Die mitochondriale DNA basaler Tracheophyten: Molekulare Evolution komplexer Genomstrukturen

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Felix Grewe
aus
Bonn

Bonn 2011

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Volker Knoop
2. Gutachter: Prof. Dr. Dietmar Quandt

Tag der Promotion: 16. Mai 2011
Erscheinungsjahr: 2011

Erklärung

Ich versichere, dass ich diese Arbeit selbständig verfasst, keine anderen als die angegebenen Hilfsmittel verwendet und Zitate kenntlich gemacht habe. Diese Dissertation wurde in gleicher oder ähnlicher Form in keinem anderen Prüfungsverfahren vorgelegt.

Bonn, 26. Mai 2011

Felix Grewe

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Zusammenfassung

Aufgrund ihrer hohen Komplexität ist die Anzahl vollständig sequenzierter pflanzlicher mitochondrialer Genome (Chondrome) noch gering. Sie beschränkt sich auf die vollständig ermittelten mtDNA-Sequenzen einiger Angiospermen und einzelner Arten der Leber-, Laub- und Hornmoose und lediglich einer Art der Gymnospermen. Aus den basalen Gefäßpflanzengruppen (Tracheophyten), also den Bärlappgewächsen (Lycophyten) und den Farnen (Monilophyten), lagen bislang sogar noch gar keine vollständigen mtDNA-Sequenzen vor. Die Evolution der vielen molekularen Besonderheiten in den Chondromen der Landpflanzen war somit nur lückenhaft nachvollziehbar. Im Rahmen dieser Arbeit wurde diese phylogenetische Lücke durch die vollständige mtDNA-Sequenzierung des Lycophyten *Isoetes engelmannii* begleitet durch eine umfassende Analyse seines mitochondrialen Transkriptoms und eine weitgehende Untersuchung der mtDNA des Monilophyten *Gleichenia dicarpa* gefüllt. Darüber hinaus wurde im Rahmen einer begleitenden Diplomarbeit das komplette mitochondriale Genom und Transkriptom des Lycophyten *Selaginella moellendorffii* ermittelt. Das Chondrom von *I. engelmannii* zeigt einen individuellen Trend zu einer kompakten genomischen Struktur mit extrem kleinen Introns und intergenischen Regionen, wohingegen die mtDNA des Schwestertaxons *S. moellendorffii* völlig gegensätzlich mit großen intergenischen Bereichen und Introns ausgestattet ist. In beiden Lycophyten allerdings wurden hoch rekombinante mitochondriale Genome gefunden. Diese führen zur Entstehung neuer mitochondrialer Eigenheiten, beispielsweise neuer *trans*-spleißender Introns - auch Gruppe I Introns in beiden Lycophyten - und dem Einbau "promisker" DNA aus dem Zellkern oder den Chloroplasten. Anhand der vollständigen mitochondrialen Transkriptomanalysen konnten in *I. engelmannii* über 1700 und in *S. moellendorffii* sogar über 2000 *RNA Editing*-Positionen identifiziert werden. Während in *S. moellendorffii* ausschließlich Cytidin-zu-Uridin-Konversionen zu finden sind, handelt es sich bei 20% der *Editing*-Ereignisse in *I. engelmannii* um Austausche in Gegenrichtung von Uridin zu Cytidin. Als weitere Besonderheit

wurde massives *RNA Editing* in *I. engelmannii* nicht nur in mRNAs sondern auch in tRNAs gefunden, während in der mtDNA von *S. moellendorffii* als völligem Novum für pflanzliche Chondrome alle Gene für tRNAs verloren gegangen sind. Die chondromale Komplexität ist offenbar in den Monilophyten noch weiter gesteigert. Große Regionen der mtDNA von *G. dicarpa* bestehen aus integrierten mobilen Elementen, u.a. acht "promisker" Retrotransposons. Diese wurden offenbar als kollaterales Ereignis massiver Transpositionen im Kerngenom auch in die mtDNA eingebaut.

2 Einleitung

2.1 Besonderheiten mitochondrialer DNA

Mitochondrien sind von einer Doppelmembran umschlossene Organellen. Ihre wichtigste Funktion ist die Energieversorgung der Zelle durch die Bereitstellung chemischer Energie in Form von ATP. Sie sind unter anderem durch eine eigene Erbsubstanz, die mitochondriale DNA (mtDNA) charakterisiert. Die mtDNA wird auf das Genom eines α-Proteobakteriums zurückgeführt, das von einem gemeinsamen Vorfahren aller eukaryotischen Zellen durch Endozytose aufgenommen wurde und sich danach zu den Mitochondrien entwickelte [Gray *et al.* 1999].

In basalen Eukaryoten, den Protisten, sind im Laufe der Evolution völlig unterschiedliche mitochondriale Genome entstanden [Gray *et al.* 2004]: Der Krankheitserreger von Malaria, *Plasmodium falciparum*, besitzt eine massiv reduzierte mtDNA von nur 6 kb, die gerade einmal fünf Gene trägt [Vaidya *et al.* 1989]. Dagegen beherbergt *Reclinomonas americana* auf einer 69 kb großen mtDNA fast 100 Gene. Diese mtDNA stellt unter allen bisher sequenzierten mtDNAs die ursprünglichste Form dar [Lang *et al.* 1997]. Weitere Studien zeigten, dass der Erreger von Fleckfieber, das α-Proteobakterium *Rickettsia prowazekii*, näher als alle anderen bisher untersuchten Mikroorganismen mit der ursprünglichen mtDNA von *R. americana* verwandt ist, sodass alle Mitochondrien möglicherweise von einem gemeinsamen Vorfahren abstammen [Andersson *et al.* 1998].

Tierische mtDNAs unterscheiden sich im Gegensatz zu den mtDNAs der Protisten nur gering in Struktur und Größe. Die mitochondrialen Genome der meisten Tiere (Metazoa) sind auf kompakte, zirkuläre DNA Moleküle mit einer Größe von 16 kb reduziert, die in den meisten Fällen dicht gepackt 13 proteinkodierende Gene sowie 22 tRNA- und 2 rRNA-Gene tragen [Gissi *et al.* 2008, Lavrov 2007].

Die mtDNA der Landpflanzen (Embryophyta) ist hingegen deutlich größer und kann von 105 kb im Laubmoos *Physcomitrella patens* [Terasawa *et al.* 2007] bis

über 2000 kb in Kürbisgewächsen betragen [Ward *et al.* 1981]. Dies ist aber bei weitem nicht die einzige außergewöhnliche Eigenschaft, die pflanzliche mtDNA auszeichnet. Weitere Besonderheiten wie (rezente) Auslagerungen von Genen ins Kerngenom, (aktive) Rekombinationen, Integrationen von fremder Kern- und Chloroplasten-DNA, Organellen-spezifische Gruppe I und II Introns, *RNA Editing* und horizontaler Gentransfer (HGT) gestalten die mitochondriale DNA der Pflanzen zu einem äußerst interessanten Thema aktueller Forschung.

2.2 Mitochondriale Genomgrößen der Landpflanzen

Während bereits hunderte der untereinander strukturell sehr ähnlichen mtDNAs von Tieren sequenziert wurden, ist die Anzahl sequenzierter pflanzlicher mitochondrialer Genome immer noch sehr gering. Die komplexe Struktur und Größenzunahme der pflanzlichen mtDNAs erschwert eine Sequenzierung, wodurch sich die Anzahl der vollständig sequenzierten Landpflanzen-mtDNAs bisher auf 25 beschränkt.

Es dauerte elf Jahre nach der ersten vollständigen Sequenzierung der menschlichen mtDNA [Anderson *et al.* 1981] bis im Jahr 1992 erstmals eine pflanzliche mtDNA des Lebermooses *Marchantia polymorpha* komplett sequenziert wurde. Mit einer Größe von 187 kb ist diese fast zwölf mal größer als die menschliche mtDNA [Oda *et al.* 1992]. Es dauerte danach weitere fünf Jahre bis mit der 367 kb großen mtDNA der Modellpflanze *Arabidopsis thaliana* ein zweites pflanzliches mitochondriales Genom sequenziert wurde, das neben einer deutlichen Größenzunahme auch mehr Komplexität aufweist [Unseld *et al.* 1997]. Mittlerweile wurden die mtDNA-Sequenzen von hauptsächlich wirtschaftlich interessanten 19 Angiospermen sowie einer Gymnosperme, zwei Hornmoosen, einem Laubmoos und zwei Lebermoosen ergänzt.

Aus den eudikotylen Angiospermen wurden neben *A. thaliana* die mtDNA zweier Brassicales sequenziert, eine Brassicaceae, *Brassica napus* (Raps), und eine Caricaceae, *Carica papaya* (Papaya). Weitere Arten stammen aus den Ordnungen der Caryophyllales, *Beta vulgaris* (Rübe) und *Silene latifolia* (Weiße Lichtnelke), Cucurbitales, *Citrullus lanatus* (Wassermelone) und *Cucurbita pepo* (Garten-Kürbis), Solanales, *Nicotiana tabacum* (Tabak) und Vitales, *Vitis vinifera* (Wein) (siehe Tabelle 2.1).

Alle sequenzierten Arten der Monokotyledonen stammen aus der Ordnung der Poales (Gräser): *Oryza sativa* (Reis), *Triticum aestivum* (Weizen), *Sorghum bicolor* (Mohrenhirse), *Tripsacum dactyloides* (Gamagras) und *Zea mays* (Mais). Sequenzierungen weiterer Reis- und Mais-Arten, *O. rufipogon*, *Z. luxurians* und *Z. perennis*, dienten Vergleichsstudien nah verwandter Arten. Zudem wurden *cytoplasmic male sterile* (CMS) Linien, z.B. des Mais [Allen *et al.* 2007], sequenziert, die von erheblicher agronomischer Bedeutung zur Herstellung von Hybridsaatgut sind.

Bei einer durchschnittlichen mtDNA-Größe der Angiospermen von 515 kb, stellt die mtDNA von *C. pepo* mit einer Größe von 983 kb eine deutliche Ausnahme dar (siehe Tabelle 2.1). Die mtDNA eines anderen Kürbisgewächs ist sogar noch größer: *Cucumis melo* wurde über mitochondriale Genomkartierungen die bisher größte bekannte mitochondriale Genomgröße von ca. 2,9 Mb nachgewiesen [Ward *et al.* 1981]. Die Vergrößerung der mtDNA geht allerdings nicht mit einem Anstieg kodierender Sequenzteile einher, sondern mit einer Zunahme an intergenischen Regionen. Durch eine Auslagerung mitochondrialer Gene in das Kerngenom (siehe Abschnitt 2.4) werden die kodierenden Sequenzen auf der mtDNA sogar noch reduziert.

Die Angiospermen bilden zusammen mit den vier Gruppen der Gymnospermen (Coniferophyten, Ginkgophyten, Gnetophyten und Cycadophyten) die Samenpflanzen (Spermatophyta). Aus den Gymnospermen wurde bisher erst die mtDNA einer Art vollständig sequenziert. Dabei handelt es sich um *Cycas taitungensis* aus der Familie der Cycadaceae mit einer Genomgröße von 415 kb.

Gegenüber den Samenpflanzen fallen die sequenzierten mtDNAs der Bryophyten mit einer Größe von durchschnittlich 171 kb deutlich kleiner aus. Bryophyten sind die ältesten Landpflanzen bestehend aus den paraphyletischen Gruppen Leber-, Laub-, und Hornmoose. 14 Jahre nach der mtDNA-Sequenzierung von *M. polymorpha* wurde 2007 die vollständige Laubmoos-mtDNA der Modellpflanze *P. patens* sequenziert. Um über die drei Gruppen der Bryophyten weitere Informationen zu erlangen und einen Vergleich ihrer mtDNAs zu ermöglichen, wurden in den darauf folgenden Jahren die mtDNAs des Lebermooses *Pleurozia purpurea* und zweier Hornmoose, *Megaceros aenigmaticus* und *Phaeoceros laevis*, vollständig sequenziert.

