The EMBO Journal vol.15 no.6 pp.1403–1411, 1996

RNA editing in bryophytes and a molecular phylogeny of land plants

Olaf Malek, Kathrin Lättig[1], Rudolf Hiesel[1], Axel Brennicke and Volker Knoop[2]

Allgemeine Botanik, Universität Ulm, Albert-Einstein-Allee, D-89069 Ulm and [1]Institut für Genbiologische Forschung, Ihnestrasse 63, D-14195 Berlin, Germany

[2]Corresponding author

RNA editing has been observed to date in all groups of vascular plants, but not in bryophytes. Its occurrence was therefore assumed to correlate with the evolution of tracheophytes. To gain more insight into both the phylogeny of early land plants and the evolution of mitochondrial RNA editing we have investigated a number of vascular and non-vascular plant species. Contrary to the belief that editing is absent from bryophytes, here we report mitochondrial RNA editing in *cox3* mRNA of the liverwort *Pellia epiphylla*, the mosses *Tetraphis pellucida* and *Ceratodon purpureus* and the hornwort *Anthoceros crispulus*. RNA editing in plants consequently predates the evolution of tracheophytes. Editing is also found in the eusporangiate ferns *Ophioglossum petiolatum* and *Angiopteris palmiformis*, the whisk fern *Tmesipteris elongata* and the gnetopsid *Ephedra gerardiana*, but was not detected in *Gnetum gnemon*. *cox3* mRNA of the lycopsid *Isoetes lacustris* shows the highest frequency of RNA editing ever observed in a plant, with 39% of all cytidine residues converted to uridines. The frequency of RNA editing correlates with the genomic GC content rather than with the phylogenetic position of a species. Phylogenetic trees derived from the slowly evolving mitochondrial sequences find external support from the assessments of classical systematics.

Keywords: land plant phylogeny/mitochondria/molecular evolution/RNA editing

Introduction

Plant mitochondrial RNA editing was initially found in monocotyledonous (Covello and Gray, 1989; Gualberto *et al.*, 1989) and dicotyledonous (Hiesel *et al.*, 1989) angiosperm species. Mostly cytidine to uridine and occasionally reverse exchanges modify the coding regions of virtually all mRNAs in plant mitochondria (Bonnard *et al.*, 1992). The unedited hybrid reading frame URF13 in male sterile maize T-cytoplasm (Ward and Levings, 1991) is a unique exception to this rule. RNA editing as manifested by cytidine–uridine exchange has also been found in chloroplasts, albeit at much lower frequency (Hoch *et al.*, 1991). Editing events in plant mitochondria are clustered in coding regions, but have also been described for 5′ and 3′ non-coding regions, for introns and for tRNAs (Maréchal-Drouard *et al.*, 1993; Binder *et al.*, 1994).

Mitochondrial RNA editing has been found in all angiosperm species investigated. The complete sequence of the liverwort *Marchantia polymorpha* mitochondrial DNA (Oda *et al.*, 1992), however, strongly suggests that RNA editing does not occur in this species; conserved amino acid positions are genomically encoded and editing events analogous to those in vascular plants are not observed in cDNA analysis of *Marchantia* (Ohyama *et al.*, 1993). Likewise, in the moss *Physcomitrella patens* conserved amino acids are encoded in the mitochondrial *cox3* reading frame at the genomic level (Marienfeld *et al.*, 1991). The mitochondrial *cox3* gene is, however, heavily edited in angiosperms and was therefore chosen to investigate the occurrence of RNA editing in the plant kingdom. A survey of major groups of land plants has started to bridge the large gap between the primitive non-vascular plants mentioned above and angiosperms (Hiesel *et al.*, 1994a). This work has shown that RNA editing is present in all vascular plants, including the 'fern allies' club mosses (Lycopsida), whisk ferns (Psilotopsida) and horsetails (Sphenopsida), the 'true' ferns (Filicopsida) and gymnosperms (Cycadopsida, Coniferopsida and Ginkgopsida). These groups are here regarded as classes for reasons of clarity, but are occasionally treated as separate divisions by some systematicists. RNA editing was not found in either green algae (*Coleochaete*, *Stichococcus*) or in the primitive moss *Sphagnum palustre*. Gymnosperm species, however, generally show an even higher frequency of RNA editing than has been found for the *cox3* gene in angiosperms. It was considered significant that a primitive tracheophyte, the club moss *Lycopodium squarrosum*, showed only one editing site in the 381 bp region of the *cox3* gene under investigation. Taken together, the findings led to speculation about the establishment of RNA editing along with the emergence of tracheophytes in land plant evolution (Hiesel *et al.*, 1994a).

The set of homologous mitochondrial sequences obtained during these studies yielded phylogenetic trees of considerable evolutionary depth (Hiesel *et al.*, 1994b). According to classical systematics a major group of land plants, the hornworts, was not represented in the set of species under investigation and the extended significance of the derived phylogenies depends upon inclusion of additional key species. We have now investigated *Anthoceros crispulus* as a hornwort species, as well as other representatives of early branches in land plant phylogeny. The surprising results obtained from these studies include evidence for RNA editing in the hornwort *A.crispulus*, the liverwort *Pellia epiphylla* and the two moss species *Ceratodon purpureus* and *Tetraphis pellucida*. Mitochondrial RNA editing in plants is therefore evolutionarily older than previously assumed. The highest frequency of

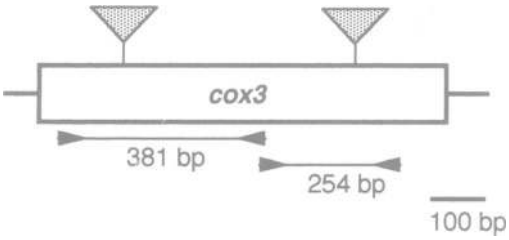

Fig. 1. Schematic representation of the mitochondrial *cox3* gene regions under investigation in this work. Arrowheads indicate oligonucleotide primers (for sequences see Materials and methods) used to amplify an upstream region (381 bp) and a downstream region (248/254 bp) of the reading frame (excluding primers). Each of these regions contains a group II intron (grey triangles) in the liverwort *M.polymorpha* (Oda *et al.*, 1992). Homologous introns are also present at identical sites in *P.epiphylla* (not shown) and the upstream intron is also present in *L.squarrosum* (Hiesel *et al.*, 1994a). RNA editing sites are deduced from comparison of cDNA and genomic sequences after PCR amplification, cloning and dideoxy sequencing.

