Phylogenetic significance of the *rpoA* loss in the chloroplast genome of mosses

Bernard Goffinet[1], Norman J. Wickett[1], A. Jonathan Shaw[2] & Cymon J. Cox[2]

[1] *Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs 06269, Connecticut, U.S.A. goffinet@uconn.edu (author for correspiondence); norman.wickett@uconn.edu*
[2] *Department of Biology, Duke University, Durham 27708, North Carolina, U.S.A. shaw@duke.edu; cymon@duke.edu*

A recent survey of arthrodontous mosses revealed that their chloroplast genome lacks the gene encoding the alpha subunit of the RNA polymerase (i.e., *rpoA*), and that at least in *Physcomitrella patens* the gene has been transferred to the nuclear genome. Subsequently the gene was recorded from the cytoplasmic genome in *Takakia* and *Sphagnum*. Here we extend the survey to representatives of all major lineages of mosses to determine when in the evolutionary history of the Bryophyta the loss took place. Amplifications using primers annealing to the flanking regions of the *rpoA* gene yield a product that contains the gene in *Takakia*, *Sphagnum*, *Andreaea*, *Oedipodium*, Polytrichaceae, and *Buxbaumia*. The gene is lacking in all arthrodontous mosses, including *Diphyscium* but also in both species of *Tetraphis*. Reconstruction of the transfer on the phylogeny of mosses suggests (a) that the *rpoA* gene was lost twice and (b) that the gene was lost after the divergence of Buxbaumiidae and prior to the divergence of *Diphyscium* from the remaining Bryopsida.

KEYWORDS: Bryophyta, chloroplast genome, gene transfer, phylogeny, *rpoA*.

INTRODUCTION

The genomes of the chloroplast and the mitochondrion have undergone severe reduction in size since their acquisition as endosymbionts by a protoeukaryotic cell (Dyall & al., 2004; Knoop, 2004; Timmis & al., 2004). In plants, genes lacking from these genomes have either been completely lost (such as genes required for a free-living existence) or have been transferred to the nucleus (Korpelainen, 2004). In fact, many of the proteins occurring in the mitochondrion and the chloroplast are synthesized in the cytosol and subsequently imported into the organelles. Analyses of the nuclear genome of *Arabidopsis* suggest that possibly as much as 18% of its genes have a cyanobacterial, and hence a plastid origin (Martin & al., 2002). By contrast, translocations between the cytoplasmic genomes seem very rare. The chloroplast genome of the angiosperm *Calycanthus* contains a gene of mitochondrial origin (Goremykin & al., 2003), and the mitochondrial genomes of maize and rice comprise various sequences of chloroplast origin (Nakazano & Hirai, 1993; Zheng & al., 1997). Transfers from the cytoplasmic to the nuclear genome characterize particular lineages of land plants (e.g., Turmel & al., 1999; Adams & al., 2002, 2003; Martin & al., 2002; Hackett & al., 2004), but their phylogenetic significance is often ambiguous, as organellar genomes have only been sequenced for few taxa, from the major lineages of green plants (see Timmis & al., 2004). One exception is the presence in the nucleus of two mitochondrial genes (i.e., *rps14* and *rpl15*) in maize and rice, and hence potentially in many or all grasses (Sandoval & al., 2003). However, the process of gene transfer is ongoing (Timmis & al., 2004), and the rate of transfer estimated from the progeny of artificial crosses (Huang & al., 2002) suggests that losses from the cytoplasmic genomes are not rare events (Richly & Leister, 2004a, b). These observations are not surprising considering that a single locus may be lost repeatedly during the evolution of a lineage, such as *infA* during the diversification of angiosperms (Millen & al., 2003).

To date the chloroplast genome has been completely sequenced for three bryophytes: *Marchantia polymorpha* L. (Ohyama & al., 1986), *Anthoceros formosae* Steph. (Kugita & al., 2003), and *Physcomitrella patens* (Hedw.) Bruch & W.P. Schimp. (Sugiura & al., 2003). The *rps16* gene is lacking from the *Marchantia* genome, and has been lost from the plastid at least four times in the evolution of green plants (Martin & al., 2002). The genome of *Anthoceros formosae* is characterized by the presence of pseudogenes for *matK* and *rps15* (Kugita & al., 2003). The structure of its inverted repeat is similar to that of vascular plants since it includes the *rps12* gene (or only the 3' exon, according to Kugita & al. [2003]) in the inverted repeat (Kelch & al., 2004). The reconstruction of the genome of *Physcomitrella patens* by Sugiura & al. (2003) revealed (1) that a fragment encompassing much of the large single copy of the genome is inverted, and (2) that the *rpoA* gene, which codes for the alpha subunit of the RNA polymerase, an enzyme essential for catalyzing the transcription of DNA into RNA, is lacking.