Tabelle 2.1: Sequenzierte mitochondriale Genome der Landpflanzen

Landpflanzengruppe **mtDNA-seq. Taxa**	**Accession** **[RefSeq]**	**Größe** **[nt]**	**Referenzen**
Angiospermen - Eudikotyledonen			
Arabidopsis thaliana	NC_001284	366924	Unseld *et al.* [1997]
Brassica napus	NC_008285	221853	Handa [2003]
Carica papaya	NC_012116	476890	nicht publiziert
Cucurbita pepo	NC_014050	982833	Alverson *et al.* [2010]
Citrullus lanatus	NC_014043	379236	Alverson *et al.* [2010]
Vitis vinifera	NC_012119	773279	Goremykin *et al.* [2009]
Silene latifolia	NC_014487	253413	Sloan *et al.* [2010a]
Beta vulgaris subsp. vulgaris	NC_002511	368801	Kubo *et al.* [2000]
Nicotiana tabacum	NC_006581	430597	Sugiyama *et al.* [2005]
Angiospermen - Monokotyledonen			
Zea mays subsp. mays	NC_007982	569630	Clifton *et al.* [2004]
Zea mays subsp. parviglumis	NC_008332	680603	Allen *et al.* [2007]
Zea luxurians	NC_008333	539368	nicht publiziert
Zea perennis	NC_008331	570354	nicht publiziert
Triticum aestivum	NC_007579	452528	Ogihara *et al.* [2005]
Sorghum bicolor	NC_008360	468628	nicht publiziert
Tripsacum dactyloides	NC_008362	704100	nicht publiziert
Oryza rufipogon	NC_013816	559045	Fujii *et al.* [2010]
Oryza sativa Indica Group	NC_007886	491515	Tian *et al.* [2006]
Oryza sativa Japonica Group	NC_011033	490520	Notsu *et al.* [2002]
Gymnospermen			
Cycas taitungensis	NC_010303	414903	Chaw *et al.* [2008]
Monilophyten			
-mtDNA Sequenzierung im Rahmen dieser Arbeit-			
Lycophyten			
-mtDNA Sequenzierung im Rahmen dieser Arbeit-			
Hornmoose (Bryophyten)			
Megaceros aenigmaticus	NC_012651	184908	Li *et al.* [2009]
Phaeoceros laevis	NC_013765	209482	Xue *et al.* [2010]
Laubmoose (Bryophyten)			
Physcomitrella patens	NC_007945	105340	Terasawa *et al.* [2007]

Tabelle 2.1: (Fortsetzung)

Landpflanzengruppe mtDNA-seq. Taxa	**Accession [RefSeq]**	**Größe [nt]**	**Referenzen**
Lebermoose (Bryophyten)			
Pleurozia purpurea	NC_013444	168526	Wang *et al.* [2009]
Marchantia polymorpha	NC_001660	186609	Oda *et al.* [1992]

Daten aus http://www.ncbi.nlm.nih.gov/genomes/

Aus den Landpflanzengruppen der Leber-, Laub- und Hornmoose, Gymnospermen und Angiospermen wurde bisher die mtDNA von mindestens einem repräsentativen Taxons sequenziert (siehe Tabelle 2.1). Hingegen lagen aus den basalen Gefäßpflanzengruppen (Tracheophyten), den Lycophyten und Monilophyten (Bärlappgewächse und Farne, siehe auch Abschnitt 2.8), noch gar keine vollständigen mtDNA Sequenzen vor. Eine 20 kb große Region, sequenziert von der mtDNA des Nestfarns *Asplenium nidus*, bildet bisher die größte zusammenhängende Sequenz einer Monilophyten-mtDNA [Panarese *et al.* 2008]. Die Evolution mitochondrialer Genome in Landpflanzen war durch die fehlenden Sequenzen der Lycophyten und Monilophyten somit bisher nur lückenhaft nachvollziehbar.

2.3 Komplexe Chondromstrukturen

Eine vollständige Sequenzierung ermöglicht unter anderem tiefe Einblicke in die molekulare Beschaffenheit und Struktur der mtDNAs.

Aufgrund des bakteriellen Ursprungs und der Ergebnisse aus den Sequenzierungen tierischer mtDNA wurde auch die mtDNA der Pflanzen stets als ein in sich geschlossenes Molekül angesehen. Die Sequenz der ersten ersten pflanzlichen mtDNA von *Marchantia polymorpha* bestätigten diese Annahme [Oda *et al.* 1992] und auch die mtDNAs aller weiteren sequenzierten Bryophyten gaben keinen eindeutigen Hinweis auf eine komplexere Struktur und werden deshalb ringförmig dargestellt [Li *et al.* 2009, Terasawa *et al.* 2007, Wang *et al.* 2009, Xue *et al.* 2010].

Es bestehen jedoch auch Annahmen, dass diese Struktur aufgrund des jeweiligen experimentellen Aufbaus, den nachträglichen computerisierten Verknüpfungen von Teilsequenzen, nur fälschlich entstand. Tatsächlich soll sich die Gesamtheit der

mtDNA (das Chondrom) aus vielen, nebeneinander existierenden, verzweigten, linearen Molekülen zusammensetzen [Bendich 2007, Oldenburg und Bendich 1998, 2001].

Aufgrund aktiver Rekombinationen innerhalb der mtDNAs der höheren Pflanzen entstehen deutlich komplexere Chondromstrukturen [Übersicht in Knoop 2004]. An Sequenzwiederholungen (*repeats*) assoziiert, werden kleinere, ringförmige Rekombinationsprodukte (*subcircle*) abgeschnürt, die wiederum mit anderen *subcirclen* fusionieren können, sodass vielfältige, koexistierende Sequenzarrangements entstehen. Die oft ringförmig dargestellte mtDNA repräsentiert in diesen Fällen einen sogenannten Standardring (*mastercircle*) [Lonsdale *et al.* 1988], der alle molekularen Subgenome zu einem hypothetischen Chondrom grafisch vereint. Es wird allerdings bezweifelt, ob rekombinante Angiospermen-Chondrome tatsächlich in einer solchen *mastercircle*-Struktur aufzufinden sind.

Erste Erkenntnisse über diese unterteilten Chondrome als Konsequenz aus homologen Rekombinationen erhielt man über physikalische Genomkartierungen [Fauron *et al.* 1995, Palmer und Shields 1984]. Spätere vollständige Sequenzierungen der mtDNAs bestätigten diese Studien [Ogihara *et al.* 2005, Sugiyama *et al.* 2005].

Die einzige sequenzierte mtDNA einer Gymnospermen-Art (*Cycas taitungensis*), besitzt viele kurze Sequenzwiederholungen, sogenannte *Bpu sequences*, von denen jedoch eine Beteiligung an homologen Rekombinationen aufgrund ihrer Länge ausgeschlossen wird [Chaw *et al.* 2008].

Ohne weitere vollständige Sequenzen aus den Gruppen der Lycophyten und Monilophyten konnte keine Aussage über den evolutiven Übergang von den homogenen mtDNAs der Bryophyten zu den rekombinanten Chondromen der Angiospermen getroffen werden. Zwar konnten den Monilophyten bereits große mtDNAs und eine sehr hohe Komplexität über Genomkartierungen nachgewiesen werden [Palmer *et al.* 1992], es gab darüber hinaus allerdings nur sehr wenige Berichte über die mtDNA Strukturen dieser Landpflanzengruppen.

2.4 Organisation mitochondrialer Gene

Nach der Aufnahme eines α-Proteobakteriums vor etwa zwei Milliarden Jahren gingen viele Gene aus dem Genom des prokaryotischen Symbionten verloren, andere wurden in das nukleäre Genom ausgelagert. Aus dem Genom des endosymbiontischen Bakteriums entstand die reduzierte mtDNA der Mitochondrien.

Alle verbleibenden Gene der mtDNAs kodieren direkt oder indirekt ausschließlich für Produkte, die in die ATP Produktion über oxidative Phosphorylierung involviert sind. Die mtDNAs der Tiere besitzen nur noch einen extrem reduzierten Satz an Genen und kodieren in der Regel 22 tRNAs, zwei rRNAs und 13 proteinkodierende Gene, Untereinheiten der Atmungskettenkomplexe I, III, IV und V (NADH Ubichinon Oxidoreduktase: *nad*1, *nad*2, *nad*4, *nad*4L, *nad*5, *nad*6, Apocytochrom b: *cob*, Cytochrom c Oxidase: *cox*1, *cox*2, *cox*3 und ATP-Synthase: *atp*6, *atp*9).

Unter den Landpflanzen hat *M. polymorpha* mit 73 Genen die genreichste mtDNA. Von diesen kodieren 41 für Proteine, 29 für tRNAs und drei für rRNAs. Das Genrepertoire des Lebermooses beinhaltet weitere Gene für Untereinheiten des Komplexes I (*nad*-Gene), Komplexes V (*atp*-Gene) sowie des Komplexes II (Succinatdehydrogenase: *sdh*3 und *sdh*4) [Oda *et al.* 1992]. Darüber hinaus liegen auf dieser mtDNA Gene, die für Proteine der Cytochrom c Biogenese (*ccb*-Gene), der kleinen und großen ribosomalen Untereinheiten (*rps*- und *rpl*-Gene) und einer Untereinheit des Tat-Transporters (*tat*C) kodieren [Übersicht in Knoop 2004].

Diese unterscheiden aber nicht nur tierische und pflanzliche mtDNA, sondern sind auch innerhalb der Landpflanzen-Chondrome sehr unterschiedlich vorhanden. Es besteht der Trend zu einer Verringerung der auf dem Chondrom kodierten genetischen Information durch eine Auslagerung von mitochondrialen Genen in den Zellkern (*endosymbiotic gene transfer, EGT*). Ausgedehnte *Southern Blot* Analysen von 280 Angiospermen verdeutlichten, dass die Anzahl ausgelagerter Gene selbst zwischen nah verwandten Arten stark variieren kann [Adams *et al.* 2002].

Abbildung 2.1 zeigt den mitochondrialen Genverlust innerhalb der Landpflanzen. Die Anzahl der verlorenen Gene variiert innerhalb der Landpflanzengruppen, wobei in den Hornmoosen und den Caryophyllales (*B. vulgaris* und *S. latifolia*) ein massiver Genverlust stattfand [Kubo *et al.* 2000, Li *et al.* 2009, Sloan *et al.* 2010a, Xue *et al.* 2010]. Im Laufe der Landpflanzenevolution verringert sich mit diesem stetigen Verlust von mitochondrial kodierter Information aber nicht die Größe der Chondrome, statt dessen nehmen diese – teilweise sogar massiv – an Größe zu (siehe Abschnitt 2.2).

In wenigen Fällen sind auch Gene aus Komponenten der Atmungskette über funktionalen EGT in den Kern ausgelagert worden: *cox*2 der Fabales [Nugent und Palmer 1991], *atp*8 von *Allium spec.* [Adams *et al.* 2002], *atp*8 und *nad*7 der