RNA editing ever observed in a plant was found for the lycopod *Isoetes lacustris cox3* transcript. Phylogenetic trees constructed from a set of 28 species representing the entire spectrum of extant land plants confirm the highly informative potential of mitochondrial sequences for evolutionary analyses.

Results

Occurrence of RNA editing in different land plant lineages

To define the occurrence of mitochondrial RNA editing at the bryophyte–tracheophyte junction more precisely and to test the usefulness of mitochondrial nucleotide sequences for derivation of phylogenetic information we investigated several species representing early branches in land plant evolution. Most importantly, the group of hornworts is now represented for the first time by the species *A.crispulus*. Furthermore, we analysed the liverwort *P.epiphylla*, the mosses *T.pellucida* and *C.purpureus*, the lycopsid *I.lacustris*, the psilotopsid *Tmesipteris elongata*, the ferns *Ophioglossum petiolatum* and *Angiopteris palmiformis* and the gnetopsids *Ephedra gerardiana* and *Gnetum gnemon*. The green alga *Chara corralina*, as a member of the Charales, regarded as the extant algal representatives most closely related to the ancestors of land plants, is now likewise included in our studies. Experimentally we have, on the one hand, extended the available data set of homologous sequences from the upstream part of the *cox3* gene and, on the other hand, investigated the 3′-terminal region of this gene to statistically evaluate the inferences made from analysis of the upstream region (Figure 1).

An alignment of the upstream *cox3* sequences obtained in the analysis of the species together with some of those previously analysed is shown in Figure 2. No length variation of this *cox3* region is observed for any of the >30 species ranging from algae to seed plants investigated.

Liverwort RNA editing

Most surprisingly, and contradicting earlier assumptions, RNA editing is detected in the liverwort species *P.epiphylla*. In *Pellia* 12 editing sites are observed, three of which are silent (Figure 2). One silent exchange is a reverse edit in a histidine codon (CAT→CAC, position 57) which is conserved in all other species as either CAC or CAT at the DNA level. Different reverse edits in *cox3* have been observed in the filicopsids *Asplenium* and *Osmunda* (Hiesel *et al.*, 1994a). In these cases the reverse edits are essential to reconstitute conserved amino acid codons and to eliminate a stop codon in the *Asplenium* reading frame (position 181). The nine non-silent C→U editing events in *Pellia* likewise reconstitute conserved amino acid codons and thus confirm the necessity of RNA editing for functional gene expression in this liverwort. Each of the nine non-silent editing positions is conserved in at least two vascular plant species and nucleotide identities are re-established at the RNA level in comparison with its sister liverwort species *M.polymorpha* in these instances.

RNA editing in 'true' mosses (Musci)

The unexpected finding of RNA editing in the liverwort *Pellia* inevitably required reconsideration of the idea that RNA editing arose in evolution only after divergence of the vascular plant line. The finding prompted us to investigate further species of another bryophyte class, the 'true' mosses (Musci). We consequently analysed the species *C.purpureus* and *T.pellucida* in this study and indeed find evidence for RNA editing. Both species have a single RNA editing site in common that is required to reconstitute a universally conserved UGG tryptophan codon (position 301; Figure 2). While this codon is genomically encoded in other bryophyte species, editing at this site is also required in some of the tracheophytes. No further editing is required in the analysed region of *cox3* in *Ceratodon* and *Tetraphis*, since all other conserved amino acids are genomically encoded.

Hornwort RNA editing

Hornworts are a rather small class of bryophytes with an unclear relationship to the other plant groups, particularly the mosses and liverworts. The species *A.crispulus* is here investigated as a first representative of this group. Nucleic acid preparations of *Anthoceros* presented us with difficulties during PCR amplification, even when CsCl-purified material was used, and were inhibitory when added to other PCR reactions. Ultimately amplification succeeded after extreme dilution of the preparations. While unambiguous sequences were obtained from amplification of *Anthoceros* cDNA (Figure 2), a population of clones with diverging sequences were obtained from *Anthoceros* genomic DNA even after repeated attempts. We ascribe this phenomenon to the presence of an inhibitory and error-inducing substance associated with *Anthoceros* DNA. The deduction of editing sites in *Anthoceros* is thus somewhat preliminary, since additional sequence differences other than the expected C→U changes observed in the comparison of DNA and cDNA are found between individual genomic clones (for details see the corresponding database entry). These differences, however, are most likely PCR artefacts, for the reasons outlined above. Due to the scarcity of biological material purification of *Anthoceros* mitochondria to possibly overcome the problem is not feasible at present.

With the above caveat 10 editing sites are observed for *Anthoceros* in the upstream *cox3* region, nine of which