Sugiura & al. (2003) demonstrated that the *rpoA* gene is present in the nucleus. They surveyed four other species of the Bryopsida sensu Goffinet & Buck (2004; i.e., the arthrodontous mosses) for the presence of the gene in cpDNA by targeting the putative *rpoA* region using primers specific to the genes flanking the *rpoA* gene in *Anthoceros* and *Marchantia*. The amplicons that they obtained comprise portion of the *rps11* and the *petD* gene but lack the *rpoA* locus. Thus, they concluded that the transfer of the *rpoA* gene from the chloroplast to the nuclear genome is diagnostic of mosses. The diversification of arthrodontous mosses is, however, preceded by the divergence of several other lineages, including *Takakia*, *Sphagnum*, *Andreaea*, *Andreaeobryum*, and the nematodontous mosses (Cox & al., 2004). Hence, Sugita & al. (2004) extended their survey to the earliest diverging lineages of mosses, *Takakia* and *Sphagnum*, and documented the presence of the *rpoA* gene in their chloroplast genome. Their result suggests that the transfer of the *rpoA* gene from the cytoplasmic to the nuclear genome occurred during and not prior to the evolution of mosses. Here we target representatives of the Andreaeopsida, Andreaeobryopsida, Oedipodiopsida, Polytrichopsida, Tetraphidopsida, and basal Bryopsida, to determine when during the early diversification of mosses, the loss of the *rpoA* gene from the chloroplast genome took place.

MATERIALS AND METHODS

Taxon sampling. — We surveyed those thirty representatives of all classes and most subclasses of mosses (sensu Goffinet & Buck, 2004) used by Cox & al. (2004) for inferring the phylogeny of the Bryophyta based on multiple genes. In a few cases, another voucher specimen or another congeneric species was used as an exemplar (Table 1). Furthermore, as preliminary amplication results suggested that the *rpoA* gene was absent from *Tetraphis pellucida*, we also tested for the presence of this gene in *Tetraphis geniculata* (Table 1).

DNA extraction, PCR amplification and sequencing. — The DNA of *T. geniculata* was extracted following a modification of the protocol by Doyle and Doyle (1987) as described in Goffinet & al. (1998). For all other new vouchers, the DNA was isolated using the DNeasy Plant Mini kit from Qiagen, following the manufacturer's instructions. The *rpoA* gene is located in the *Marchantia* and the *Anthoceros* genome between the *rps11* and the *petD* gene. We targeted this region using the primers P1 and P3 designed by Sugiura & al. (2003; Table 2, Fig. 1). The original reverse primer designed by these authors anneals to the 3' end of the *petD* gene; we replaced P3 with a new primer *petD*2 (Table 2), which anneals closer to the 5' end of the coding region (Fig. 1). We further tested for the presence of the *rpoA* gene by using P1 in combination with a reverse primer (*rpoA*1; Table 2) that anneals within the gene itself (Fig. 1). Finally we amplified the virtually complete *rpoA* gene based on new internal primers (Table 2) that annealed near the ends of the coding sequence. These primers were used with an annealing of 48°C to amplify the *rpoA* gene in taxa where this gene is presumed to be located in the nucleus. The polymerase chain reaction was performed in a 25 µL volume with one unit of Hot Master Taq polymerase (Eppendorf AG), one µL each of a 10 mM solution of each primer, a 10 mM solution of the dNTPs, and a 99.9% pure solution of dimethyl sulfoxide

Table 1. Taxa sampled in this study and vouchers used for DNA extraction. Asterisks indicate that the voucher specimen or the generic exemplar differs from that used by Cox & al. (2004).

Taxon, Family, Voucher

Alophosia azoricum (Ren. & Card.) Card., Polytrichaceae, *Rumsey s.n.* (DUKE). **Andreaea ruspetris* Hedw. (instead of *A. wilsonii* Hook. f.), Andreaeaceae, *Schofield 110817* (DUKE). *Andreaeobryum macrosporum* Steere & B. M. Murray, Andreaeobryaceae, *Schofield 78094* (DUKE). *Aulacomnium turgidum* (Wahlenb.) Schwägr., Aulacomniaceae, *Hedderson 6385* (RNG). **Bartramia pomiformis* Hedw. (instead of *B. stricta* Brid.), Bartramiaceae, *Hax CT1* (CONN). *Brachythecium salebrosum* (Hoffm. ex F. Weber & D. Mohr) Schimp., Brachytheciaceae, *Goffinet 4723* (DUKE). *Buxbaumia aphylla* Hedw., Buxbaumiaceae, *Belland 16889* (DUKE). *Dendroligotrichum dendroides* (Brid. ex Hedw.) Broth., Polytrichaceae, *Goffinet 5425* (DUKE). **Diphyscium foliosum* (Hedw.) D. Mohr, Diphysciaceae, *Goffinet 4595* (DUKE). *Encalypta ciliata* Hedw., Encalyptaceae, *Schofield 98872* (DUKE). *Entosthodon laevis* (Mitt.) Fife, Funariaceae, *Goffinet 5601* (DUKE). *Fissidens subbasilaris* Hedw., Fissidentaceae, *Goffinet 5263* (DUKE). *Funaria hygrometrica* Hedw., Funariaceae, *Cox 148* (RNG). **Hedwigia ciliata* (Hedw.) P. Beauv., Hedwigiaceae, *Goffinet 8955* (CONN). **Hookeria lucens* (Hedw.) Sm., Hookeriaceae, *Goffinet 8030* (CONN). *Mielichhoferia elongata* (Hoppe & Hornsch.) Nees & Hornsch., Mniaceae, *Shaw s.n.* (RNG). **Mnium hornum* Hedw., Mniaceae, *Goffinet 8830* (CONN). *Oedipodium griffithianum* (Dicks.) Schwägr., Oedipodiaceae, *Schofield 98670* (DUKE). *Orthodontium lineare* Schwägr., Orthodontaceae, *Hedderson s.n.* (RNG). **Orthotrichum lyellii* Hook. & Taylor, Orthotrichaceae, *Goffinet 9126* (CONN). *Polytrichadelphus purpureus* Mitt., Polytrichaceae, *Cox 84/01* (DUKE). *Polytrichum pallidisetum* Funck, Polytrichaceae, *Goffinet 4581* (DUKE). **Pyrrhobryum spiniforme* (Hedw.) Mitten, Rhizogoniaceae, *Goffinet 5107-b* (CONN). *Rhodobryum giganteum* (Schwägr.) Paris, Bryaceae, *Longton 5073* (RNG). **Scouleria aquatica* Hook., Scouleriaceae, *Goffinet 7969* (CONN). **Sphagnum fimbriatum* Wils. (instead of *S. palustre* L.), Sphagnaceae, *Goffinet 5758* (CONN). **Takakia lepidozioides* S. Hatt. & Inoue, Takakiaceae, *Rupp 10/8/1990* (CONN). *Tetraphis geniculata* Girgensohn ex Milde, Tetraphidaceae, *Schofield 103022* (DUKE). **Tetraphis pellucida* Hedw., Tetraphidaceae, *Goffinet 4542* (CONN). **Tetraplodon mnioides* (Sw. ex Hedw.) Bruch & Schimp., Splachnaceae, *Shaw 9024* (DUKE). *Timmia megapolitana* Hedw., Timmiaceae, *Schofield 97597* (DUKE).