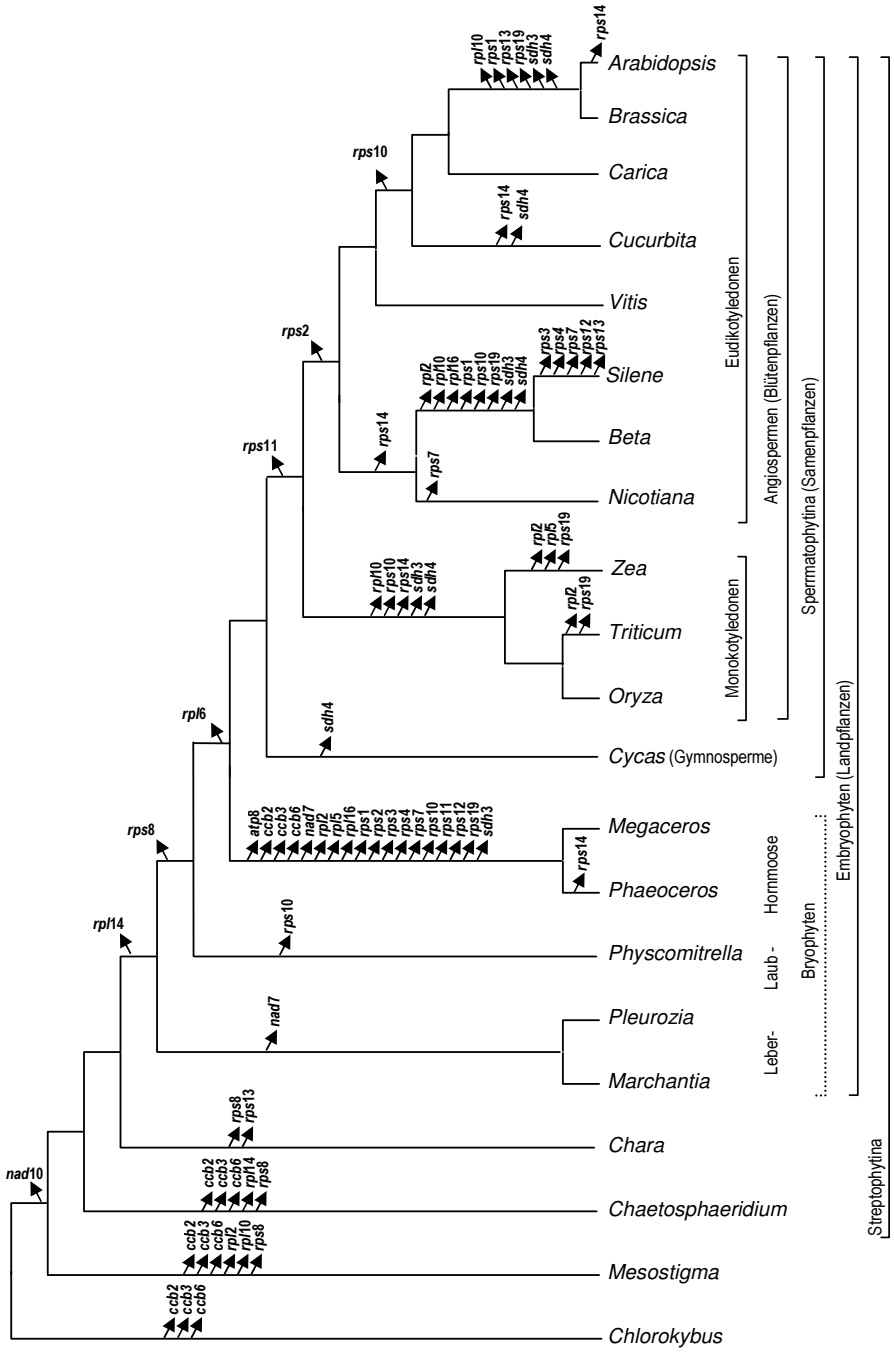


Abbildung 2.1: Mitochondrialer Genverlust innerhalb der Landpflanzen. Dargestellt ist die Verwandtschaft aller vollständig sequenzierten mtDNAs der Landpflanzen und vier Grünalgen (*Chlorokybus atmophyticus* [Turmel *et al.* 2007], *Mesostigma viride* [Turmel *et al.* 2002b], *Chaetosphaeridium globosum* [Turmel *et al.* 2002a] und *Chara vulgaris* [Turmel *et al.* 2003]) nach neusten Erkenntnissen der Pflanzenphylogenie [Qiu *et al.* 2006]. An den Ästen des Kladogramms zeigen Pfeile den Verlust der entsprechenden Gene an. Die Arten *C. lanatus*, *T. aestivum* und *S. bicolor* wurden aus Gründen der Übersicht nicht aufgelistet, da ihr Gengehalt dem ihrer nahen Verwandten gleicht. [modifiziert nach Knoop *et al.* 2011]

Hornmoose [Li *et al.* 2009, Xue *et al.* 2010] und *nad*7 der Lebermoose [Groth-Malonek *et al.* 2007, Kobayashi *et al.* 1997]. Interessanterweise ist *nad*7 immer noch als Pseudogen in den Lebermoosen enthalten, obwohl selbst nach rezentem EGT innerhalb der Angiospermen die mitochondriale Genkopie schnell degeneriert und meist nicht mehr auf der mtDNA vorzufinden ist. Die im Kern lokalisierten Gene weisen eine größere Ähnlichkeit mit den post-transkritionell bearbeiteten Transkripten als mit den mitochondrial lokalisierten Genen anderer Arten auf. Offensichtlich sind die Gene beim EGT über ein RNA-Intermediat aus dem Chondrom in den Zellkern transferiert worden [Brennicke *et al.* 1993].

Komplexe der Atmungskette bestehen also meist aus Produkten nukleärer und (noch) auf dem Chondrom kodierter Gene. Sie sind dann zum einen aufgebaut aus Protein-Untereinheiten, die im Mitochondrium gebildet werden, und zum anderen aus Protein-Untereinheiten, die im Zellkern kodiert und im Cytoplasma synthetisiert ins Organell eingeschleust werden. Die vorhandenen tRNAs, rRNAs und Gene ribosomaler Proteine kodieren auf die der mtDNA verbliebenen Teile des ehemaligen prokaryotischen Translationssystems. Dessen vollständige Funktion ist ebenfalls nur durch ergänzende Zellkern-kodierte Faktoren, u.a. ribosomale Proteinuntereinheiten, auf der mtDNA fehlenden tRNAs und Aminoacyl-tRNA-Synthetasen, gewährleistet.

2.5 Integration fremder DNA

Neben der Auslagerung von mitochondrialen Genen in das Kerngenom über EGT, werden auch DNA-Segmente aus der Kern- und der Plastiden-DNA (cpDNA) in das mitochondriale Genom integriert. Im Zusammenhang mit der ersten Entdeckung einer 12 kb großen Region chloroplastidärer DNA im Mais-Chondrom [Stern und Lonsdale 1982] wurden derartige fremde Fragmente "promiske" DNA benannt (*promiscuous DNA*) [Ellis 1982]. Diese Art der "promisken" DNA ist in allen Angiospermen vorzufinden und wurde zudem auch in der Gymnospermen-mtDNA von *C. taitungensis* identifiziert [Wang *et al.* 2007], dagegen bisher in keiner Bryophyten-mtDNA.

In der Regel sind die transferierten Fragmente inaktiv, selten kommt es jedoch auch zu einem funktionellen Transfer von Chloroplasten-tRNAs oder einer Promotorsequenz [Joyce und Gray 1989, Nakazono *et al.* 1996]. Die tRNA-Population mancher Mitochondrien besteht damit aus den Produkten von drei unterschiedlich entstandenen Genklassen: Die erste Klasse sind vom Endosymbionten abstammende,

ursprünglich mitochondriale tRNAs. Darüber hinaus gibt es die Chloroplasten-ähnlichen tRNAs, die zwar auch mitochondrial kodiert vorliegen, jedoch von einem funktionalen Gentransfer aus dem Chloroplasten-Genom abstammen. Die dritte Klasse bilden Kern-kodierte tRNAs, die aus dem Cytoplasma assoziiert an Aminoacyl-tRNA-Synthetasen, importiert werden und die Funktion eines verlorenen paralogen Gens ausüben [Duchêne *et al.* 2011, 2009, Glover *et al.* 2001].

Neben der Integration von plastidärer DNA, wurden auch mitochondrial lokalisierte Ty1-*copia*-, Ty3-*gypsy*- und non-LTR-Retrotransposon-Fragmente sowie Transposon-Fragmente mit offensichtlicher nukleärer Herkunft identifiziert [Knoop *et al.* 1996, Notsu *et al.* 2002].

2.6 Gruppe I und Gruppe II Introns

In der mitochondrialen DNA findet man besondere Typen von Introns, die Gruppe I und Gruppe II Introns. 1981 wurde das erste mitochondrial kodierte Intron innerhalb der Angiospermen im *cox*2 Gen des Mais entdeckt [Fox und Leaver 1981]. Inzwischen sind über 100 verschiedene Loci mitochondrialer Introns in Landpflanzen bekannt [Übersicht in Bonen 2008, Knoop 2004]. Beide Introntypen sind in der Natur weit verbreitet und wurden zudem in der mtDNA von Pilzen und Protisten, in der cpDNA und auch in Prokaryoten (Bakterien und Archaeen) vorkommend beschrieben [Haugen *et al.* 2005, Simon *et al.* 2008].

Die Typisierung der Introns beruht auf charakteristischen Sekundärstrukturen, die durch spezifische Basenpaarungen der RNAs zustande kommen [Michel *et al.* 1989, Übersicht in Cech *et al.* 1994]. Die Introns beider Gruppen werden ohne den Einfluss eines Spleißosoms (*spliceosome*) aus dem Primärtranskript (prä-RNA) entfernt. Manchen Introns konnte *in vitro* ein sogenanntes "Selbst-spleißen" nachgewiesen werden. Diese entfernten sich selbst durch ihre katalytische Aktivität der RNA ohne Anwesenheit von Proteinen [Kruger *et al.* 1982, Übersicht in Michel und Ferat 1995]. Ein solches "Selbst-spleißen" konnte für pflanzliche Introns aber bisher nicht gezeigt werden. Zum Spleißen der Introns werden *in vivo* Proteinfaktoren benötigt, die in zwei Gruppen eingeteilt werden können: Intron-kodierte und Kern-kodierte Faktoren.

Einige Introns enthalten einen offenen Leserahmen (*open reading frames, ORF*), der für Proteine kodiert, die die Mobilität oder das Spleißen der Introns beeinflussen [Haugen *et al.* 2005, Robart und Zimmerly 2005]. Gruppe I Introns kodieren

Endonukleasen (*homing endonuclease*) und Gruppe II Introns kodieren für Reverse Transkriptasen, wodurch die Introns in andere Orte des Genoms oder sogar in andere Genome migrieren können. Beide Intron ORFs kodieren zudem für RNA Maturasen, die Introns aus der prä-RNA schneiden können. Damit bringt das Intron bei einer Migration in einen neuen kodierenden Locus gleich alle Voraussetzungen mit, sodass dessen Funktion danach nicht unterbunden ist [Bonen 2011].

Intron-ORFs sind innerhalb von Prokaryoten, Pilzen und auch den Bryophyten verbreitet, gingen aber in der Evolution zu den höheren Pflanzen weitgehend verloren. Das Fehlen dieser für die Mobilität der Introns verantwortlichen Proteine, erklärt womöglich auch die stabilen Intronpositionen innerhalb der Samenpflanzen [Knoop *et al.* 2011]. Mehreren kernkodierten Proteinen konnte bereits eine Beteiligung am Spleißen der ORF-freien Introns nachgewiesen werden [Übersicht in Bonen 2011, Knoop *et al.* 2011]. Beispielsweise wird das Produkt eines in den Kern ausgelagerten Maturase-Gens für das Spleißen mehrerer Introns in das Mitochondrium von *A. thaliana* reimportiert [Keren *et al.* 2009].

Innerhalb der Landpflanzen liegen in den rekombinanten Chondromen der Angiospermen manche Gruppe II Introns in *trans*-konfigurierter Form vor. Dabei ist ein Gen innerhalb der Intronsequenz unterbrochen und liegt in zwei voneinander getrennten Teilen auf der Organellen-DNA vor. Erst durch einen *trans*-Spleißvorgang werden die Exons des Gens miteinander verbunden, wodurch eine aus mehreren Transkripten zusammenhängende mRNA entsteht. Die Gene *nad*1, *nad*2 und *nad*5 der Samenpflanzen werden von bis zu sechs *trans*-konfigurierten Introns unterbrochen [Binder *et al.* 1992, Knoop *et al.* 1991, Wissinger *et al.* 1991]. Besonders interessant ist deren evolutive Entstehung, denn den *trans*-gespleißten Introns der Samenpflanzen konnten entsprechende *cis*-gespleißte Vorläufer in Laub- und Hornmoosen, Lycophyten und Farnen nachgewiesen werden [Groth-Malonek *et al.* 2005, Malek und Knoop 1998]. Auch außerhalb der *nad*-Gene wurde kürzlich ein *trans*-gespleißtes Intron im *cox*2 Gen der Zwiebel entdeckt [Kim und Yoon 2010]. Auch von diesem Intron cox2i373g2 gibt es *cis*-gespleißte Varianten in den Laub- und Hornmoosen sowie in manchen Samenpflanzen. Eine umfassende Analyse eines *trans*-gespleißten Introns der Angiospermen zeigte 15 unabhängige Wechsel von der *cis*- zur *trans*-gespleißten Form [Qiu und Palmer 2004].

2.7 *RNA Editing* der Landpflanzen

Eine post-transkriptionelle Reifung beinhaltet oft auch verschiedene Sequenzänderungen durch *RNA Editing* [Knoop 2010]. In den meisten Mitochondrien (und auch den Chloroplasten) der Landpflanzen ist die dominierende Form des *RNA Editings* der Austausch spezifischer Cytidin-Nukleotide in Uridin-Nukleotide (C-zu-U *Editing*) sowie – jedoch viel seltener – eine Änderung in umgekehrter Richtung (U-zu-C *Editing*, *reverse editing*). Auch mehr als zwanzig Jahre nach der Entdeckung [Covello und Gray 1989, Gualberto *et al.* 1989, Hiesel *et al.* 1989] ist der Grund für diesen Prozess weiterhin umstritten. Es gibt bisher keine Indizien für einen funktionellen Nutzen, zum Beispiel einer physiologischen Regulation oder aber der Herstellung unterschiedlicher Proteine [Knoop 2010, Mower und Palmer 2006]. Somit wird die Hauptaufgabe des *RNA Editings* der Organellen wahrscheinlich darin bestehen, evolutiv konservierte Sequenzen wiederherzustellen, damit bei einer anschließenden Translation intakte Proteine entstehen. Aber auch die Pyrimidine von Introns [Carrillo *et al.* 2001] und tRNAs [Marchfelder *et al.* 1996] werden durch *RNA Editing* geändert, sodass diese ihre charakteristische Sekundärstruktur ausbilden.