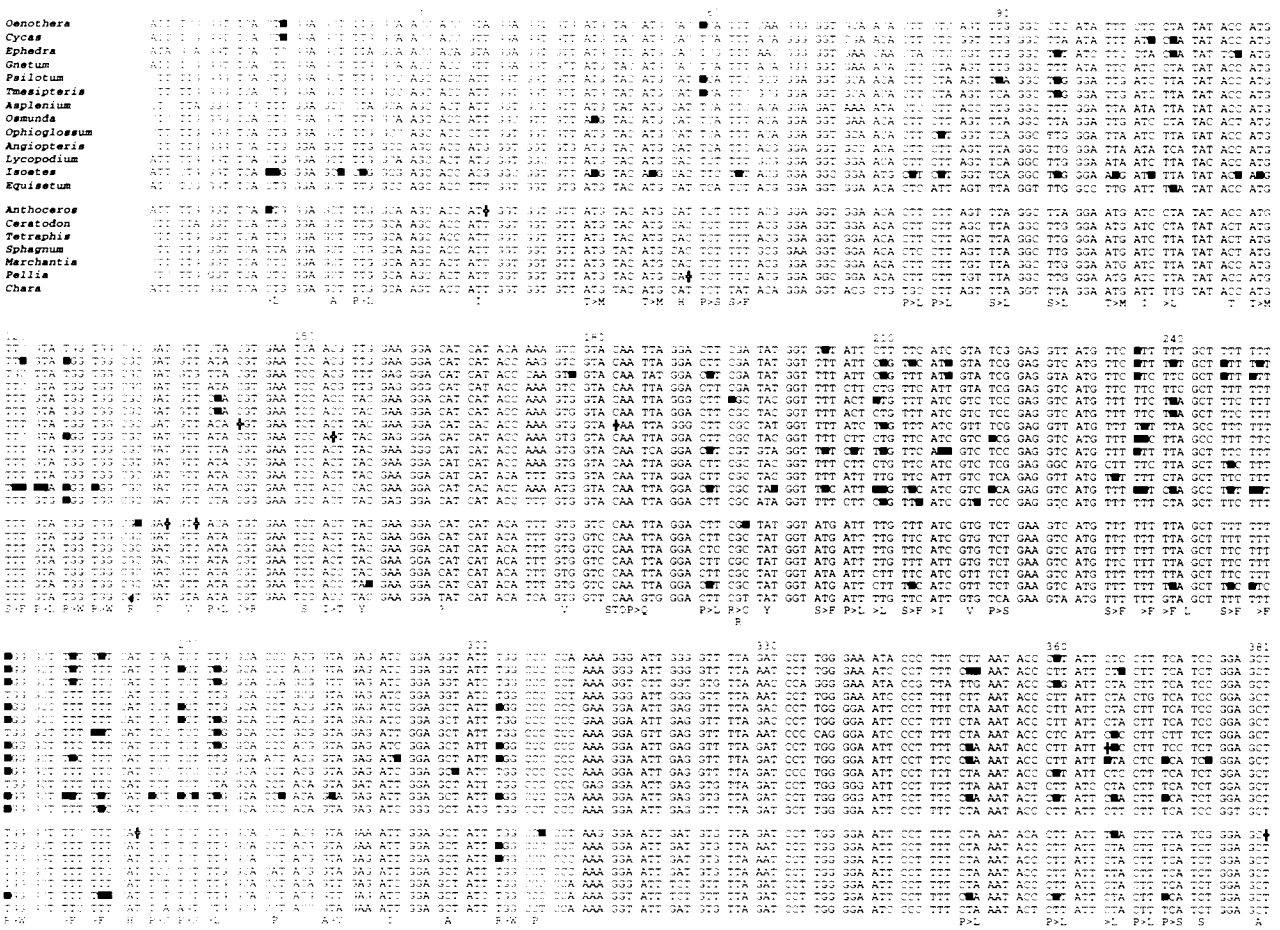

Fig. 2. Alignment of the 381 nt upstream region of the mitochondrial *cox3* gene amplified from the species indicated. Amino acid codons are given as triplets and the corresponding amino acids are indicated in the one letter code where changes are introduced by RNA editing. C→U editing events are indicated by black boxes, reverse edits by double crosses. The horizontal gap separates vascular (upper part) from non-vascular species (lower part). Homologous group II introns interrupt the *cox3* reading frame in *Lycopodium*, *Marchantia* and *Pellia* between positions 117 and 118. Database accession nos for the nucleotide sequences from top to bottom are: *O.berteriana* X76275; *Cycas revoluta* X76279; *E.gerardiana* X92735; *G.gnemon* X92722; *P.nudum* X76276; *T.elongata* X92738; *A.nidus* X76274; *O.claytoniana* X76277; *O.petiolatum* X92739; *A.palmiformis* X92731; *L.squarrosum* X76273; *I.lacustris* X92736; *E.arvense* X76282; *A.crispulus* X92721, *C.purpureus* X92733, *T.pellucida* X92737; *S.palustre* X76271; *M.polymorpha* M68929; *P.epiphylla* X92740; *C.corallina* X92734.

are silent exchanges. Eight editing events are observed in the third codon position. The single first codon position edit (at position 13, in common with *Isoetes* and the only editing site conserved in another species) leaves the affected leucine codon identity unchanged. Five cases each of conventional C→U edits and reverse exchanges are found. The only non-silent C→U exchange (position 365) reconstitutes a well-conserved leucine codon. An additional edit should be postulated for *Anthoceros* at position 142, where a conserved arginine codon is reconstituted in *Asplenium* by a reverse (U→C) exchange. While uncertainties about the genomic sequence remain, analysis of the 3′-terminal *cox3* sequence confirms the presence of editing in the hornwort (see below).

Tracheophyte RNA editing

The earlier hypothesis that RNA editing evolved in the tracheophyte line of land plant evolution was strengthened by the observation of only a single editing site in the *cox3* sequence of *L.squarrosum*, a tracheophyte at a basal phylogenetic position. The discovery of RNA editing in bryophytes prompted us to include the lycopsid *I.lacustris* as a sister group to the previously investigated club moss *Lycopodium*. In sharp contrast to the single editing site of *Lycopodium*, in this *cox3* region *Isoetes* shows the highest frequency of RNA editing ever observed in a plant. Of 122 cytidine residues, 48 (39%) are converted to uridines at the RNA level and only nine of these edits leave the amino acid identity unchanged. At positions 119–122 four consecutive C→U edits are found.

RNA editing is also identified in the psilotopsid *T.elongata*. Seven RNA editing sites are found in the investigated *cox3* coding region, six of which are conserved in the sister whisk fern *Psilotum nudum*. A thymidine is observed at position 272 of the genomic *Psilotum* sequence, where editing is required in the *Tmesipteris* RNA sequence to reconstitute a conserved leucine codon.

Both true fern species included in the earlier studies, *Asplenium nidus* and *Osmunda claytoniana*, are systematically classified as leptosporangiate ferns. We have now included the species *O.petiolatum* and *A.palmiformis* as members of the eusporangiate ferns. Editing patterns and frequency are highly divergent in the two species, with 16 sites identified in *Ophioglossum* and only

Fig. 3. Alignment of the downstream (248/254 bp) region of the *cox3* gene. Graphic details of the presentation are as in Figure 2. Homologous group II introns interrupt the *cox3* reading frame in the liverworts *Marchantia* and *Pellia* between positions 182 and 183. No editing sites are as yet determined for *Psilotum*, where only the cDNA sequence is available, as shown. Editing sites indicated in the *Equisetum* and *Cycas* sequences are deduced from sequence comparison and not yet confirmed by cDNA sequencing. Database accession Nos are: *C.revoluta* X93553; *G.gnemon* X92722; *P.nudum* X92724; *A.nidus* X92732; *O.claytoniana* X92725; *O.petiolatum* X92723; *E.arvense* X93552; *A.crispulus* X92730; *P.epiphylla* X92741.

four sites in *Angiopteris*. Only the editing event at position 250, reconstituting a tryptophan codon, is conserved between the two eusporangiate ferns. Five of the *Ophioglossum* base exchanges and one in *Angiopteris* are silent with respect to the amino acid encoded. Additional editing sites are postulated for *Angiopteris* at position 107 and for *Ophioglossum* at 185 to reconstitute conserved leucine codons and for *Angiopteris* at position 232 to re-establish a phenylalanine codon, but these are not found to be edited in the respective cDNA sequences.