Table 2. Primers used to test for the presence of the *rpoA* gene in the chloroplast genome of mosses.

P1	TTTTGTTCGTGATGTAACTCCTATG	Sugiura & al., 2003
P2	CTACCATAGCATCCTCAGTAGATT	Sugiura & al., 2003
P3	CTAAATTAGCTAAAGGTATGGGTC	Sugiura & al., 2003
*petD*2	CTTTGGTTGGGTATTGGAGCTG	this study
*rpoA*1	GGTTAAACTTCCATTGGTCC	this study
*rpoA*F-47	GTGGAGGTGTGTTGAATC	this study
*rpoA*R-965	GCTTCTAAAACTTGTTCWAC	this study

(DMSO).

The amplification profile when using P1 and P3 is: 95°C for 1 min followed by 30 cycles of denaturation (1 min at 95°C), annealing (1 min at 48°C), extension (1.5 min at 72°C), and a final extension at 72°C for 7 min. Amplicons were purified using the QIAquick PCR purification kit from Qiagen (Valencia, California, U.S.A.). The annealing temperature was raised to 55°C when *petD*2 was used as reverse primer, but retained at 48°C when the *rpoA*1 internal reverse primer was used.

All amplicons were sequenced using the PCR primers and these reactions were performed using the ABI PRISM BigDye Terminators v 3.0 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, U.S.A.) optimized for half- or quarter-size reactions. Sequencing products were purified using Sephadex G-50 (Amersham, Piscataway, New Jersey, U.S.A.) gel filters, and then separated by capillary electrophoresis using the ABI Prism 3100 Genetic Analyzer. Nucleotide sequences were edited using Sequencher 3.1.1 (Gene Codes Corporation, Ann Arbor, Michigan, U.S.A.), entered in PAUP* version 4.0b10 for Macintosh-PPC (Swofford, 2002), and manually aligned. Sequences were submitted to GenBank when reliable sequencing products were obtained (Table 3). Nucleotide sequences of the *rpoA* gene were translated to amino acid sequences using MacClade 4.05 (Maddison & Maddison, 2002).

Ancestral character-state reconstruction. — The loss of the *rpoA* gene from the chloroplast genome was reconstructed using both maximum parsimony (MP) and maximum likelihood (ML) as the optimality criterion based on the phylogenetic hypothesis proposed by Cox & al. (2004). Outgroup taxa (i.e., liverworts) were pruned from the tree. Under MP, transformations were reconstructed using PAUP (Swofford, 2002), with either the delayed transformation (DELTRAN) or accelerated transformation (ACCTRAN) option invoked. For ML reconstructions, *Andreaeobryum macrosporum* and *Alophozia azoricum* were also pruned from the tree, as no amplicon compatible with either the presence or absence of the *rpoA* gene was obtained, and because missing data prevent reconstructions under ML using the software Mesquite 1.0 (Maddison & Maddison, 2003). Branch lengths estimates under the General-Time-Reversible model of evolution (Rodriguez & al., 1990) with a proportion of invariable sites and a discrete gamma distribution of rates were taken from Cox & al. (2004). Rates of evolution for the loss of the *rpoA* gene were estimated under the simple Mk1 model (one rate) and the AsymmMk model (two rates, estimating the bias of gains versus losses; Pagel, 1999; Lewis, 2001). The difference in likelihoods was tested for significance based on the likelihood ratio test. Ancestral character-state reconstructions will be considered significant if their likelihood equals or exceeds 95% (Schluter & al., 1997).