Durch cDNA Sequenzdaten mitochondrialer Transkripte wurde *RNA Editing* in allen untersuchten Landpflanzengruppen mit Ausnahme der thallösen marchantiiden Lebermoose, zum Beispiel *M. polymorpha*, nachgewiesen [Hiesel *et al.* 1994, Malek *et al.* 1996, Steinhauser *et al.* 1999]. Es wird daher angenommen, dass diese Fähigkeit in einem gemeinsamen Vorfahren aller Landpflanzen entstand, jedoch auch einmal sekundär verloren ging [Groth-Malonek *et al.* 2007]. Die Anzahl der *Editing*-Positionen der vollständig analysierten mitochondrialen Transkriptome ist sehr unterschiedlich. Im Laubmoos *P. patens* wurden nur elf mitochondriale *Editing* Positionen identifiziert [Rüdinger *et al.* 2009], wohingegen die Anzahl der sechs weiteren analysierten Transkriptome – alle aus der Gruppe der Angiospermen – zwischen 189 in *Silene noctiflora* und 491 in *O. sativa* variiert [Handa 2003, Mower und Palmer 2006, Notsu *et al.* 2002, Picardi *et al.* 2010, Sloan *et al.* 2010b]. Kürzlich wurden in 25 von 39 proteinkodierenden Genen der Gymnosperme *C. taitungensis* 565 *Editing*-Positionen gezählt – das bisher höchste identifizierte Vorkommen an *Editing* [Salmans *et al.* 2010]. Doch ohne vollständige Transkriptomdaten aus den Landpflanzengruppen der Lebermoose, Hornmoose, Lycophyten, Monilophyten und Gymnospermen besteht weiterhin eine große Lücke, die eine umfassende Analyse des Landpflanzen-*RNA Editings*

erschwert. Hinzu kommt, dass in den Mitochondrien der bisher vollständig analysierten Pflanzen kein U-zu-C *Editing* vorkommt und darüber bisher auch nur spärliche Informationen verfügbar sind.

Die Mechanismen des *RNA Editings* sind bisher erst lückenhaft geklärt. Eine eindeutige Beteiligung konnte den RNA-bindenden PPR-Proteinen zuerst in den Chloroplasten [Kotera *et al.* 2005], später auch in den Mitochondrien [Zehrmann *et al.* 2009] nachgewiesen werden. Einige Typen dieser Proteine besitzen eine DYW-Domäne, die aufgrund ihrer Ähnlichkeit zu Cytidindeaminasen als mögliche katalytische Einheit für das *RNA Editing* gesehen wird [Salone *et al.* 2007]. Unterstützt wurde diese Annahme durch die Entdeckung, dass ein Vorkommen von PPR-Proteinen des DYW-Typs mit dem Vorkommen von *RNA Editing* korreliert [Rüdinger *et al.* 2008]. Bemerkenswerterweise wurden diese Proteine kürzlich auch im amöboiden Einzeller *Naegleria gruberi* entdeckt; danach kann *Editing* auch im Mitochondrium dieses Protisten stark vermutet werden [Knoop und Rüdinger 2010].

Der umgekehrte *Editing* Prozess (*reverse editing*) wurde als erstes für jeweils eine Position im *cob*- und *cox*3-Gen zweier Angiospermen beschrieben [Gualberto *et al.* 1990, Schuster *et al.* 1990]. Wo hingegen bei keiner späteren Angiospermen-Transkriptomanalyse weiteres U-zu-C *Editing* gesehen wurde, identifizierte man vier heterologe Positionen in den *cox*3-Transkripten zweier Monilophyten [Hiesel *et al.* 1994]. Darauf folgende umfassendere Untersuchungen von Regionen des *cox*3- und *nad*5-Gens bestätigten das Vorkommen von U-zu-C *Editing* in weiteren Monilophyten und zudem auch in Lycophyten und Hornmoosen [Groth-Malonek *et al.* 2005, Malek *et al.* 1996, Steinhauser *et al.* 1999]. In den basalen Landpflanzengruppen, den Leber- und Laubmoosen, gibt es dagegen kein U-zu-C *Editing*. Dies deutet darauf hin, dass das U-zu-C *Editing* möglicherweise in einem gemeinsamen Vorfahren der Tracheophyten und Hornmoose entstanden ist und in den Samenpflanzen größtenteils wieder verloren ging.

Das *RNA Editing* arbeitet an unterschiedlichen Positionen verschieden effizient. Bei einem Vergleich des *Editings* in mehreren Transkriptkopien liegen manche von ihnen, die sogenannten partiellen *Editing*-Positionen, nur teilweise in ihrer geänderten Form vor [Mower und Palmer 2006]. In diesem Zusammenhang wurde beobachtet, dass "stille" *Editing* Positionen, deren Pyrimidinänderung später keinen Einfluss auf die Aminosäuresequenz hat, weniger effizient geändert werden als Positionen, die eine Änderung der Proteinübersetzung zur Folge haben.

2.8 *Isoetes engelmannii*, *Selaginella moellendorffii* und *Gleichenia dicarpa*

Phylogenetische Studien beweisen, dass die Bärlappgewächse (Lycophyten) eine monophyletische Schwestergruppe aller anderen Gefäßpflanzen (Tracheophyten) bilden. Diese setzen sich wiederum aus zwei Gruppen zusammen, den Monilophyten (Farne) und den Samenpflanzen (Angiospermen und Gymnospermen) [Qiu *et al.* 2007]. In der Entwicklung der Landpflanzen bilden die Lycophyten ein Intermediat zwischen den ersten Landpflanzen, also den Leber-, Laub- und Hornmoosen, und den Monilophyten.

Fossile Funde aus dem späten Silur und frühen Devon bestätigen, dass die Lycophyten die älteste Tracheophytengruppe bilden [Gensel und Berry 2001]. Die Erscheinungsformen der meisten rezenten Arten ist im Laufe von hunderten Millionen Jahren im Wesentlichen unverändert geblieben [Bresinsky *et al.* 2008]. Ihre größte Vielfalt erreichten sie vor etwa 310 Millionen Jahren in den kohlebildenden Sumpfwäldern des späten Karbon [Phillips und DiMichele 1992]. Eine Karbonisierung ihrer Überreste bildet die heutigen Kohlevorkommen der Welt.

In der Gruppe der Bärlappgewächse stellen die heterosporen Isoetales und Selaginellales zusammen eine Schwestergruppe der isosporen Lycopodiales dar [Qiu *et al.* 2007].

In der Ordnung Isoetales bildet *Isoetes* die einzige rezente Gattung in einer ansonsten ausgestorbenen Entwicklungslinie der Bärlappgewächse. Zwei südamerikanische terrestrische Arten, die früher als eigene Gattung *Stylites* geführt wurden, werden heute der Gattung *Isoetes* zugeordnet [Bresinsky *et al.* 2008]. Die Gattung *Isoetes* ist heute nur noch durch ungefähr 150 Arten vertreten [Taylor und Hickey 1992], die fast weltweit vorkommen. Deren Verbreitungsschwerpunkte liegen in Nordamerika und in Mitteleuropa. Die meisten *Isoetes*-Arten sind submers oder saisonal aquatisch. Sie leben in einer Wassertiefe bis zu zehn Metern oder besiedeln Randzonen klarer Teiche und Seen. Es gibt aber auch terrestrische Arten, die recht trockene Standorte besiedeln können. Die Erscheinungsform von *Isoetes* unterscheidet sich sehr von der anderer Lycophyten (Bärlappgewächse). Nach der Beschreibung im Strasburger entspringen der Achse "aus 2-3 Längsfurchen Reihen von gabelig verzweigten Wurzeln und oben lange pfriemenförmige, eine Rosette bildende Blätter (bei bestimmten Arten bis zu 1 m lang!). Die von vier Luftkanälen

durchzogenen Blätter sind auf der Oberseite ihres verbreiterten Grundes mit einer länglichen, grubenartigen Vertiefung ("Fovea") versehen. Die meisten Blätter sind Sporophylle mit je einem Sporangium in der Fovea. [...] An den äußeren Blättern der Rosette bilden sich Megasporangien mit zahlreichen Megasporen, an den nach innen folgenden jüngeren Blättern Mikrosporangien mit jeweils sehr vielen Mikrosporen."[Bresinsky *et al.* 2008]

Isoetes engelmannii ist eine aquatische *Isoetes*-Art und kommt im Osten Nordamerikas vor. Wie alle *Isoetes*-Arten reagiert sie besonders anfällig auf Störungen der Umwelt und gilt daher als vom Aussterben bedroht (MNFI Rare Species Explorer: http://web4.msue.msu.edu/mnfi/explorer/species.cfm?id=15925).

Die Ordnung der ebenfalls heterosporen Selaginellales beinhaltet *Selaginella* als einzige Gattung. Die etwa 700 Arten leben überwiegend als Bodendecker in feuchten Tropenwäldern. Wenige Arten sind allerdings auch frosttolerant oder an trockene Standorte angepasst. Die Erscheinungsform von *Selaginella* ist durch verschiedenartige Sprosse geprägt. Im Strasburger wird hierzu beschrieben, dass "die Sprossachse [...] mit kleinen, zunächst schraubig, später meist dekussiert stehenden schuppenartigen Blättern in vier Zeilen dorsiventral besetzt [ist], und zwar meist mit zwei Reihen kleiner so genannter Oberblätter und zwei Reihen diesen gegenüberstehenden größeren Unterblättern [...]. An den Gablungsstellen der Sprosse entstehen bei vielen Arten [...] abwärts wachsende [...] Wurzelträger (Rhizophore), an deren freien Ende endogen Büschel von Wurzeln entspringen [...]." Die Sporangien sind zu endständigen Sporophyllständen vereint, wobei jedem Sporophyll ein Sporangium mit Megasporen oder Mikrosporen entspringt [Bresinsky *et al.* 2008].

Selaginella moellendorffii besitzt das kleinste bekannte Pflanzengenom von etwa 110 Mb Größe [Wang *et al.* 2005] und wurde vor kurzem vom Energy Joint Genome Institute (JGI) für eine vollständige Sequenzierung ausgewählt [Banks 2009].

Farne werden nach neuesten molekular-phylogenetischen Erkenntnissen mit den Psilotophytina und Equisetophytina in die Landpflanzengruppe der Monilophyten eingeordnet [Pryer *et al.* 2001]. Monilophyten treten in den unterschiedlichsten Erscheinungsformen auf und bilden eine Schwestergruppe der Samenpflanzen innerhalb der Euphyllophyten (alle Landpflanzen außer den Bryophyten und Lycophyten). Die Phylogenie innerhalb der Gruppe der Monilophyten ist noch nicht detailliert aufgeklärt. Daher bleibt die exakte Position der Equisetophytina,

das Verwandtschaftsverhältnis der Baum- und Wasserfarne sowie der Hymenophyllaceae, Matoniaceae und Gleicheniaceae unklar [Pryer *et al.* 2004].

Die Gleicheniaceae sind – wie die meisten Monilophyten – isospor. Die Familie umfasst sechs Gattungen (*Dicranopteris*, *Diplopterygium*, *Gleichenella*, *Gleichenia*, *Sticherus*, *Stromatopteris*) mit etwa 125 Arten und bildet innerhalb der leptosporangiaten Monilophyten eine monophyletische Gruppe [Hasebe *et al.* 1995, Pryer *et al.* 2001, 2004, 1995]. Gleicheniaceae besitzen oft geteilte Megaphylle (Wedel), die in der Jugend an der Spitze eingerollt sind. Die Sori befinden sich abaxial mit jeweils 5-15 Sporangien, die wiederum 120 bis 800 Sporen tragen [Smith *et al.* 2006].