A group of gymnosperms with few extant species, the gnetopsids, is considered to have a unique evolutionary position among the gymnosperms and has been suggested to be closely related to angiosperm plants. In *E.gerardiana* we find 14 conventional C→U editing sites, four of which are silent. No reverse editing is observed in this species and no further potential edits can be postulated from the sequence comparison to increase protein similarity. No RNA editing is observed in the *cox3* mRNA of the other gnetopsid investigated, *G.gnemon*, where well-conserved codon identities are already encoded in the genomic sequence.

Is plant mitochondrial RNA editing older than land plants?

The DNA sequence of the now included green alga *C.corallina* shows one site (position 163) that is expected to be edited at the RNA level by a reverse (U→C) editing event. A genomically encoded tyrosine UAU codon should be changed into the CAU codon for histidine, which is universally found encoded at this position in all other species. Remarkably, the histidine residue is even conserved in vertebrate *cox3* sequences. However, no evidence for RNA editing has been found in the analysis of cDNA sequences of *Chara*. The genomic sequence of *Chara* may point to the former existence and subsequent loss of RNA editing in this alga. More mitochondrial sequence data for *Chara* and related algae (e.g. *Nitella* spp.) will help to clarify this possibility.

The 3′-terminal part of the cox3 gene

To evaluate the results obtained for RNA editing frequencies and the phylogenetic tree construction (see below) derived from analysis of the upstream *cox3* region we have started a parallel investigation of the 3′-region of this gene (Figure 1). Additional interest in the 3′-terminal part of the *cox3* gene is based on the presence of a second group II intron in *cox3* of *M.polymorpha* (Figure 1) that has significant homology to the intron in the mitochondrial *rps10* gene of angiosperms described recently (Knoop *et al.*, 1995; Zanlungo *et al.*, 1995). A 248 bp region from the 3′-terminal half of the gene has been amplified from nucleic acid preparations of diverse species. A length variation is observed in this part of the *cox3* reading frame due to insertion of two codons after position 22 of the alignment in *Pellia* (Figure 3). The occurrence and frequency of RNA editing in the different species as deduced from analysis of the upstream *cox3* region are corroborated. Definite conclusions about editing can be drawn for *Oenothera*, *Pellia*, *Asplenium*, *Gnetum*, *Osmunda*, *Ophioglossum* and *Anthoceros*, where both DNA and cDNA sequences are available.

In the fern *Asplenium* a total of 18 editing events are observed in the downstream *cox3* region, with only eight conventional C→U edits and 10 reverse exchanges. As in the 5′-region, genomically encoded stop codons are removed from the *cox3* reading frame in three instances (positions 35, 95 and 230) in this part of the *Asplenium cox3* gene. Comparable with the observations made for the upstream *cox3* region, the second leptosporangiate fern species *Osmunda* also shows both types of editing. The two reverse edits (positions 176 and 185) in the *Osmunda* sequence are in common with *Asplenium* and are needed to restore well-conserved histidine codon identities. Three of the seven conventional C→U editing events observed for this part of *Osmunda cox3* mRNA are silent. Reverse edits have not been observed for the eusporangiate ferns in the upstream region, but one such event is now found in the 3′-region of *cox3* mRNA of *Ophioglossum*. This editing event (position 230) is needed to remove a genomic stop codon and is shared with *Asplenium* and *Anthoceros*. Additionally, eight conventional C→U editing events are observed in *Ophioglossum*, six of which reconstitute conserved amino acid codons.