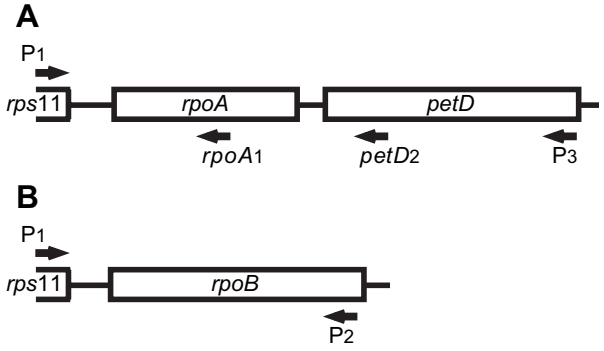

Fig. 1. Location of primers used to target the *rpoA* region of the chloroplast genome in mosses. A, typical gene arrangement in land plants; B, gene arrangement in *Physcomitrella*.

RESULTS

Amplification of the *rpoA* region using the *rpoA*1 primer yielded products for *Takakia*, *Sphagnum*, *Andreaea*, *Oedipodium*, *Polytrichum* and *Buxbaumia*. This amplicon comprised approximately 600 bp of the *rpoA* gene that could be fully aligned with that of *Marchantia* (Ohyama & al., 1986). The presence of the *rpoA* gene was confirmed for all these taxa except *Andreaea*, by amplifying the whole region using the reverse primer P3 or *petD*2 (Fig. 2). Although the *rpoA* gene could not be fully amplified for *Andreaea*, the sequence of the amplicon obtained with the reverse primer *rpoA*1, aligns with that of other taxa, and does not include any stop codons. In the case of *Polytrichadelphus*, the combination of P1 and *petD*2 yielded a product, but any attempt to sequence it with the reverse primer failed, and hence we only obtained a partial

Table 3. Size characterization of the *rpoA* region in the chloroplast genome of mosses (Funariales not included as gene arrangement differs from other mosses).

Taxon	*rps11-rpoA*	*rpoA*	*rpoA-petD*	*rps11-petD*	GenBank accession
Marchantia polymorpha	33	1023	111	-	NC001319
Sphagnum fimbriatum	26	1017	256	-	AY885232
Takakia lepidozioides	30	1149	204	-	AY884000
Andreaea rupestris[1]	30	Partial	-	-	AY886745
Oedipodium griffithianum	30	1023	171	-	AY886744
Dendroligotrichum dendroides	30	1023	165	-	AY886746
Polytrichadelphus purpureus[1]	-	Partial	161	-	AY911400
Polytrichum pallidisetum	30	1017	163	-	AY886747
Tetraphis geniculata	-	-	-	690	AY911406
Tetraphis pellucida	-	-	-	720	AY911407
Buxbaumia aphylla	30	1023	154	-	AY886743
Diphyscium foliosum	-	-	-	526	AY911401
Fissidens subbasilaris	-	-	-	267	AY911402
Scouleria aquatica	-	-	-	279	AY911403
Bartramia pomiformis[2]	-	-	-	280*	AY911408
Tetraplodon mnioides	-	-	-	197	AY914175
Orthotrichum lyellii	-	-	-	82	AY911405
Brachythecium salebrosum	-	-	-	309	AY911404

[1]Attempts to sequence the full P1-*petD*2 amplicon with the P1 primer failed, and only the 3' half of the product could be sequenced.
[2]Attempts to sequence the full P1-*petD*2 amplicon with the *petD*2 primer failed, and only the 5' half of the product could be sequenced.

sequence, which matches the 5' half of the *rpoA* gene of other mosses. The complete *rpoA* sequences obtained for other early diverging mosses varied between 1017 bps in *Polytrichum* and *Sphagnum*, and 1149 in *Takakia*, with most taxa having a *rpoA* gene that is 1023 bps long (Table 3). Closer examination revealed further cryptic length variation. For example, *Sphagnum* differs from *Marchantia* by four codon insertions and six deletions and *Polytrichum* by one insertion and three deletions. The sequence of *Sphagnum fimbriatum* obtained here aligns completely with that published by these authors for *S. girgensohnii* Russow (GenBank accession AB193122), except for three additional codons in the latter species. Aligning the sequence of the *rpoA* gene of *Takakia* against that of *Marchantia* requires a priori no insertion of gaps, but the gene is 126 bps longer in the former taxon. The sequence obtained here for *Takakia* is identical to that reported by Sugita & al. (2004) based on comparisons with GenBank accession AB193121. None of the *rpoA* sequences contained any internal stop codons.

Using the *rpoA*1 primer with P1 did not result in any amplicon for either species of *Tetraphis*, whereas P1 and P3 lead to a PCR product. The P1-P3 sequences obtained for *Tetraphis* and other mosses, could be aligned at both extremities with either the 3' end of the *rps11* gene or the 5' end of the *petD* gene, respectively. Furthermore, attempts to amplify the *rpoA* gene using two internal primers failed.

Representatives of the arthrodontous mosses, including *Diphyscium*, all tested negative for the *rpoA* gene (i.e., yielded no amplification with *rpoA*1 primer, or yielded a small fragment using the P3). The primer combination P1-P3 did not yield an amplicon for *Funaria*, *Entosthodon* and *Encalypta*, a result compatible with a rearrangement of genes as in *Physcomitrella*. Although amplicons were obtained for all other arthrodontous mosses screened, sequencing reactions using the reverse primer failed for several species (i.e., *Aulacomnium turgidum*, *Bartramia pomiformis*, *Hedwigia ciliata*, *Hookeria lucens*, *Mielichhoferia elongata*, *Mnium hornum*, *Orthodontium lineare*, *Pyrrhobryum vallis-gratiae*, *Rhodobryum giganteum*, and *Timmia megapolitana*). However, the sequence obtained for *Bartramia pomiformis* here matches that obtained by Sugiura & al. (2003) except for lacking the last 83 nucleotides of the spacer. The sequence flanked by the *rps11* and *petD* genes in the remaining arthrodontous mosses ranged between 82 in *Orthotrichum lyellii* and 526 bps in *Diphyscium foliosum*, respectively. This intergenic spacer is much shorter than that of *Tetraphis pellucida*, which is 720 bps long (Table 3). Targeting the whole *rpoA* gene using internal primers did not yield any product for any arthrodontous mosses.