Die Art *Gleichenia dicarpa* wächst bei hoher Luftfeuchtigkeit und ist in den Tropen und Subtropen Australiens und Neuseelands verbreitet. Die Fieder wachsen nicht flächig, wie es von anderen Farnarten bekannt ist, sondern bilden eine feste, schmale Struktur.

3 Zielsetzung der Promotionsarbeit

Die mtDNA der Landpflanzen umfasst ein großes Repertoire besonderer Charakteristika. Beinahe aus jedem vollständig sequenzierten Chondrome lassen sich neue interessante Rückschlüsse auf die Evolution dieser Besonderheiten sowie auf die Entwicklung dieser Organellen-DNA ziehen. Bis heute wurde die mtDNA von fünf Bryophyten und 20 Samenpflanzen sequenziert, es gibt jedoch keine vollständige Sequenzinformation aus den Landpflanzengruppen der Lycophyten und Monilophyten. Hinsichtlich vollständiger Transkriptomanalysen liegen außer der des Laubmooses *P. patens* und von sechs Angiospermen noch keine Daten vor. Die Evolution der mitochondrialen post-transkriptionellen Eigenheiten der Landpflanzen ist daher noch nicht kontinuierlich nachvollziehbar.

Ziel der vorliegenden Dissertation war es daher, durch vollständige mtDNA-Sequenzierungen der Lycophyten *Isoetes engelmannii*, *Selaginella moellendorffii* und umfassender Analysen der mtDNA des Monilophyten *Gleichenia dicarpa* diese phylogenetische Lücke zu schließen. Aufgrund der hohen Komplexität dieser mtDNAs wurde im Vorfeld eine sortierte Fosmidbank angelegt, für *S. moellendorffii* wurde unter Anwendung dieser Methode eine *next-generation* Sequenzierung vervollständigt. Mit diesen Sequenzinformationen als Basis soll der evolutive Weg der Chondrome von den homogenen Molekülen der Bryophyten zu den extrem variablen Strukturen der Samenpflanzen und die Entwicklung der besonderen Charakteristika pflanzlicher mtDNA nachvollzogen werden. Zudem sollen vollständige Transkriptomanalysen die Evolution des *RNA Editing* und des Spleißens innerhalb der Lycophyten- und Monilophyten-mtDNA eingehender beleuchten.

4 Ergebnisse und Diskussion

4.1 Die mitochondriale DNA von *Isoetes engelmannii*

Die mitochondriale DNA (mtDNA) der Lycophytengattung *Isoetes* war schon des öfteren Mittelpunkt molekularbiologischer Evolutionsstudien. Denn neben dem besonderen Auftreten und Habitus der Pflanze (siehe Abschnitt 2.8), konnten *Isoetes* auch im molekularen Bereich viele Besonderheiten nachgewiesen werden.

In frühen Studien über unterschiedliche *RNA Editing* Frequenzen in Landpflanzen zeigte *Isoetes lacustris* bereits ein außergewöhnlich hohes Aufkommen an *RNA Editing* [Malek *et al.* 1996]. Eine weitere Kuriosität wurde bei Untersuchungen *cis*-arrangierter Vorfahren *trans*-gespleißter Introns der Angiospermen offenkundig [Malek und Knoop 1998]. In diesem Zusammenhang entdeckte man das bis dahin kleinste mitochondriale Gruppe II Intron im *nad*2 Gen von *I. lacustris*, das damals schon ein Hinweis auf eine extrem kompakte Chondromstruktur gab.

Diese und viele weitere Besonderheiten, die bei Analysen partieller mtDNA Sequenzen von *Isoetes* auffielen, begründeten das Vorhaben einer vollständigen Sequenzierung. Eine *shotgun*-Sequenzierung aus isolierten Mitochondrien schied aufgrund der Seltenheit des Lycophyten und keiner Möglichkeit zu einer Vermehrung durch Kultivierung aus. Erste Versuche unternahm man daher über eine der Sequenzierung vorangehenden Cosmid-Klonierung von Gesamt-DNA [Karolin Kalmbach, Yesim Kümetepe, Julia Neuwirt und Patrick Johner, Universität Ulm, Diplomarbeiten]. Diese wurde jedoch aufgrund einer großen Instabilität der Cosmid-Konstrukte nicht weiter fortgeführt. Mit der Anwendung eines stabileren Fosmid-Klonierungssystems konnte dieses Problem dann überwunden werden. Aufgrund der extrem komplexen Chondromstruktur von *I. engelmannii* stellte sich die Anwendung dieser Methode zudem als besonders vorteilhaft heraus. Eine Zusammensetzung von Teilsequenzen zu einer Gesamtsequenz (*Sequenzassembly*),

stellte bereits für die einzelnen Fosmid-*Inserts* eine große Herausforderung dar – ein Zusammensetzen von *shotgun*-Sequenzteilen wäre wahrscheinlich ganz unmöglich gewesen.

Für die Fosmid-Klonierung wurde Pflanzenmaterial der Art *I. engelmannii* verwendet. Die Pflanze wurde im US-Bundesstaat Indiana von Jerry Gastony gesammelt und anschließend im Gewächshaus kultiviert. Für unser Projekt wurde das Material freundlicherweise von Dr. Jeff Palmer und Dr. Erin Badenhop (Bloomington, IN) zur Verfügung gestellt.

Ergebnisse aus Fragmentierungen mitochondrialer Fosmide mittels Endonukleasen und Teile einer ersten Fosmid-Sequenzierung deuteten auf bemerkenswerte Eigenheiten der mtDNA von *I. engelmannii* hin: Viele Rekombinationen führen zu koexistierenden Genen, Genfragmenten und Pseudogenen, die sich jeweils in verschiedenen Genarrangements befinden. Darüber hinaus wurden extreme *RNA-Editing* Frequenzen vorhergesagt. Zwei Gene konnten vollständig und ein weiteres, das *cox*1 Gen, fast vollständig sequenziert werden. Zu einer Vervollständigung fehlten stets Bereiche des cox1i1305g1 Gruppe I Introns [Felix Grewe, Universität Bonn, Diplomarbeit].

In weiterführenden Analysen wurde später erkenntlich, dass es gar nicht möglich war dieses *cox*1 Gen innerhalb der DNA Sequenz zu schließen. Es wurde gezeigt, dass das Intron cox1i1305g1 in einer besonderen trans-gespleißten Konfiguration vorliegt, die für Gruppe I Introns damit erstmals beobachtet wurde. Doch dies blieb nicht die einzige neu entdeckte Besonderheit, denn durch die Sequenzierung weiterer Fosmide bis hin zur Fertigstellung der ersten mitochondrialen Gesamtsequenz eines Lycophyten wurden noch viele weitere Eigenheiten erkannt, die in der Arbeit *A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii* in *Nucleic Acids Research* veröffentlicht worden sind.

Published online 23 June 2009

Nucleic Acids Research, 2009, Vol. 37, No. 15 **5093–5104**
doi:10.1093/nar/gkp532

A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*

Felix Grewe[1], Prisca Viehoever[2], Bernd Weisshaar[2] and Volker Knoop[1,*]

[1]Institut für Zelluläre und Molekulare Botanik (IZMB), Universität Bonn, Kirschallee 1, 53115 Bonn and [2]Institut für Genomforschung und Systembiologie (IGS), Universität Bielefeld, Universitätsstraße 25, 33594 Bielefeld, Germany

Received March 19, 2009; Revised June 5, 2009; Accepted June 6, 2009

ABSTRACT

Plant mitochondrial genomes show much more evolutionary plasticity than those of animals. We analysed the first mitochondrial DNA (mtDNA) of a lycophyte, the quillwort *Isoetes engelmannii*, which is separated from seed plants by more than 350 million years of evolution. The *Isoetes* mtDNA is particularly rich in recombination events, and chloroplast as well as nuclear DNA inserts document the incorporation of foreign sequences already in this most ancestral vascular plant lineage. On the other hand, particularly small group II introns and short intergenic regions reveal a tendency of evolution towards a compact mitochondrial genome. RNA editing reaches extreme levels exceeding 100 pyrimidine exchanges in individual mRNAs and, hitherto unobserved in such frequency, also in tRNAs with 18 C-to-U conversions in the tRNA for proline. In total, some 1500 sites of RNA editing can be expected for the *Isoetes* mitochondrial transcriptome. As a unique molecular novelty, the *Isoetes cox*1 gene requires *trans*-splicing via a discontinuous group I intron demonstrating disrupted, but functional, RNAs for yet another class of natural ribozymes.

INTRODUCTION

Mitochondrial DNAs (mtDNAs) trace back in evolution to the genome of an α-proteobacterial endosymbiont which gave rise to the mitochondria of eukaryotic cells (1). The mitochondrial genomes in most animal (metazoa) lineages are compact, circular DNAs of some 16 kb which encode a standard set of 37 or fewer tightly packed genes (2). The mtDNAs of other eukaryotes, however, are significantly more diversified, most notably between different protist lineages, which reflect most of the evolutionary history and diversity of eukaryotic cells (3). These, for example, include obvious evolutionary ancestral states such as the gene-rich 69-kb mtDNA of the jakobid protist *Reclinomonas americana* with nearly 100 mitochondrial genes (4) as well as the massively reduced 6-kb mtDNA of the malaria parasite *Plasmodium falciparum* with only five genes (5), reflecting a massive gene transfer into the nuclear genome.

Land plant (embryophyte) mtDNAs in contrast are significantly extended in size and may exceed 2000 kb in certain flowering plant (angiosperm) families (6). The embryophyte mtDNAs encode some of the genes for protein subunits of the respiratory chain complexes, for ribosomal proteins and for proteins involved in cytochrome *c* biogenesis which are found in protists but are generally absent from animal or fungal mitochondrial genomes. Many plant mitochondrial genes are interrupted by introns belonging to either of the two classes of ribozyme-type group I or group II introns, which are commonly encountered in fungal, algal and plant organelle genomes and occasionally also in bacteria, phages and exceptionally also in the mtDNAs of primitive metazoan lineages (2,7). Besides intron gains, size increases of plant mitochondrial genomes have occurred mainly through the extension of non-coding intergenic regions. This becomes immediately apparent when the available mtDNAs of land plants (Figure 1) are compared to those of the charophyte algae (8,9) phylogenetically related to the embryophyte lineage (10). Some of the additional sequences, at least in flowering plant (angiosperm) mtDNA, have been identified as copies of chloroplast or nuclear DNA (11,12) or even as gained via horizontal gene transfer (13–16).

Moreover, plant mitochondria have evolved complex features with respect to genome arrangements and gene expression that are contrary to the general evolutionary trend of compaction and streamlining of endosymbiotic genomes (17,18). The gain of RNA editing activity in the organelles (19,20) to correct gene sequences by

*To whom correspondence should be addressed. Tel: +49 228 73 6466; Fax: +49 228 73 6467; Email: volker.knoop@uni-bonn.de

5094 Nucleic Acids Research, 2009, Vol. 37, No. 15

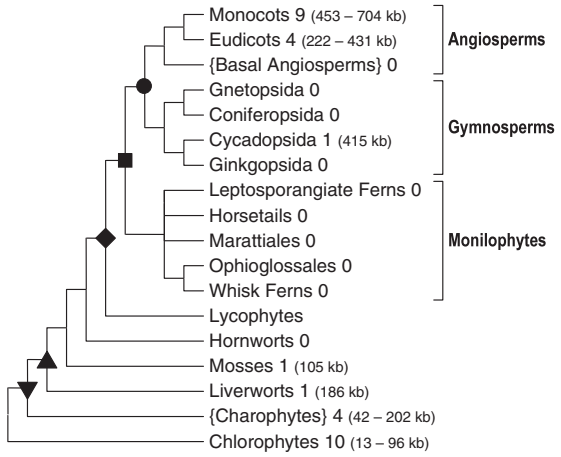


Figure 1. Current view of a simplified phylogeny of extant Viridiplantae (green plants *sensu lato*). The cladogram shown summarizes insights from recent molecular studies of land plant phylogeny (e.g. Qiu *et al.*, 2006). Numbers of completely sequenced mtDNAs (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/plants_tax.html) are indicated for each group. Brackets indicate paraphyletic grades, all other designations indicate reasonably well supported monophyletic groups. Further well supported monophyletic clades of higher order are the spermatophytes (seed plants, circle), the euphyllophytes (square), the tracheophytes (vascular plants, rhomb), the embryophytes (land plants, up triangle) and the streptophytes (down triangle) whereas the bryophytes (liverworts, mosses and hornworts) are paraphyletic.

pyrimidine exchanges (mainly cytidine to uridine) at the transcript level likewise appears to be gained with the earliest embryophytes, although this phenomenon is suspiciously absent in the subclass of marchantiid liverworts (21). Despite size increase to more than 100 kb in early embryophyte evolution (Figure 1), the two so far available completely sequenced mtDNAs of bryophytes—those of the liverwort *Marchantia polymorpha* (22) and the moss *Physcomitrella patens* (23)—are recognized as simple, circular-mapping genomes. However, linear DNAs may in fact contribute significantly to the population of mtDNA molecules actually present in the mitochondria (24). Flowering plant (angiosperm) mtDNAs are rich in active recombination resulting in co-existing alternative mitochondrial genome arrangements (e.g. 25–27), the stoichiometries of which are now understood to be regulated by nuclear-encoded protein factors related to bacterial *rec* proteins (28). A 'master-circle' representing the full mitochondrial genome complexity in a single circular DNA molecule may be entirely hypothetical in these cases (29).