The liverwort *Pellia* shows six conventional C→U edits

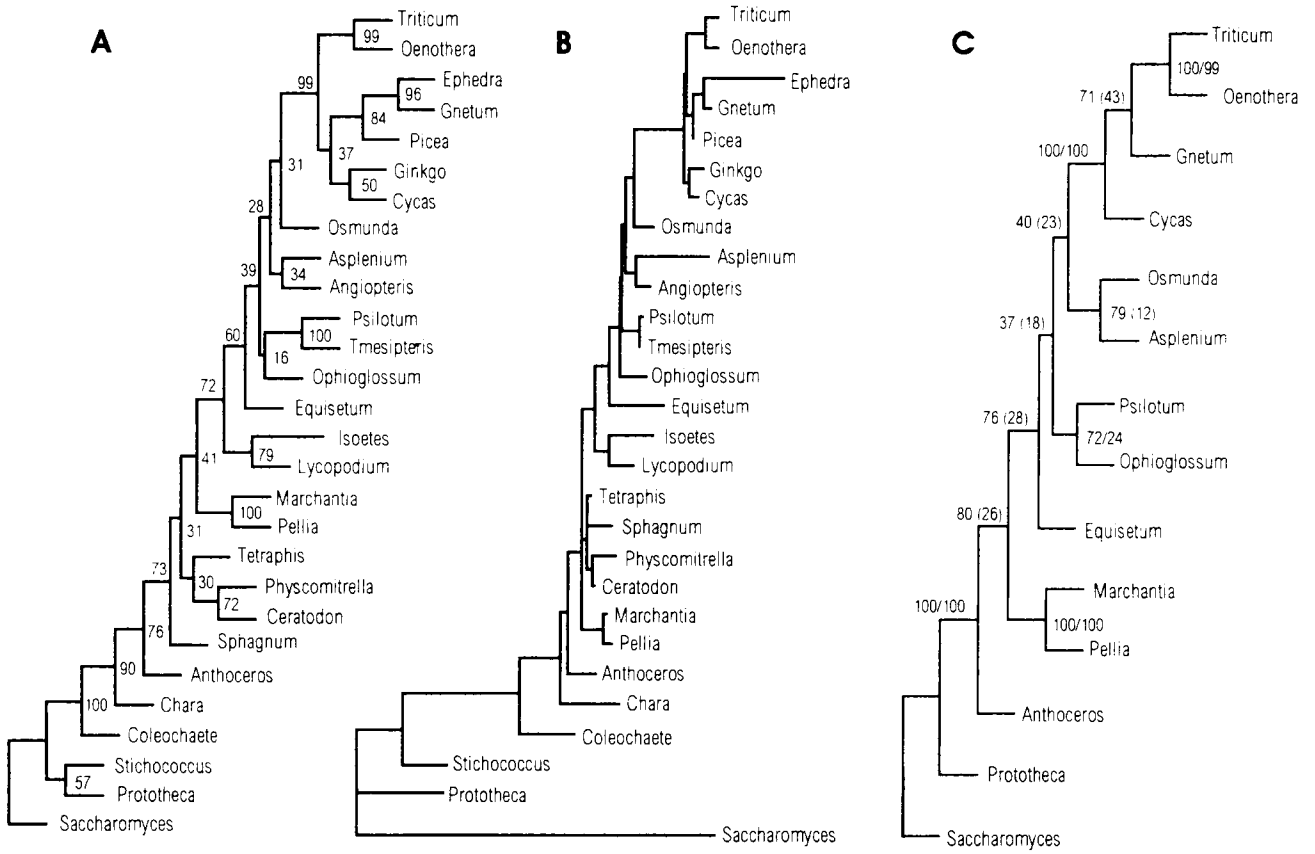

Fig. 4. Phylogenetic trees obtained with the PHYLIP program package. The yeast sequence was chosen as outgroup to root the trees in each case. Trees were reorganized by flipping certain branches using RETREE for the sake of a more convenient comparison. Sequences (cDNA) used were as shown in Figures 2 and 3 and accession nos for the species not included in these figures are: *Ginkgo biloba* X76280; *Triticum aestivum* X52539; *Picea abies* X76278; *Physcomitrella patens* X53679; *Stichococcus* X76272; *Coleochaete* X76270; *Prototheca wickerhamii* U02970; *Saccharomyces cerevisiae* M62622. Further data are available for a number of angiosperm species, but were not included due to lack of additional phylogenetic information. (**A**) Consensus derived from 100 trees constructed by the parsimony method (DNAPARS) after bootstrapping based on the alignment as shown in Figure 2. Numbers at the branches indicate the percentage of trees in which a group is present in the 100 bootstrap replicates of the data set. (**B**) Tree from the identical data set constructed by the maximum likelihood method (DNAML) with the default transition/transversion ratio of 2.0. (**C**) Tree as in (A) but based on the combined set of sequences (381 + 254 nt) shown in Figures 2 and 3, with bootstrap values indicated as in (A). Additionally, bootstrap values are given for the nucleotide data restricted to the upstream region after the oblique or in parentheses. Parentheses indicate that the given bifurcation is not present in the consensus tree topology of the smaller data set.

in this region, five of which are needed to re-establish conserved codon identities.

Comparison of DNA and cDNA sequences confirm RNA editing in *Anthoceros* in the 3′-terminal part of *cox3*. Sequences of independent genomic clones from *Anthoceros* DNA are homogeneous in this gene region. Both conventional and reverse edits at a total of 10 sites are observed in cDNA of *Anthoceros*. In the 3′-terminal part of the *cox3* gene, however, these edits are all non-silent and are required to re-establish amino acid identities or (in two cases) to remove in-frame stop codons, as observed for the fern *Asplenium*.

No conclusions about the distribution of editing sites in this region of *cox3* can as yet be made for *Psilotum*, where currently only the cDNA sequence is available. For *Cycas* and *Equisetum* only genomic sequences are at present available for this *cox3* region. In *Equisetum* one editing event is predicted at position 159, to reconstitute a phenylalanine codon universally conserved among other plants. In *Cycas*, however, 10 sites are predicted to be altered by C→U exchanges in order to re-establish conserved amino acid codons and this observation is in correspondence with the high editing frequency (19 sites) observed for *Cycas* in the 5′-region of the gene.

Phylogenetic aspects of land plant evolution

The cDNA sequences as listed in Figure 2 and others obtained earlier for the *cox3* gene have been used to construct phylogenetic trees by different methods implemented in the PHYLIP program package (Felsenstein, 1994). A consensus tree derived from 100 bootstrap replicates by the parsimony method (DNAPARS) is shown in Figure 4A. The maximum likelihood method (DNAML) results in a very similar tree topology for the data set of sequences from 28 species (Figure 4B). High bootstrap values for the parsimony consensus tree correspond to long branch lengths at the relevant nodes in the maximum likelihood tree and thus concurrently indicate branchings of high reliability. Consistently low bootstrap values in the parsimony analysis correspond to short node branch lengths in the maximum likelihood topology. Sister genera of the classically defined embryophyte groups angiosperms (*Oenothera berteriana* and *Triticum aestivum*), gnetopsids (*E.gerardiana* and *G.gnemon*), whisk ferns (*P.nudum* and

T.elongata), club mosses (*L.squarrosum* and *I.lacustris*) and liverworts (*M.polymorpha* and *P.epiphylla*) cluster as distinct clades in our trees in each case with high bootstrap reliability.

While the position of *Lycopodium* in relation to the bryophytes was not clearly resolved in earlier phylogenetic trees, the lycopsids now clearly branch off as the most basal tracheophyte group. This grouping is now widely supported by classical systematics, which also place lycopsids at the base of the tracheophyte phylogeny. Only one extant genus (*Equisetum*) is described for sphenopsids. The horsetail species under investigation, *E.arvense,* branches off after the lycopsids in the tracheophyte section of the tree.

The fern species expectedly occupy intermediate positions between the horsetail and the seed plants. The evolutionarily advanced leptosporangiate fern *Osmunda* groups as a neighbour to the seed plants and the eusporangiate fern *Ophioglossum* branches off next to the psilotopsids; the classical separation of eusporangiate and leptosporangiate ferns is otherwise not reflected in the trees. The leptosporangiate fern *Asplenium* and the eusporangiate fern *Angiopteris* are joined as sister groups (albeit with low bootstrap reliability) and branch off between the other ferns. While the branching of the eusporangiate fern *Ophioglossum* together with the whisk ferns observed with both tree-building methods may appear unusual, it should be noted that psilotopsids and Ophioglossales indeed share several classical developmental characteristics, such as a persistent prothallus and obligate mycorrhiza during this stadium.