Reconstructing the evolutionary history of the *rpoA* gene on the phylogeny presented by Cox & al. (2004) suggests that the gene was lost twice under DELTRAN optimization and once under ACCTRAN optimization, in which case a reversal took place in Buxbaumiidae (Fig. 2). On the same phylogeny, the likelihood (-logL = 5.84202197) for a two-rate model of evolution (i.e., AsymmMk) is higher than that (-log L = 7.422298653) of a one-rate model (i.e., Mk1) but not significantly (P = 0.08) under the likelihood ratio test. Hence, the one rate model cannot be rejected. Under this model, the nodes leading to *Tetraphis* and *Buxbaumia* are reconstructed with the *rpoA* gene most likely present (0.85% and 0.82%, respectively; Fig. 3), although neither value

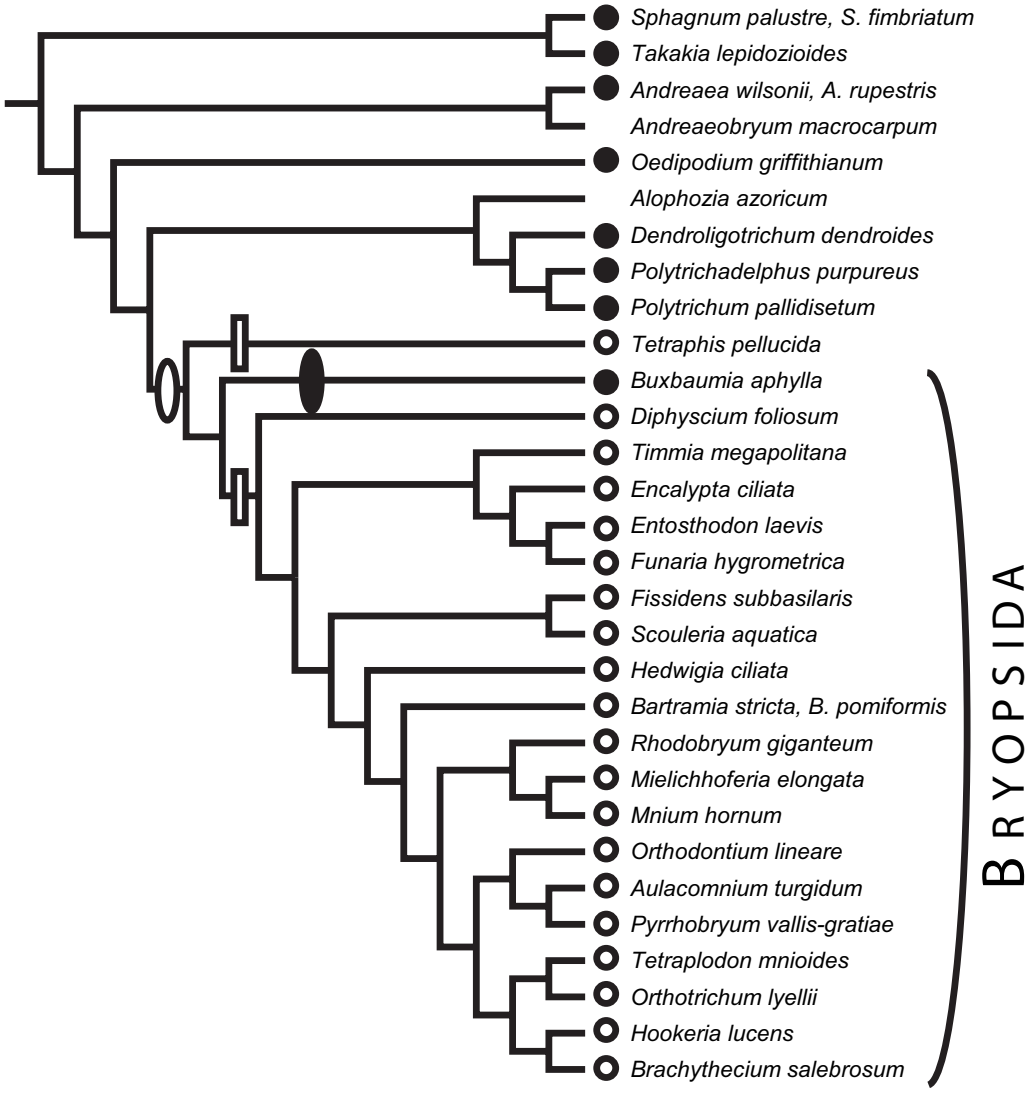

Fig. 2. Reconstruction under parsimony of the loss of the *rpoA* gene from the chloroplast genome during the diversification of major lineages of mosses based on the phylogenetic hypothesis proposed by Cox & al. (2004) with outgroup taxa pruned. The reference to two names at a terminal node indicates that the presence or loss of the *rpoA* gene was determined for an exemplar differing from the one used by Cox & al. (2004; see Table 1). Filled circles next to taxon indicate that the *rpoA* gene is present between the *rps11* and *petD* genes in the chloroplast genome; open circles mark taxa lacking the *rpoA* gene. Ancestral character states were reconstructed under maximum parsimony. Boxes identify the transformations under DELTRAN and ellipses under ACCTRAN with the open ellipse marking the loss and the filled one the gain.