Evidently coinciding with the rise of recombinational activity during the evolution of plant mitochondrial genomes is the appearance of *trans*-splicing group II introns producing peculiar arrangements of the affected genes with exons distributed across wide distances in the mtDNA. The origins of *trans*-splicing group II introns have been traced back through plant evolution (Figure 1) as having arisen through disruption of ancestral, conventional group II introns that can still be identified as their respective orthologues in ferns, hornworts and mosses (30,31).

In the absence of complete mtDNA information for ferns, horsetails, lycophytes or hornworts (Figure 1), there is currently a large phylogenetic gap remaining between the available mtDNA sequences of the liverwort *Marchantia* or the moss *Physcomitrella* and the first recently completed mtDNA of a gymnosperm, the cycad *Cycas taitungensis* (32). Accordingly, we have investigated the mtDNA of the quillwort *Isoetes engelmannii*. As a lycophyte, *Isoetes* represents the most ancestral lineage of recent vascular plants (tracheophytes). The mtDNA of *I. engelmannii* offers a plethora of surprising findings, which include particularly small group II introns, extreme frequencies of DNA recombination and RNA editing also in tRNAs, insertions of chloroplast and nuclear DNA and, most notably, a *trans*-splicing group I intron.

MATERIALS AND METHODS

Fosmid analyses

Isoetes engelmannii plant material originally collected in South Central Indiana (USA) by Jerry Gastony, and subsequently greenhouse cultivated, was kindly made available through Jeff Palmer and Erin Badenhop (Bloomington, IN). The non-green bulb tissue of plants was used to enrich for mitochondrial vs. chloroplast DNA. Total genomic DNA was isolated using a CTAB protocol. After size-fractionation into ~38 kb fragments, DNA was blunt-ended and cloned into the fosmid vector pCC1FOS using the CopyControl Fosmid Library Production Kit (EPICENTRE, Madison, Wisconsin). A library of 11 700 fosmid clones was sorted and filter-spotted for successive rounds of hybridization initially using a mixture of PCR-derived gene probes of *cox*3, *nad*2, *nad*5 and *nad*7 and subsequently with probes derived from the sequence-verified mitochondrial fosmids. Identity of fosmid clones was initially verified through terminal insert sequencing and positive clones were used for sub-library production. Fosmid DNAs were isolated using NucleoBond Xtra Midi EF Kit (Macherey Nagel, Düren, Germany), sheared by Nebulizers (Invitrogen, Carlsbad, California), blunted using a End-It DNA End-Repair Kit (EPICENTRE, Madison, Wisconsin), A-tailed with Taq-Polymerase (Genaxxon, Biberach, Germany), and fractionated by preparative electrophoresis in 0.8% agarose. Fragments of 2–2.5 kb in size were recovered using the NucleoSpin Extract II Kit (Macherey Nagel, Düren, Germany) and cloned into pGEM-T Easy vector (Promega, Madison, Wisconsin). Minimally 400 plasmid clones were sequenced for each fosmid to reach ~8-fold sequencing coverage. Five fosmid clones (11P20, 19N12, 26A6, 28M14 and 30K18) were validated as native mtDNA. Graphical maps of the fosmid clones created with OGDRAW v1.1 (33) are given in Supplementary Figure 1. The respective fosmid insert sequences were annotated and deposited in the database under accession numbers FJ010859, FJ536259, FJ390841, FJ176330 and FJ628360, respectively.

Nucleic Acids Research, 2009, Vol. 37, No. 15 5095

Sequence analyses

Sequence handling and analysis of final fosmid assemblies was essentially done using the alignment explorer of the MEGA software (34). Identification of loci was essentially done using similarity searches with Basic Local Alignment Search Tool (BLAST) service at the NCBI (35). Candidate sites of RNA editing in *Isoetes* were identified manually in alignments of deduced protein sequences with homologues in *Chara* (AY267353) and *Marchantia* (M68 929), species devoid of RNA editing.

Trancript analyses

Total *I. engelmannii* RNA was isolated using the NucleoSpin RNA Plant Kit (Macherey Nagel, Düren, Germany); cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas, Burlington, Ontario) in the presence of random hexamer primers as recommended by the manufacturer. Oligonucleotide pairs (all sequence information available from the authors upon request) were used for RT–PCR amplification according to the standard protocol of GoTaq DNA Polymerase (Promega, Madison, Wisconsin) in a GeneAmp PCR System 2700 (Applied Biosystems, Foster City, California) with annealing temperatures between 50°C and 55°C. Amplicons were recovered from agarose gel and cloned into pGEM T Easy vector as described above. On average, 10 cDNA clones per locus were sequenced and analysed by comparison with the corresponding DNA sequences. RNA self ligation for *cox*1 transcript end mapping followed published procedures (36). Total *I. engelmannii* RNA was ligated by T4 RNA ligase (New England Biolabs, Ipswich, Massachusetts) and cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas, Burlington, Ontario) in the presence of 200 pmol of outward directed primers cox1leftdo1 and cox1rightup (1 and 2, respectively, in Figure 2). The same oligonucleotide pairs were used for first PCR amplification according to the standard protocol of BD Advantage 2 polymerase (BD Bioscience, Franklin Lakes, New Jersey) in a GeneAmp PCR System 2700 (Applied Biosystems, Foster City, California) with annealing temperatures at 45°C. Amplification results in several molecules of different sizes which were recovered from an agarose gel using the NucleoSpin Extract II Kit (Macherey Nagel, Düren, Germany). These molecules served as templates for subsequent nested PCR in presence of primers cox1leftdo and cox1rightup (1 and 3, respectively, in Figure 2) with same preferences as initial PCR.

Southern blotting

For Southern blotting, ~10 µg of *I. engelmannii* total DNA was digested with combinations of restriction enzymes (either *Eco*RI and *Cfr*9I or *Eco*RI and *Eco*RV, see Figure 6) and separated on a 0.8% agarose gel prior to blotting following established procedures (37). Approximately 100 ng of PCR-derived probes (Figure 6) were radioactively labelled with 50 µCi of α-P^{32}-dCTP. Hybridization of the nylon blotting membranes was overnight at 65°C in 50 mM sodium phosphate buffer containing 0.9 M NaCl, followed by washing in 2× SSC with 0.1% SDS at 65°C before exposure on a phosphor imager.

RESULTS

Genomic features and gene complement of the *Isoetes* mtDNA

The *I. engelmannii* mtDNA sequence was assembled from fosmid clones, identified in an arrayed library by hybridization with mitochondrial gene probes and verified in their mitochondrial nature through complete sequencing of the inserts. As more fosmid sequences were analyzed in the course of our studies, it became apparent that the same genes were repeatedly identified. Mitochondrial genes were found in different genomic environments, indicating a particularly high frequency of recombination events resulting in co-existing alternative gene arrangements (Figure 2). A total of 24 recombination breakpoints were identified, making the physical existence of a potential mtDNA master-circle encompassing the full mtDNA complexity highly unlikely. Different fosmid inserts reflected different pathways through the recombination points and the resulting products of DNA recombination were exemplarily verified as co-existing (see below). The net mtDNA sequence complexity of the analyzed *I. engelmannii* fosmid clones is 57 571 bp, with an overall A + T content of 51.3% and a percentage of 46.2% coding sequences.

We identified a typical complement of plant mitochondrial genes (Table 1) encoding subunits of respiratory chain complex I (*nad* genes *nad*1, 2, 3, 4, 4L, 5, 6, 7 and 9), complex II (*sdh*3), complex III (*cob*), complex IV (*cox*1, 2 and 3) and of complex V, the ATP synthase (*atp*1, 4, 6, 8 and 9). Likewise present are the genes for the large, small and 5S rRNAs (*rrn*L, *rrn*S, *rrn*5), for four ribosomal proteins (*rpl*5, *rps*2, *rps*3 and *rps*4) as well as the *tat*C gene encoding a subunit of the *sec*-independent transport pathway, and thirteen intact tRNA genes. Hence, on the one hand, four ribosomal protein genes demonstrated to be frequently transferred to the nucleus in angiosperms (38) are present in the *Isoetes* mtDNA. On the other hand, genes encoding cytochrome biogenesis components (*ccm*B, *ccm*C, *ccm*F) are completely lacking as had previously been observed for the land plant lineage only in the mtDNA of the green alga *Chaetosphaeridium* (8). To exclude the possibility that the *ccm* genes were accidently missed through yet a further recombination event, we have used oligonucleotide primers directed against conserved *ccm*B, *ccm*C and *ccm*F sequences but were unable to retrieve them in PCR approaches using *I. engelmannii* DNA. In addition, we identified small pseudogene fragments of three tRNA genes, of the *rrn* genes and of the *nad*4 and *rps*1 genes.

Recombination points (Figure 2) were identified both in intergenic regions (R1, R2, R8, R11, R20 and R21) as well as in coding regions (of *atp*6: R5, *atp*8: R12, *atp*9: R19, *cob*: R3, R22, *cox*1: R12, R13, R4, R14, *rpl*5: R10, *rps*2: R23, *rps*3: R4, *trn*F: R18, *nad*1: R6, R10, *nad*2: R24, R9, R15, *nad*3: R17, *nad*5: R16, R3b and *sdh*3: R7,

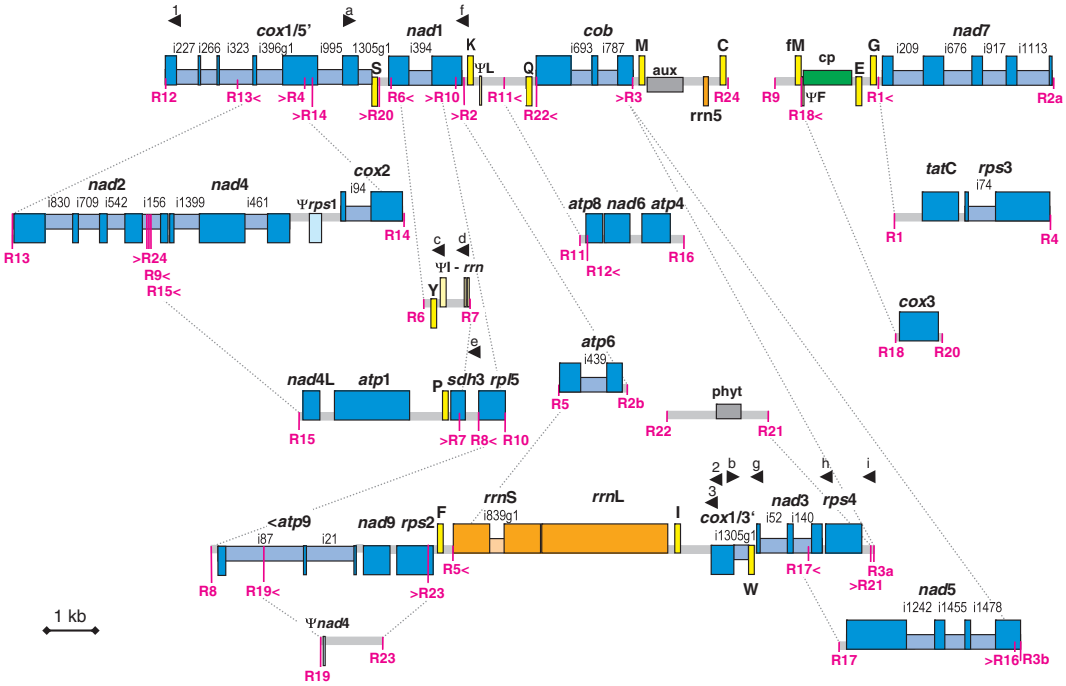


Figure 2. The *Isoetes engelmannii* mtDNA with protein-coding genes shown in blue, tRNAs in yellow, rRNAs in orange, pseudogenes in the respective lighter colours and the cp and nuc DNA inserts indicated with green and grey boxes, respectively. Drawing is approximately to scale. Genes shown above or below the lines indicate directions of transcription to the right or to the left, respectively. Recombination points (R1–R24) are highlighted in magenta with arrows indicating recombination forks. Selected connections between islands of recombination are exemplarily shown with stippled grey lines. Unique net mtDNA sequences add up to 57 571 bp. Arrowheads indicate oligonucleotide primers (1–3 and a–i) anchoring in regions not affected by RNA editing to analyze the arrangement of *cox*1 and its transcript maturation.

respectively), which accordingly result in fragmented pseudogene copies co-existing with the functional genes. In agreement with the observation of highly frequent recombination, only two evolutionary ancient gene linkages (*trn*P–*sdh*3 and *nad*4–*nad*2) are conserved as mere traces of much more extended syntenies that were identified when the liverwort *Marchantia* and the moss *Physcomitrella* mtDNAs were compared (23) to those of the streptophyte algae *Chaetosphaeridium globosum* (8) and *Chara vulgaris* (9). When not affected by recombination, intergenic regions between functional genes are small (only 17 bp between *nad*4 and *nad*2 and 8 bp between *atp*8 and *nad*6, respectively) with the exception of the spacer between *trn*K and *trn*Q carrying the pseudo-*trn*L fragment and the non-coding regions extended through the insertions of foreign DNA fragments.