Both the maximum likelihood and parsimony analyses include the hornwort *Anthoceros* as the most basal land plant lineage of the data set. External support for this branching comes from investigation of some morphological traits, e.g. chloroplast number and structure, which indicate significant similarity between *Anthoceros* and algal chloroplasts. Whether liverworts or mosses are to be regarded as the sister group to the tracheophytes remains largely unresolved in the maximum likelihood analysis (Figure 4B) and the inclusion of liverworts as a sister group to the tracheophytes is statistically insignificant in the parsimony analysis (Figure 4A). Among the four moss sequences the primitive moss *Sphagnum* occupies the basal position, as is also suggested by classical systematics. The two gnetopsids *E.gerardiana* and *G.gnemon* are expectedly joined as sister groups in the trees, but a specific affiliation of the gnetopsids with angiosperms remains statistically unsupported. The algal sequences from *Prototheca wickerhamii*, *Stichococcus*, *Coleochaete* and *Chara* assemble with high bootstrap reliability at the base of the embryophyte phylogeny in this ordering.

While some bootstrap values in Figure 4A clearly indicate affiliations of high reliability, other junctions are poorly supported statistically. To this end we have (upon suggestions made during the reviewing process) used the data available for the downstream *cox3* region (Figure 3) to base our trees on a larger data set. A parsimony tree based on the combined sequences of the upstream (381 nt) and downstream (254 nt) alignments from the set of species for which sequences of both regions are available is shown in Figure 4C. The resulting tree shows no essential topological differences from the trees based on a smaller number of informative sites (Figure 4A and B) with respect to the branching order of the major plant phyla (not considering the restricted species set). Some branchings are, however, now supported with higher bootstrap reliability. Most interestingly, *Ophioglossum* is still included as the nearest neighbour to the psilotopsid *Psilotum*, but with a greatly increased bootstrap reliability (72). *Equisetum* expectedly still occupies an intermediate position between bryophytes and the other vascular plants. Also, the inclusion of *Anthoceros* as the most basal genus in the land plant phylogeny remains unaltered, albeit with a higher bootstrap reliability. Likewise, *Gnetum* is included as the sister group of the angiosperms with higher reliability, but the restricted species sampling for the set of extended sequences must be kept in mind as a cautionary reservation.

Discussion

The phylogeny of land plants

The macrosystematics and phylogeny of early land plants are still being debated. An excellent summary reviewing classical and molecular approaches to define a comprehensive land plant phylogeny is provided in a recent paper by Manhart (1994). It is generally assumed that a green alga related to the class Charophyceae gave rise to the embryophytes. Different systematic classifications have been suggested for green algae that are distinguished by the number of orders belonging to the Charophyceae. In agreement with the recently established view that an alga specifically related to the order Charales gave rise to land plants, our phylogenies place *Chara* as a sister group to the latter. *Coleochaete* and *Stichococcus* are included in different orders of the Charophyceae (Coleochaetales and Klebsormidiales respectively) by some systematicists or even in different classes by others and are found at more basal positions in our trees. The inference that *Chara* is more closely related to embryophytes than *Coleochaete* contradicts a phylogeny obtained recently from nuclear rRNA sequences (Kranz *et al.*, 1995).

Bryophytes are considered a paraphyletic group at the root of the embryophyte tree with only the true mosses (Musci) or the liverworts giving rise to tracheophytes and hornworts as a sister group to all other land plants. The seven bryophyte species included here indeed cluster as a paraphyletic group at the bottom of the embryophyte phylogeny, but it remains unresolved whether mosses or liverworts are more likely as the sister clade to the tracheophytes.

Due to the lack of useful morphological or physiological characters a definite phylogeny for the early vascular plants (and here mainly for the three fern ally groups psilotopsids, lycopsids and sphenopsids) remains to be established. Molecular studies have supported the view that lycopsids are most closely related to bryophytes (Raubeson and Jansen, 1992). Our studies clearly support the placement of lycopsids at the root of the tracheophyte tree, since the members of the two lycopsid genera *Isoetes* and *Lycopodium* are placed together as a sister group to all other vascular plants. Likewise, the two psilotopsid genera *Psilotum* and *Tmesipteris* constitute a dichotomy of high reliability branching off after the sphenopsid *Equisetum* in the vascular plant phylogeny. The unequi-

vocal affiliation of *Tmesipteris* with *Psilotum* questions a suggested close relation to hornworts based on a morphological trait (Frey *et al.*, 1994) and rather confirms its earlier classification as a psilotopsid. Clearly though, the data set has to be extended to allow statistically meaningful assessments concerning the grouping of leptosporangiate and eusporangiate ferns and the here observed affiliation of *Ophioglossum* with the psilotopsids. Molecular data may in the future render these classically defined groups phylogenetically meaningless. An extension of the data set should take into account both a wider sampling of key species and more phylogenetically informative positions, possibly based on additional and slightly faster evolving mitochondrial protein genes.

Different molecular markers for phylogenetic reconstruction

In the contribution of Manhart (1994) the chloroplast *rbcL* gene (encoding the large subunit of the ribulose bisphosphate carboxylase) was used as a molecular marker in an attempt to deduce a similarly comprehensive phylogeny including representatives of all land plant groups and diverse green algae. The phylogenetic trees obtained, however, were in conflict with phylogenetic assessments derived from classical characters, even after modifying diverse parameters of tree construction. The chloroplast phylogenies (Manhart, 1994) relied on ~1.5 kb of *rbcL* gene sequence, i.e. nearly four times the mitochondrial sequence data reported here. The unsuitability of a special molecular sequence for a given purpose, inadaequate taxon sampling or unequal rates of evolution are generally responsible for unsatisfactory results in phylogenetic tree construction. Chloroplast sequences in general appear to be better suited to phylogenetic analysis close to the familial level (Clegg and Zurawski, 1992). Additionally, as also discussed by Manhart (1994), RNA editing may play an important role in determining the phylogenetically informative nucleotide sequence in the chloroplast *rbcL* gene and may thus render attempts to obtain reliable phylogenies directly from genomic DNA sequences questionable. In this respect a survey of land plants for the occurrence of editing in chloroplasts, as reported here for mitochondria, may give informative insights into both the evolution of RNA editing and the employment of chloroplast nucleotide sequences in phylogenetic studies. It will be interesting to see whether the occurrence of RNA editing in the two plant organelles correlates, i.e. whether all vascular plants and some bryophytes show chloroplast RNA editing. Such a finding would consequently prompt speculation about a common evolutionary origin and maybe even a common mechanism for the RNA editing process in chloroplasts and mitochondria.

