean phytochrome genes appearing at their base. Further support for the basic position of the Gnetatae may also come from the fact that affinities to angiosperms cannot wholly be denied (see above).

The present analyses speak in favor of an almost-trifurcation of angiosperms, Gnetatae, and non-Gnetatae gymnosperms. Efforts to order these events conclusively will remain difficult because of their proximity to each other in evolutionary ancient times.

Nonetheless, a rough time estimate of the events giving rise to angiosperms, Gnetatae, and gymnosperms may be given. Using molecular clocks constructed from the chloroplastic gene coding for the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbc*L) and the nuclear gene coding for the small subunit of rRNA (*Rrn*18), Savard et al. (1994) conclude that the last common ancestor of the five extant taxa Cycadatae, *Ginkgo*, Gnetatae, conifers, and angiosperms lived 275 to 290 Myr b.p. Another estimate (Goremykin et al. 1997), based on a concatenated 14,295-amino acid alignment of chloroplast genes, is 350 ± 35 million years. The simple estimate we got using a distance tree based on full-length phytochrome genes is well within these limits. Assuming approximately constant evolutionary rates on individual branches and calibrating the tree by the divergence point of mosses (*Physcomitrella*) and the first vascular plants (*Selaginella*) (450 Myr b.p.) localizes the divergence point of both the O and the P phytochrome lineages in the Carboniferous (345 to 280 Myr b.p.). [A distortion by the more rapidly evolving angiosperm phytochrome genes is unlikely because this acceleration appears to have taken place after the separation of gymnosperms and angiosperms (see Fig. 1).] Thus, according to the succession inferred by phytochrome genes, the divergence period for Gnetatae, Ginkgoatae, and Cycadatae could well be in the Permian (280 to 225 Myr b.p.) and that of the Pinatae in the Lower Triassic (Triassic 225 to 195 Myr b.p.) as Savard et al. (1994) have suggested. There is hope that molecular relationships will gradually make it easier to integrate data from the fossil and morphological record into phylogenies.

Data on nuclear protein genes of gymnosperms are presently sparse. However, they would be needed to examine the outcome of this study, specifically regarding the position of the Gnetatae within the gymnosperms, even though similarities to phytochrome gene trees would still not permit deciding which type of genes reveals the true tree.

References

Basu D, Dehesh K, Schneider-Poetsch HAW, Harrington SE, McCouch SR, Quail PH (2000) Rice *PHYC* gene: Structure, expression, map position and evolution. Plant Mol Biol 44:27–42

Bowe ML, Coat G, de Pamphilis CW (2000) Phylogeny of seed plants based on all three genomic compartments: Extant gymnosperms are monophyletic and Gnetales closest relatives to conifers. Proc Natl Acad Sci USA 97:4092–4097

Cai C, Ouyang S, Wang Y, Fang Z, Rong J, Geng L, Li X (1996) An early Silurian vascular plant. Nature 379:592

Chaw S-M, Zharkikh A, Sung H-M, Lau T-C, Li W-H (1997) Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol Biol Evol 14:56–68

Chaw S-M, Parkinson C-L, Cheng Y, Vincent TM, Palmer JD (2000) Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of gnetales from conifers. Proc Natl Acad Sci USA 97:4086–4091

Clapham DH, Kolukisaoglu HÜ, Larsson C-L, Qamaruddin M, Ekberg I, Wiegmann-Eirund C, Schneider-Poetsch HAW, v. Arnold S (1999) Phytochrome types in *Picea* and *Pinus*. Expression patterns of *PHYA*-related types. Plant Mol Biol 40:669–678

Crane PR, Friis EM, Pedersen KR (1995) The origin and early diversification of angiosperms. Nature 374:27–33

Crepet WL (1998) The abominable mystery. Science 282:1653–1654

Donoghue MJ, Doyle AD (2000) Seed plant phylogeny: Demise of the anthophyte hypothesis? Curr Biol 10:R106–R109

Doyle JA (1994) Origin of the angiosperm flower: A phylogenetic perspective. Plant Syst Evol 8:7–29

Doyle JA (1996) Seed plant phylogeny and the relationships of the gnetales. Int J Plant Sci 157:S3–S39