Insertions of foreign DNA

Three 'promiscuous' DNA inserts of foreign origin were identified in the *I. engelmannii* mtDNA. A 1208 bp fragment of chloroplast DNA located between *trn*E and *trn*fM (Figure 2) covers parts of the chloroplast *trn*A and 23S rRNA genes. Highest similarity of this chloroplast sequence insert is found with the corresponding chloroplast sequence of another *Isoetes* species deposited in the database (*I. malinverniana*, DQ629281) indicating (recent) inter-organellar rather than horizontal gene transfer. Sequence deviations of the chloroplast insert in the *I. engelmannii* mtDNA from the native chloroplast homologue show striking pattern of degeneration with only two base changes within 1080 nt of 23S rRNA but indels of exclusively 5 or 6 bp (Supplementary Figure 2).

A 735 bp sequence stretch in the intergenic region between *rrn*5 and *trn*M bears strong similarity with nuclear encoded auxin-responsive transcription factors and a 533-bp sequence with similarity to phytochrome genes occurs between *rps*4 and *trn*Q (Figure 2). Like the chloroplast insertion, both nuclear sequence inserts are non-functional pseudogene fragments.

RNA editing in mRNAs

The protein-encoding genes in the *I. engelmannii* mtDNA show a very strong requirement for mRNA editing via pyrimidine exchanges to reconstitute evolutionary conserved codons. Altogether more than 1420 positions (over 1200 C-to-U and 220 U-to-C changes) in the *Isoetes* mitochondrial transcriptome appear to be subject to editing (Supplementary Table 1). This includes the reconstitution of appropriate AUG start codons from ACG threonine codons which is required in 12 cases and

Nucleic Acids Research, 2009, Vol. 37, No. 15 **5097**

Table 1. The gene and intron complement of *Isoetes engelmannii* (Ie) mtDNA in comparison to the mtDNAs of the liverwort *Marchantia polymorpha* (Mp) the moss *Physcomitrella patens* (Pp), the gymnosperm *Cycas taitungensis* (Ct) and the angiosperm *Arabidopsis thaliana* (At)

Genes/introns	*Mp*	*Pp*	*Ie*	*Ct*	*At*
atp1	+	+	+	+	+
atp1i989g2	+ +				
atp1i1050g2	+ +				
atp1i1128g2		+			
atp4	+	+	+	+	+
atp6	+	+	+	+	+
atp6i80g2		+			
atp6i439g2			+		
atp8	+	+	+	+	+
atp9	+	+	+	+	+
atp9i21g2		+	+		
atp9i87g2	+ +	+	+		
atp9i95g2		+			
ccmB	+	+		+	+
ccmC	+	+		+	+
ccmF	+	+		+	+
ccmFCi829g2		+		+	+
cob	+	+	+	+	+
cobi372g2	+				
cobi420g1		+			
cobi693g2			+		
cobi783g2	+				
cobi787g2			+		
cobi824g2	+ +				
cox1	+	+	+	+	+
cox1i44g2	+ +				
cox1i178g2	+ +				
cox1i227g2			+		
cox1i266g2			+		
cox1i323g2			+		
cox1i375g1	+				
cox1i395g1	+ +		+		
cox1i511g2	+ +	+			
cox1i624g1	+	+			
cox1i730g1	+ +				
cox1i732g2		+ +			
cox1i995g2			+		
cox1i1064g2		+			
cox1i1116g1	+ +				
cox1i1305g1	+		*trans*		
cox2	+	+	+	+	+
cox2i94g2			+		
cox2i97g2	+				
cox2i104g2		+			
cox2i250g2	+ +				
cox2i373g2		+		+	
cox2i691g2		+		+	+
cox3	+	+	+	+	+
cox3i171g2	+				
cox3i506g2		+			
cox3i625g2	+				
nad1	+	+	+	+	+
nad1i287g2		+			
nad1i394g2			+	*trans*	*trans*
nad1i477g2				+	+
nad1i669g2				*trans*	*trans*
nad1i728g2		+		+ +	+ +
nad2	+	+	+	+	+
nad2i156g2		+	+	+	+
nad2i542g2			+	*trans*	*trans*
nad2i709g2	+		+	+	+
nad2i830g2			+		
nad2i1282g2				+	+
nad3	+	+	+	+	+
nad3i52g2			+		
nad3i140g2	+		+		

(continued)

Table 1. Continued

Genes/introns	*Mp*	*Pp*	*Ie*	*Ct*	*At*
nad4	+	+	+	+	+
nad4i461g2		+	+	+	+
nad4i548g2	+				
nad4i976g2				+	+
nad4i1399g2			+	+	+
nad4L	+	+	+	+	+
nad4Li100g2	+				
nad4Li283g2	+	+			
nad5	+	+	+	+	+
nad5i230g2		+		+	+
nad5i753g1	+	+			
nad5i1242g2			+		
nad5i1455g2		+ +	+	*trans*	*trans*
nad5i1477g2			+	*trans*	*trans*
nad5i1872g2				+	+
nad6	+	+	+	+	+
nad7	Ψ	+	+	+	+
nad7i140g2		+			+
nad7i209g2		+	+	+	+
nad7i336g2	+				
nad7i676g2			+	+	+
nad7i917g2			+	+	+
nad7i1113g2	+		+	+	
nad9	+	+	+	+	+
nad9i283g2		+			
rpl2	+	+		+	+
rpl2i28g2	+				
rpl2i917g2				+	+
rpl5	+	+	+	+	+
rpl6	+	+			
rpl16	+	+		+	+
rps1	+	+	Ψ	+	
rps2	+	+	+	+	
rps3	+	+	+	+	+
rps3i74g2			+	+	+
rps3i257g2				+ +	
rps4	+	+	+	+	+
rps7	+	+		+	+
rps8	+				
rps10	+			+	
rps10i235g2				+	
rps11	+	+		+	
rps12	+	+		+	+
rps13	+	+		+	
rps14	+	+		+	Ψ
rps14i114g2	+				
rps19	+	+		+	Ψ
rrn5	+	+	+	+	+
rrnL	+	+	+	+	+
rrnLi827g2	+				
rrnS	+	+	+	+	+
rrnSi839g1			+		
rrnSi1065g2	+ +				
sdh3	+	+	+	+	
sdh3i100g2		+			
sdh4	+	+			Ψ
tatC	+	+	+	+	+
trnA(ugc)	+	+			
trnC(gca)	+	+	+	+	+
trnD(guc)	+	+		+	
trnD(guc) cp					+
trnE(uuc)	+	+	+	+	+
trnF(gaa)	+	+	+	+	
trnG(gcc)	+	+	+	+	+
trnG(ucc)	+	+			
trnH(gug)	+	+			
trnH(gug) cp				+	+
trnI(cau)	+	+	+	+	+
trnK(uuu)	+	+	+	+	+
trnL(caa)	+	+		+	

(continued)

5098 *Nucleic Acids Research, 2009, Vol. 37, No. 15*

Table 1. Continued

Genes/introns	*Mp*	*Pp*	*Ie*	*Ct*	*At*
trnL(uag)	+	+	Ψ	+	
trnL(uaa)	+	+			
trnM(cau)	+	+	+		
trnM(cau) cp				+	+
trnMf(cau)	+	+	+	+	+
trnN(guu)	+			+	
trnN(guu) cp					+
trnP(ugg)	+	+	+	+	+
trnP(ugg) cp					Ψ
trnQ(uug)	+	+	+	+	+
trnR(acg)	+	+			
trnR(ucg)	+				
trnR(ucu)	+	+		+	
trnS(gcu)	+			+	+
trnSi43g2	+				
trnS(uga)	+	+	+	+	+
trnS(gga) cp				+	+
trnT(ggu)	+	+			
trnV(uac)	+	+			
trnV(uac) cp				+	
trnVi39g2 cp				+	
trnW(cca)	+	+	+	+	
trnW(cca) cp					+
trnY(gua)	+	+	+	+	+

Plus signs indicate presence of a gene or intron, respectively and Ψ indicates recognizable pseudogene remnants (indicated only where not accompanied by a co-existing functional gene copy). Intron designations are according to the position of the preceding nucleotide in the uninterrupted coding sequence in the reading frames of *M. polymorpha* with the addendum g1 or g2 indicating group I or group II introns, respectively. Double plus signs (++) indicate presence of intron-borne ORFs.

the introduction of stop codons which is required for nine reading frames, respectively. In fact, the introduction of both the start and stop codons at the same time to correctly define the reading frames is necessary in five genes: *atp*6, *atp*9, *cox*3, *nad*4 and *nad*4L. Reverse U-to-C editing is required in the majority of mRNAs to convert genomically encoded stop codons into conserved glutamine or arginine codons. Indeed, only three out of the 24 protein encoding genes in the *Isoetes* mtDNA (*atp*4, *nad*3, *nad*4L) are without any stop codons on DNA level. To confirm the expectations on RNA editing, we performed exemplary cDNA analyses (Supplementary Table 1). RNA editing was indeed found to affect more than one out of five amino acid identities in the *atp*1 reading frame (115 of 515 codons) through 105 C-to-U exchanges and 23 U-to-C exchanges in the mRNA, including a change of six codons in a row with the sequence motif H-C-STOP-T-P-S changed into Y-R-Q-M-S-L by four C-to-U and two U-to-C exchanges in the transcript (not shown). A full 14 stop codons are removed through U-to-C RNA editing in the case of the *atp*1 mRNA. Approximately one out of seven nucleotides is affected by RNA editing in the *atp*9 reading frame (Figure 3A) resulting in sense changes in one out of three codons exactly as predicted to reconstitute evolutionary conserved codons, including introduction of both the start and the stop codon. Typically, the steady state pool of flowering plant mitochondrial

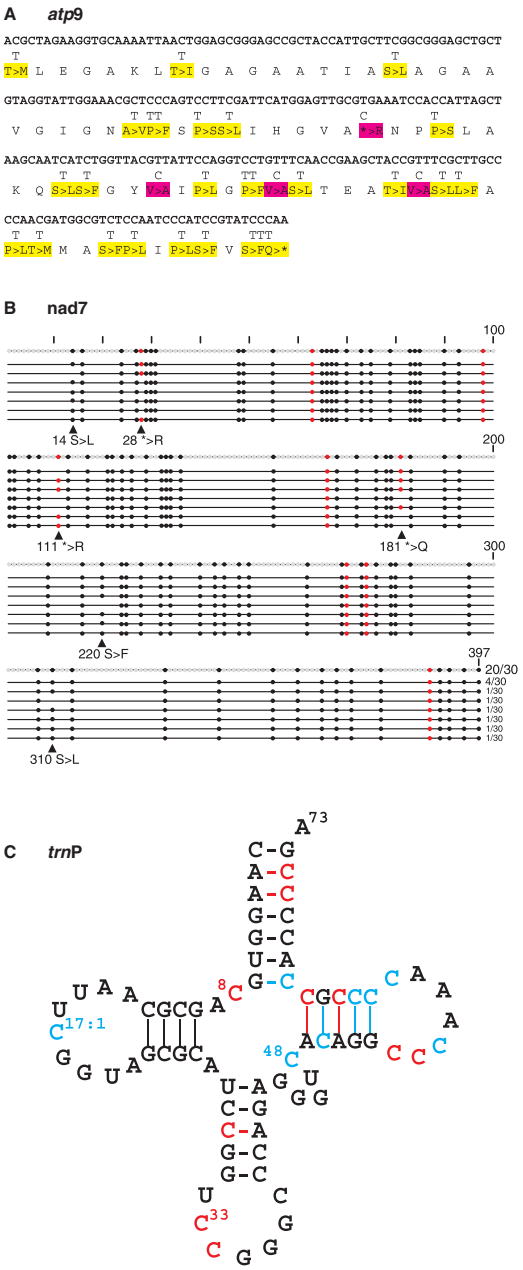


Figure 3. Examples of cDNA analyses for RNA editing: *atp*9 (A), *nad*7 (B) and *trn*P (C). (**A**) Yellow and magenta shading indicate changes introduced through C-to-U or U-to-C editing events in the *atp*9 mRNA. (**B**) The degree of partial editing of *nad*7 transcripts was investigated with 30 cDNA clones covering the complete reading frame of 397 codons. Black dots indicate codon sense changes, red dots the removal of stop codons. Twenty of 30 cDNA clones reflected complete editing at all predicted sites, the remaining clones lacked editing at certain positions. (**C**) Ten expected (red) and eight additional, non-predicted RNA editing sites (cyan) were found in *trn*P. Numbering of tRNA positions follows the standard convention with numbers after the colons indicating optional nucleotides not present in all tRNAs.