Evolution and raison d'être of RNA editing

The existence of RNA editing in all three classes of bryophytes is probably the most surprising finding reported here; plant mitochondrial RNA editing is consequently older than previously assumed. The differences in RNA editing between the two liverworts *Marchantia* and *Pellia*, the two lycopsids *Isoetes* and *Lycopodium* and also the two gnetopsids *Gnetum* and *Ephedra* show that RNA editing frequency does not correlate with the phylogenetic position of a species. Correlating the occurrence of editing in plant mitochondria with the phylogentic trees suggests more than one independent loss or gain of editing activity. The suspicious case of a TAT (tyrosine) codon in *Chara* where a conserved CAT (histidine) codon is found in all other species may indicate that RNA editing even predates the occurrence of land plants and was only later lost in certain lines of land plant evolution, such as thalloid liverworts (*Marchantia*). Additional sequence data of the Charales and Musci will verify or falsify the assumptions about the absence of RNA editing in distinct species of these groups.

Very much like the splicing of introns, RNA editing is an apparently superfluous detour in the cell to establish an RNA sequence that could just as well be transcribed directly from a co-linear DNA sequence. Phenomena like alternative splicing and regulatory influences may 'justify' the existence of introns in the eukaryotic nucleus and similar speculations about the reasons for RNA editing in plant organelles have been made. RNA editing may add additional dimensions to gene expression, allowing for different proteins to be made from only one gene by differential editing or providing additional features for gene regulation. Still more sophisticated speculations include the proposal that RNA editing may help to avoid certain DNA sequences incompatible with other processes in the organelles. There is little support for this idea, however, due to the lack of convincing common sequence features surrounding editing sites.

As a simple alternative rationale editing may just compensate at the RNA level for genetic drift from thymidine to cytidine in DNA (i.e. towards an increased G+C content) to re-establish the required codon identities. Consequently, a correlation is to be expected between the G+C content of a sequence and the observed frequency of editing. To test this hypothesis we plotted the genomic G+C content in *cox3* against the number of C→U RNA

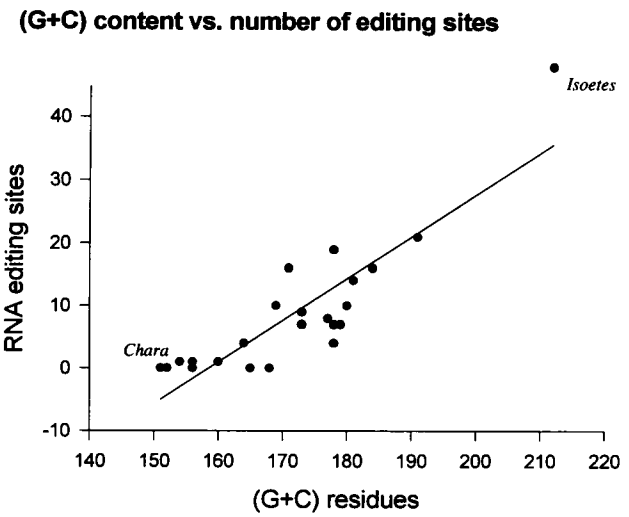

Fig. 5. Number of G+C residues in the upstream 381 nt *cox3* region (*x* axis) plotted against the number of RNA editing sites (*y* axis). Each C→U editing event is counted as +1 and each reverse editing (U→C) as −1. The set of land plant species is as in Figure 4A (excluding *Anthoceros*, but including the alga *Chara*). Linear regression results in a correlation coefficient r^2 of 0.77 and the slope of the regression curve is 0.67. Changes in G+C content are mainly due to changes in C residue number on the coding strand, with only minor variations in the number of G residues (87 ± 6).

editing events for the species investigated (Figure 5). A positive linear correlation (r^2 = 0.77) between editing frequency and G+C content is observed. This observation is in agreement with the idea that RNA editing may allow a drift of mitochondrial DNA sequences from U to C which can be compensated for at the RNA level once the RNA editing machinery had been established. A model has been proposed in which a RNA editing activity (e.g. a deaminase activity fortuitously operating on polyribonucleotides) becomes fixed by genetic drift after mutations at the DNA level require RNA alterations to express the conserved protein sequence from the edited RNA (Covello and Gray, 1993). This idea is in accord with the observed proportionality of editing frequency and genomic G+C content observed here for the *cox3* gene. Inherent in this idea, however, is the possibility that editing can be lost in the evolution of certain plant lineages (e.g. *Marchantia*) when the genomic G+C content drifts back to lower values, as suggested by the phylogeny presented here.

Materials and methods

Biological material

When available, fresh plant material was kindly provided by the Berlin Botanical Garden. *Tmesipteris elongata* was a kind gift from Drs Frey and Kürschner (Berlin) and *Isoetes lacustus* was kindly provided by Drs Moberg and Martinsson (Uppsala). *Anthoceros crispulus* and *P.epiphylla* were kindly made available by Dr Binding (Kiel). *Anthoceros* and *Pellia* were cultivated on MS agar. *Chara corralina* grown in culture was kindly provided by Dr Fisahn (Berlin) and *T.pellucida* was obtained with the kind help of Dr Muhle (Ulm). Nucleic acids were extracted either immediately after harvesting or after storage at −20°C.

Molecular biological procedures

Nucleic acids were prepared from 0.5–3 g plant material by the CTAB method (Doyle and Doyle, 1990). Crude nucleic acid preparations were separated into RNA and DNA by differential precipitation in the presence of lithium acetate and subsequent digestion with DNase and RNase respectively. The Boehringer Mannheim kit was used for cDNA synthesis with random hexamer primers. An aliquot of the cDNA synthesis reaction was used directly for PCR amplification. PCR amplification was at 50 or 45°C annealing temperature with oligonucleotide primers (5′→3′) cox315 (GTAGATCCAAGTCCATGGCCT), cox313 (GCATGATGGGCCCAAGTTACGGC), cox325 (CATCCGGAGCTCCCGTGACTTGGGC) and cox323 (GCGAATTCAAAGCCAACGTGATGCT). Oligonucleotides were purchased from TIB Molbiol (Berlin). PCR assays were routinely purified on Microcon-30 columns (Amicon Inc.) and eluted from preparative agarose gels after restriction digestion. PCR products were cloned into Bluescript (Stratagene) vectors via restriction sites in the primer sequences. Approximately four clones were sequenced for each DNA and cDNA cloning of a species by the dideoxy chain termination procedure using the Pharmacia sequencing kit. Alternatively, PCR was carried out with one biotinylated primer and subsequently one strand of the PCR product was purified using streptavidin-coated Dynabeads M-280 (Dynal) and used directly as the template for dideoxy sequencing. Comparison of cDNA and DNA sequences was used as a control for PCR errors where generally only C↔U exchanges attributable to RNA editing were observed. The majority of editing events can be predicted from comparison of the genomic sequences and were generally confirmed by cDNA sequencing. Rare sequence divergencies in one of a set of clones were considered PCR artefacts. A higher rate of such divergence (mostly G↔A) was observed for independent genomic clones in *Anthoceros* (see the corresponding database entry).