Doyle JA (1998) Molecules, morphology, fossils, and the relationship of angiosperms and gnetales. Mol Phylogenet Evol 9:448–462

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

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

Fischer H, Chen L, Wallisch S (1996) The evolution of angiosperm seed proteins: A methionine rich legumin subfamily present in lower angiosperm clades. J Mol Evol 43:399–404

Goremykin V, Bobrova V, Pahnke J, Troitsky A, Antonov A, Martin W (1996) Noncoding sequences of the slowly evolving chloroplast

inverted repeat in addition to *rbc*L data do not support gnetalean affinities of angiosperms. Mol Biol Evol 13:383–396

Goremykin VV, Hansmann S, Martin WF (1997) Evolutionary analysis of 59 proteins encoded in six completely sequenced chloroplast genomes: Revised molecular estimates of two seed plant divergence times. Plant Syst Evol 206:337–351

Hansen A, Hansmann S, Samigullin T, Antonov A, Martin W (1999) Gnetum and the angiosperms: Molecular evidence that their shared morphological characters are convergent rather than homologous. Mol Biol Evol 16:1006–1009

Kimura M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kramer KU, Green PS (1993) Pteridophytes and gymnosperms. In: Kubitzki K (ed) The families and genera of vascular plants, Vol 1. Springer Verlag, Berlin, pp 282–361

Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15:1403–1411

Mathews S, Sharrock RA (1996) The phytochrome gene family in grasses (Poaceae): A phylogeny and evidence that grasses have a subset of the loci found in dicot angiosperms. Mol Biol Evol 13: 1141–1150

Page CN (1990) Coniferophytina. In: Kramer KU, Green PS (eds) Pteridophytes and gymnosperms [Vol 1, The families and genera of vascular plants. (Kubitzki K, ed)]. Springer Verlag, Berlin, p 292

Qiu Y-L, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastic and nuclear genomes. Nature 402:404–407

Raubeson LA, Jansen RK (1992) A rare chloroplast-DNA structural mutation is shared by all conifers. Biochem System Ecol 20:17–24

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Samigullin TK, Martin WF, Troitsky AV, Antonov AS (1999) Molecular data from the chloroplast *rpo*C1 gene suggest a deep and distinct dichotomy of contemporary spermatophytes into two monophyla: gymnosperms (including gnetales) and angiosperms. J Mol Evol 49:310–315

Sanderson MJ, Wojciechowski MF, Hu J-M, Sher Khan T, Brady SG (2000) Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Mol Biol Evol 17:782–797

Savard L, Li P, Strauss SH, Chase MW, Michaud M, Bousquet J (1994) Chloroplast and nuclear gene sequences indicate late Pennsylvanian time for the last common ancestor of extent seed plants. Proc Natl Acad Sci USA 91:5163–5167

Schmidt M (2000) Phytochrome in Gymnospermen: Genfamilien, Evolution und Expression in *Pinus sylvestris* L, Thesis. University of Cologne, Köln, Germany

Schneider-Poetsch HAW, Kolukisaoglu Ü, Clapham DH, Hughes J, Lamparter T (1998) Non-angiosperm phytochromes and the evolution of vascular plants. Physiol Plant 102:612–622

Shear WA (1991) The early development of terrestrial ecosystems. Nature 351:283–289

Sitte P, Ziegler H, Ehrendorfer F, Bresinsky A (1999) Evolution und Systematik. In: Strasburger: Lehrbuch der Botanik für Hochschulen, 34th ed. Spektrum Akademischer Verlag, Heidelberg, pp 459–819

Stefanovic S, Jager M, Deutsch J, Broutin J, Masselot M (1998) Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. Am J Bot 85:688–697

Stewart WN, Rothwell GW (1993) Paleobotany and the evolution of plants, 2nd ed. Cambridge University Press, Cambridge

Strickberger MW (1996) Evolution, 2nd ed. Jones and Bartlet, London, p 298

Swofford DL (1999) PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA

Winter K-U, Becker A, Münster T, Kim JT, Saedler H, Theissen G (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc Natl Acad Sci USA 96:7342–7347