should be considered significant. The node leading to *Diphyscium* is characterized by a likelihood of 0.90 for lacking the gene. The maximum likelihood reconstruction thus supports the hypothesis of two independent losses, whereby the *rpoA* gene was deleted from the chloroplast genome once in *Tetraphis* and once in the ancestor of *Diphyscium* and the arthrodontous mosses (Fig. 3).

DISCUSSION

The chloroplast genome of green plants typically comprises a gene encoding the alpha subunit of the RNA polymerase. In arthrodontous mosses, however, it is located in the nuclear genome, and this transfer had initially been interpreted as a synapomorphy for all mosses (Sugiura & al., 2003). Subsequent investigations revealed that this locus was, however, present in the plastid genome of two early diverging mosses, namely *Takakia* and *Sphagnum* (Sugita & al., 2004). Our survey

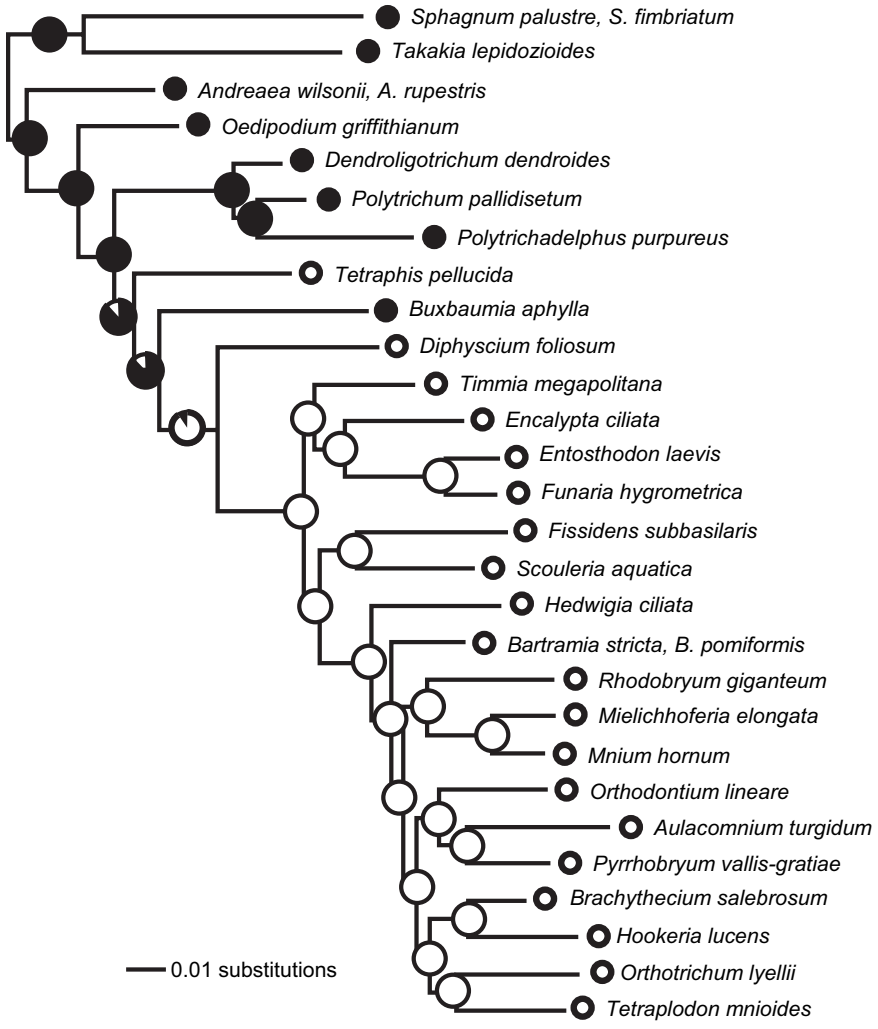

Fig. 3. Reconstruction under likelihood of the loss of the *rpoA* gene from the chloroplast genome during the diversification of major lineages of mosses based on the phylogenetic hypothesis proposed by Cox & al. (2004). Outgroups and exemplars for which no positive or negative amplicon could be obtained were pruned (see text). The reference to two names at a terminal node indicates that the presence or loss of the *rpoA* gene was determined for an exemplar differing from the one used by Cox & al. (2004; see Table 1). Maximum likelihood estimates of presence or absence of the gene in ancestral nodes under the Mk1 model are presented as proportions (white: absence, black: presence). Estimates for the nodes with *Tetraphis*, *Buxbaumia* and *Diphyscium* are 0.85, 0.82 and 0.90, respectively. Small circles next to taxon name indicate whether the *rpoA* gene is present (black) or absent (white) from the chloroplast genome.

further indicates that the *rpoA* gene is located in the chloroplast genome of other mosses, representing lineages that diverged after *Takakia* and *Sphagnum* but prior to the origin of true arthrodontous mosses. Exemplars of Andreaeopsida, Oedipodiopsida, Polytrichopsida, and Buxbaumiidae all share the ancestral condition, with the *rpoA* gene located at the end of the ribosomal protein cluster (Fukuzawaa & al, 1988), that is, between the *rps11* and the *petD* genes in the chloroplast genome. By contrast, the gene is lacking from the plastid genome in *Diphyscium* and all other arthrodontous mosses. This pattern is congruent with a shared ancestry between these two lineages as repeatedly hypothesized from phylogenetic inferences based on DNA sequence data (Goffinet & Buck, 2004).