Computer analysis

The UWGCG program package for the VAX/VMS (Genetics Computer Group, 1994) was used for initial sequence handling and analysis. Phylogenetic tree construction (Felsenstein, 1988) was with the PHYLIP 3.56 package in the Power Macintosh version obtained via ftp from evolution.genetics.washington.edu (Felsenstein, 1994). Default values were used for run parameters; further details are given in Results and the figure legends.

Acknowledgements

We are very grateful to our colleagues for providing plant material and helping with classical systematics, most importantly to Drs Moberg and Martinsson (Uppsala), Frey and Kürschner (Berlin), Muhle (Ulm), Binding (Kiel) and Fisahn (Berlin) and T.Dürbye and his colleagues from the Berlin Botanical Garden. Additional thanks are due to Herrn Fritsche for taking care of the greenhouse plants. Finally, we would like to thank three anonymous referees for helpful comments to improve this manuscript.

References

Binder,S., Marchfelder,A. and Brennicke,A. (1994) RNA editing of tRNAPhe and tRNACys in mitochondria of *Oenothera berteriana* is initiated in precursor molecules. *Mol. Gen. Genet.*, **244**, 67–74.

Bonnard,G., Gualberto,J.M., Lamattina,L. and Grienenberger,J.-M. (1992) RNA editing in plant mitochondria. *Crit. Rev. Plant Sci.*, **10**, 503–524.

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

Covello,P.S. and Gray,M.W. (1989) RNA editing in plant mitochondria. *Nature*, **341**, 662–666.

Covello,P.S. and Gray,M.W. (1993) On the evolution of RNA editing. *Trends Genet.*, **9**, 265–268.

Doyle,J.J. and Doyle,J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Felsenstein,J. (1988) Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.*, **22**, 521–565.

Felsenstein,J. (1994) *Program Manual for the PHYLIP 3.5c Package*. Washington University, Seattle, WA.

Frey,W., Campbell,E.O. and Hilger,H.H. (1994) Structure of the sporophyte–gametophyte junction in *Tmesipteris elongata* P.A. Dangeard (Psilotaceae, Psilotopsida) and its phylogenetic implications—a SEM analysis. *Nova Hedwigia*, **59**, 21–32.

Genetics Computer Group (1994) *Program Manual for the Wisconsin Package Version 8*. Genetics Computer Group, University of Wisconsin, Madison, WI.

Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.-H. and Grienenberger, J.-M. (1989) RNA editing in wheat results in the conservation of protein sequences. *Nature*, **341**, 660–662.

Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) RNA editing in plant mitochondria. *Science*, **246**, 1632–1634.

Hiesel,R., Combettes,B. and Brennicke,A. (1994a) Evidence for RNA editing in mitochondria of all major groups of land plants except the bryophyta. *Proc. Natl Acad. Sci. USA*, **91**, 629–633.

Hiesel,R., von Haeseler,A. and Brennicke,A. (1994b) Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl Acad. Sci. USA*, **91**, 634–638.

Hoch,B., Maier,R.M., Appel,K., Igloi,G.L. and Kössel,H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, **353**, 178–180.

Knoop,V., Ehrhardt,T., Lättig,K. and Brennicke,A. (1995) The gene for ribosomal protein S10 is present in mitochondria of pea and potato but absent from those of *Arabidopsis* and *Oenothera*. *Curr. Genet.*, **27**, 559–564.

Kranz,H.D., Miks,D., Siegler,M.-L., Capesius,I., Sensen,C.W. and Huss,V.A.R. (1995) The origin of land plants: phylogenetic relationships among charophytes, bryophytes and vascular plants inferred from complete small-subunit ribosomal RNA sequences. *J. Mol. Evol.*, **41**, 74–84.

Manhart,J.R. (1994) Phylogenetic analysis of green plant *rbcL* sequences. *Mol. Phylogenet. Evol.*, **3**, 114–127.

Maréchal-Drouard,L., Ramamonjisoa,D., Cosset,A., Weil,J.-H. and Dietrich,A. (1993) Editing corrects mispairing in the acceptor stem of bean and potato mitochondrial phenylalanine transfer RNAs. *Nucleic Acids Res.*, **21**, 4909–4914.

Marienfeld,J.R., Reski,R. and Abel,W.O. (1991) The first analysed archegoniate mitochondrial gene (*cox3*) exhibits extraordinary features. *Curr. Genet.*, **20**, 319–329.

Oda,K. *et al.* (1992) Gene organization deduced from the complete

sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. *J. Mol. Biol.*, **223**, 1–7.

Ohyama,K., Oda,K., Ohta,E. and Takemura,M. (1993) Gene organization and evolution of introns of a liverwort, *Marchantia polymorpha*, mitochondrial genome. In Brennicke,A. and Kück,U. (eds), *Plant Mitochondria*. VCH Verlagsgesellschaft, Weinheim, Germany, pp.115–129.

Raubeson,L.A. and Jansen,R.K. (1992) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science*, **255**, 1697–1699.

Ward,G.C. and Levings,C.S. (1991) The protein-encoding gene T-*urf13* is not edited in maize mitochondria. *Plant Mol. Biol.*, **17**, 1083–1088.

Zanlungo,S., Quinones,V., Moenne, A., Holuigue,L. and Jordana,X. (1995) Splicing and editing of *rps10* transcripts in potato mitochondria. *Curr. Genet.*, **27**, 565–571.

Received on September 6, 1995; revised on November 24, 1995