The transfer of the *rpoA* gene from the chloroplast to the nuclear genome does not, however, solely diagnose the origin of the lineage sister to Buxbaumiidae. Indeed, both species of *Tetraphis* lack the *rpoA* gene between the *rps11* and the *petD* genes. The absence of the gene in the ancestral position can be explained either by a transfer to the nucleus, as in *Physcomitrella*, or by a relocation within the chloroplast genome. The *rpoA* gene present in the chloroplast could be amplified using two internal primers

for all basal mosses where the gene was located in a position homologous to *Marchantia* and *Anthoceros*. By contrast, all attempts to target the copy present in the nuclear genome failed. Similarly, no amplification products were obtained for *Tetraphis*. We interpret this pattern as evidence that the *rpoA* gene has not been relocated within the plastid genome of *Tetraphis* but has instead been transferred to the nucleus. Although we had expected to be able to amplify the nuclear homologue of the *rpoA* gene using the two internal primers, the failure to do so is not surprising considering that the nuclear copy sequenced from *Physcomitrella* exhibits only 50% similarity in amino acid sequence with *Marchantia* (Sugiura & al., 2003). Furthermore the nuclear homologue appears to be 330 bp longer. Hence, it is possible that the sequence of the *rpoA* has diverged significantly following its transfer to the nucleus. This hypothesis can, however, only be formally tested as nuclear homologues of more arthrodontous mosses are sequenced.

The loss of the *rpoA* gene from the chloroplast genome may have occurred once in the ancestor to the "*Tetraphis*-*Buxbaumia*-arthrodontous moss" clade and subsequently regained in *Buxbaumia* or twice, namely in the ancestor to *Tetraphis* and again in the ancestor to *Diphyscium* and the remaining Bryopsida. We cannot unambiguously distinguish these two scenarios. Moreover, the interpretation of two losses rests on the assumption that the phylogenetic topology used to reconstruct the history of transformations is correct. If *Tetraphis* is in fact the sister taxon to the *Diphyscium*-arthrodontous mosses clade, any reconstruction would favor a single loss of the gene. The phylogenetic position of *Buxbaumia* as sister to the clade anchored with *Diphyscium* is, however, well supported by analyses of eight nuclear, plastid and mitochondrial genes (i.e., bootstrap proportions under ML of 83, and posterior probabilities of 0.99 or 1; Cox & al., 2004). Hence, at least two transformations are needed to explain the distribution of the *rpoA* gene in the context of our best phylogenetic hypothesis for mosses.

A reversal, that is a secondary insertion of the *rpoA* gene in the chloroplast genome of *Buxbaumia* in a position homologous to the ancestral location (i.e., between *rps11* and *petD*) seems unlikely, although this hypothesis cannot *a priori* be rejected. The mechanics of gene transfer from the plastid to the nuclear genome are still unclear, but the process seems to take place by recombination in the sperm cell following degradation of the plastid prior to syngamy (Timmis & al., 2004), although RNA mediated transfer is also a possibility (Nugent & Palmer, 1991; Kobayashi & al., 1997). We have not found any hypothesis explaining how the reverse transfer could take place. At present, we favor a scenario of two independent losses, on the sole basis that a reversal seems unlikely. Note, however, that Kelch & al. (2004) explained the presence of the *ycf15* gene I in *Arabidopsis* as the result of a secondary gain, rather than as a retained ancestral character.

Parallel losses of genes from the chloroplast genome are not uncommon (Martin & al., 1998; Timmis & al., 2004), with at least one gene (i.e., *minE*) having been lost seven times during the evolution of the chloroplast genome (Martin & al., 2002). For two other genes, for which six losses are hypothesized, homologues have been detected in the nucleus (Martin & al., 2002). The diversification of angiosperms is characterized by at least 20 independent losses of the *infA* gene, and in four cases, expressed copies were detected in the nucleus (Millen & al., 2001). Hence, a dual loss of the *rpoA* gene from the chloroplast genome in mosses is not incongruent with patterns observed in other green plants.

ACKNOWLEDGEMENTS

This study was made possible by grants from the U.S. National Science Foundation through awards DEB-0089633 to BG and DEB-0089131 to AJS. We thank Harald Schneider for the valuable comments he provided on an earlier draft of this manuscript.

LITERATURE CITED

Adams, K. L. & Palmer, J. D. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molec. Phylogenet. Evol.* 29: 380–395.

Adams, K. L., Qiu Y.-L., Stoutemyer, M. & Palmer, J. D. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl. Acad. Sci. U.S.A.* 99: 9905–9912.

Cox, C. J., Goffinet, B., Shaw, A. J. & Boles, S. B. 2004. Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genomic compartments. *Syst. Bot.* 29: 234–250.

Doyle, J. J. & Doyle, J. L. 187. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11–15.

Dyall, S. D., Brown, M. T. & Johnson, P. J. 2004. Ancient invasions: from endosymbionts to organelles. *Science* 304: 253–257.

Fukuzawaa, H., Kohchia, T., Sanoa, T., Shiraia, H., Umesonob, K., Inokuchib, H., Ozekib, H. & Ohyama, K. 1988. Structure and organization of *Marchantia polymorpha* chloroplast genome III. Gene organization of the large single copy region from *rbcL* to *trnL* (CAU). *J. Molec. Biol.* 203: 333–351.

Goffinet, B., Bayer, R. J. & Vitt, D. H. 1998. Circumscription and phylogeny of the *Orthotrichales* (Bryopsida) inferred from *rbcL* sequence analyses. *Amer. J. Bot.* 85:

1324–1337.

Goffinet, B. & Buck, W. R. 2004. Systematics of the Bryophyta (mosses); from molecules to a revised classification. *Monogr. Syst. Bot. Missouri Bot. Gard.* 98: 205–209.

Goremykin, V., Hirsch-Ernst, K. I., Wölfl, S. & Hellwig, F. H. 2003. The chloroplast genome of the "basal" angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Pl. Syst. Evol.* 242: 119–135.

Hackett, J. D., Yoon, H. S., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Nosenko, T. & Bhattacharya, D. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. *Curr. Biol.* 14: 213–218.

Huang, C. Y., Ayliffe, M. A. & Timmis, N. J. 2002. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422: 72–76.

Kelch, D. G., Kriskell, A. & Mishler, B. D. 2004. Inferring phylogeny using genomic characters: a case study using land plant plastomes. *Monogr. Syst. Bot. Missouri Bot. Gard.* 98: 3–12.

Knoop, V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.* 46: 123–139.

Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A. & Ohyama, K. Interorganellar gene transfer in bryophytes: the functional *nad7* gene is nuclear encoded in *Marchantia polymorpha*. *Molec. Gen. Genet.* 256: 589–592.

Korpelainen, H. 2004. The evolutionary processes of mitochondrial and chloroplast genomes differ from those of nuclear genomes. *Naturwissenschaften* 91: 505–518.

Kugita, M., Kaneko, A., Yamamoto, Y., Takeya, Y., Matsumoto, T. & Yoshinaga, K. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucl. Acids Res.* 31: 716–721.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological characters. *Syst. Biol.* 50: 913–925.

Maddison, D. R. & Maddison, W. P. 2002. *MacClade 4: Analysis of Phylogeny and Character Evolution,* version 4.05. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P. & Maddison D. R. 2004. *Mesquite: A Modular System for Evolutionary Analysis,* version 1.05. http://mesquiteproject.org.

Martin, W., Rujan, T., Richly, E., Hansen, A., Cornelsen, S., Lins, T., Leister, D., Stoebe, B., Hasegawa, M. & Penny, D. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 99: 12246–12251.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. & Kowallik, K. V. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393: 162–165.

Millen, R. S., Olmstead, R. G., Adams, K. L., Palmer, J. D., Lao, N. T., Heggie, L., Kavanagh, T. A., Hibberd, J. M., Gray J. C., Morden, C.W., Calie, P. J., Jermiin L. S. & Wolfe, K. H. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Pl. Cell* 13: 645–658.

Nakazano, M. & Hirai, A. 1993. Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. *Molec. Gen. Genet.* 236: 341–346.

Nugent, J. M. & Palmer, J. D. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66: 473–481.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umezono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.-I., Inokuchi, H. & Ozeki, H. 1986. Chloroplast gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322: 572–574.

Pagel, M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* 48: 612–622.

Richly, E. & Leister, D. 2004a. NUMTs in sequenced eukaryotic genomes. *Molec. Biol. Evol.* 21: 1081–1084.

Richly, E. & Leister, D. 2004b. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Molec. Biol. Evol.* 21: 1972–1980.

Rodriguez, F., Oliver, J. L., Marin, A. & Medina, J. R. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142: 485–501.

Sandoval, P., León, G., Gómez, I., Carmona, R., Figueroa, P., Houigue, L., Araya, A. & Jordana, X. 2003. Transfer of *rps14* and *rpl15* from the mitochondrion to the nucleus in grasses. *Gene* 324: 139–147.

Schluter, D., Price, T., Mooers, A. Ø. & Ludwig, D. 1997. Likelihoods of ancestor states in adaptive radiations. *Evolution* 51: 1699–1711.

Sugita, M., Sugiura, C., Arikawa, T. & Higuchi, M. 2004. Molecular evidence of an *rpoA* gene in the basal moss chloroplast genomes: *rpoA* is a useful molecular marker for phylogenetic analysis of mosses. *Hikobia* 14: 171–175.

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C. & Sugita, M. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella* patens: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. *Nucl. Acids Res.* 31: 5324–5331.

Swofford, D. L. 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods),* version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* 5: 123–135.

Turmel, M., Otis, C. & Lemieux, C. 1999. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: insights into the architecture of ancestral chloroplast genomes. *Proc. Natl. Acad. Sci. U.S.A.* 96: 10248–10253.

Zheng, D., Nielsen, B. L. & Daniell, H. 1997. A 7.5-kbp region of the maize (T cytoplasm) mitochondrial genome contains a chloroplast-like *trnI*(CAT) pseudogene and many short segments homologous to chloroplast and other known genes. *Curr. Genet.* 32: 125–131.