Nucleic Acids Research, 2009, Vol. 37, No. 15 **5099**

mRNAs contains transcripts edited to different degrees, reflecting only partial editing of some sites. To investigate this for *I. engelmannii* we examined 30 cDNA clones for the *nad*7 gene, for which we postulated RNA editing to correct 92 codon identities, including removal of nine stop codons. Complete editing of all the sites exactly as predicted was observed in 20 of the 30 cDNA clones, whereas four cDNAs lacked one editing to remove one of the stop codons (Figure 3B). The remaining six cDNA clones showed individual patterns lacking editing at this or another of one of five codons in total affected by partial editing (Figure 3B).

RNA editing in tRNAs

Cloverleaf modelling of the 13 tRNAs present in the *Isoetes* mtDNA strongly suggested frequent RNA editing activity to act on tRNAs as well. Several base-pairings in the four conserved stems and unpaired conserved uridines need to be re-established through C-to-U RNA editing in eleven tRNAs. The number of sites with predicted RNA editing events varied from single positions each in tRNA-fM and tRNA-G to six in tRNA-Q and even ten in tRNA-P (Figure 3C), respectively. Assuming that tRNA editing may take place in a precursor-transcript before processing we targeted a likely co-transcript of *trn*P with *sdh*3 (Figure 2) by RT–PCR, one of the rare cases of an ancient, conserved gene arrangement. Sequencing the cDNA product revealed not only the ten postulated positions of C-to-U exchanges, but also eight additional sites of C-to-U editing (but no reverse U-to-C changes), i.e. a total of 18 RNA editing positions in the *trn*P coding sequence (Figure 3C).

Introns

A total of 27 group II introns were identified in the *I. engelmannii* mtDNA (Table 1, nomenclature according to ref. (39), all of which are located in protein coding genes and most of which are particularly small. In fact, the *I. engelmannii* intron cox1i266 has a size of only 327 bp—to our knowledge, the smallest known group II intron as yet identified in any organism. Despite the strong recombinational activity in *I. engelmannii* mtDNA, none of the group II introns is in a *trans*-splicing arrangement. On the contrary, four of the known *trans*-splicing group II introns in angiosperms have *cis*-arranged counterparts in *I. engelmannii* (nad1i394, nad2i542, nad5i1455 and nad5i1477), seed plant introns nad1i669 and nad1i728 (in e.g. *Beta vulgaris* and *Oryza sativa*) obviously only appear later in evolution and get disrupted into *trans*-arrangements. A total of nine group II introns appear at novel insertion sites not yet observed in green algae (Charophytes or Chlorophytes), bryophytes or seed plants: atp6i439, cobi693, cobi787, cox1i227, cox1i266, cox1i323, cox1i995, cox2i94 and nad2i830 (Table 1).

Three group I introns were found in the *Isoetes* mtDNA, one in *rrn*S and two in *cox*1 (Figure 2). Orthologues of group I intron cox1i395 had previously been identified the liverwort *Marchantia* and in the alga *Chaetosphaeridium*, both of which carry endonuclease ORFs typical for this intron class. The *I. engelmannii* counterpart now identified is a small group I intron of only 328 bp without an ORF and hence similarly size-reduced as the group II introns. Yet smaller with a size of only 237 bp is a group I intron (rrnSi839g1) in the small ribosomal RNA gene *rrn*S.

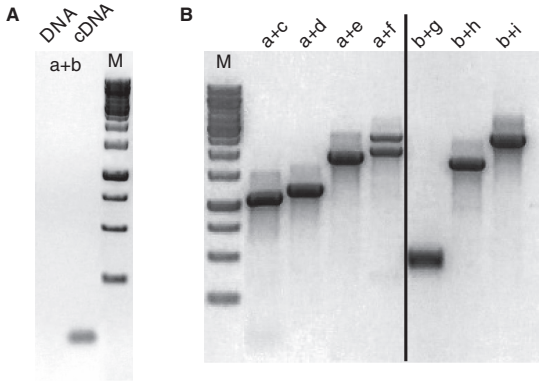


Figure 4. (**A**) Primers anchoring in the sub-terminal (a) and terminal (b) exons of *cox*1 fail to detect a genomic continuity across cox1i1305 or an alternative, intron-less *cox*1 copy with DNA but readily amplify the expected, spliced product with cDNA. (**B**) Each of the primers a and b combined with other primers anchoring in distant mtDNA regions (see Figure 2) downstream of *trn*Y (a + c), downstream of Ψ*trn*I (a + d), downstream of *sdh*3 (a + e), downstream of *nad*1 (a + f) or upstream of *trn*W (b + g), downstream of *nad*3 (b + h) or downstream of *rps*4 (b + i), respectively, all reveal amplicon products as expected. Two PCR products obtained simultaneously with primer combination a–f (lane 4) faithfully reflect the co-existing alternative genomic arrangements downstream of cox1/5′ to *nad*1 either via the *trn*S-*nad*1 continuity directly or alternatively through the *trn*S-R6-*trn*Y-Ψ*trn*I-R7-*sdh*3-*rpl*5-R10-*nad*1 pathway.

The most notable genomic peculiarity of the *Isoetes* mtDNA resides in the 3′ part of the *cox*1 gene (Figure 2). A group I intron (coxi1305) with a known homologue in the *Marchantia* mtDNA interrupts the *cox*1 coding region. However, intron homology breaks off sharply 210 bp after the splice donor site, 130 bp upstream of the *trn*S(uga) gene located downstream in inverted orientation. The seemingly missing terminal *cox*1 coding sequence was found elsewhere, preceded by the adequate splice acceptor site for joining the exons appropriately, 290 bp downstream of the *trn*W(cca) gene. Both parts of the *cox*1 coding regions have comparable similarities to other *cox*1 sequences in the database making independent, foreign origins, of the one or the other part of the gene, for example through horizontal gene transfer (15,16), unlikely.

To elucidate whether we had failed to identify a *cox*1 sequence continuity, we used primers anchoring in the directly flanking and also in distant *cox*1 exons, respectively, for PCR amplification assays on *I. engelmannii* DNA but failed to retrieve products (Figure 4A). To exclude potential malfunctions of the primers we used them individually in combinations with other primers anchoring in genomic distances ~2 kb apart in each case, as predicted from the recombinational mtDNA map (Figure 2). Expected products were retrieved

5100 *Nucleic Acids Research, 2009, Vol. 37, No. 15*

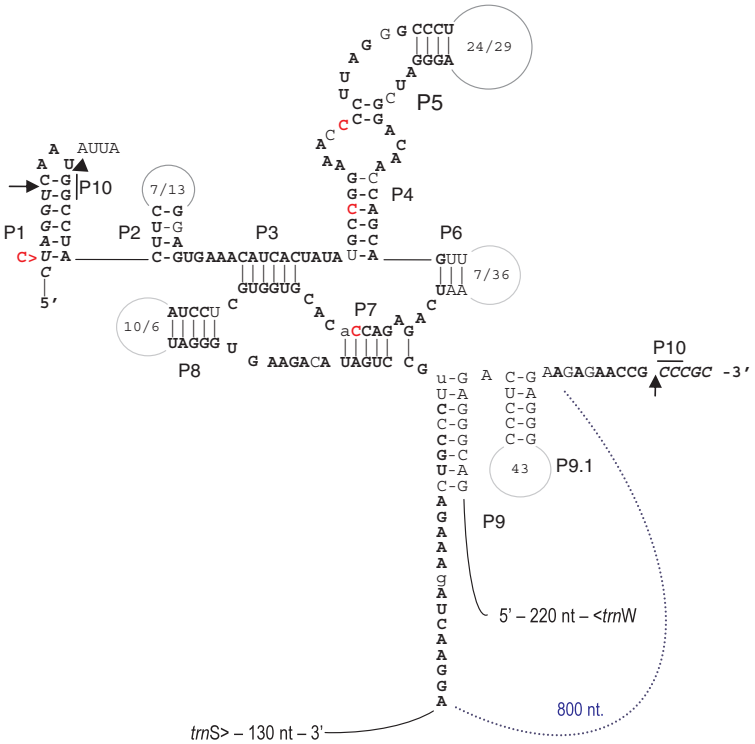


Figure 5. The *trans*-splicing group I intron cox1i1305 in *I. engelmannii*. The ribozymic core in the 5′-half (paired regions P1–P8) is conserved with the *cis*-arranged orthologue in *Marchantia polymorpha*. Italic letters indicate flanking exon sequences, bold letters indicate nucleotide identities in *Marchantia*, non-bold letters indicate transitions, lower case letters indicate transversions and numbers indicate loop sizes L2, L5, L6 and L8 in *Isoetes* and *Marchantia* (after the oblique), respectively. The intron discontinuity in *Isoetes* coincides with rearrangements in P9 which embraces a large continuous L9 loop of 800-nt in *Marchantia* (dotted line, blue). The *Marchantia* orthologue carries only traces of a formerly functional intron-encoded ORF. A C-to-U RNA editing event in the upstream exon is shown, three further cytidines in the intron core may likewise be subject to editings which could improve conservation of base-pairings.

both for the upstream part of *cox*1 extending downstream across several other genes and recombination points (*trn*S-R6-*trn*Y-*trn*I-R7-*sdh*3-*rpl*5-R10-*nad*1) as well as the downstream part of *cox*1 extending upstream across other genes (*trn*W-*nad*3-*rps*4), respectively (Figure 4B). Most notably, the different genomic routes downstream of *cox*1/5′ (Figure 2) identified through fosmid mapping were found to be faithfully reflected by two PCR products confirming the coexisting gene arrangements (Figure 4B).

RT–PCR products across the cox1i1305 discontinuity were easily retrieved from cDNA (Figure 4A). Cloning and sequencing verified correct splicing of cox1i1305 and all five additional upstream *cox*1 introns and showed differences to the genomic sequence exclusively at 106 positions of RNA editing, exactly as expected. Modelling the discontinuous group I intron sequences of cox1i1305 flanking the distantly located terminal *cox*1 exons reveals that the two sequence halves can combine for a classic group I secondary structure (Figure 5), to our knowledge the first example of a *trans*-splicing group I intron identified in nature. The typical ribozymic intron core structure of group I introns (40–42) is well conserved in comparison to its conventionally *cis*-arranged homologue in *Marchantia* (43).

To complement the PCR approaches outlined above in targeting potential alternatively arranged *cox*1 loci, we used *cox*1 cDNA as well as a mixture of the cox1i1305 intron halves as new probes. Rehybridization into our fosmid library, however, identified only those fosmids that had been identified and sequenced before.

To independently investigate the *cox*1 gene arrangement in *I. engelmannii* as deduced from the mtDNA map (Figure 2), we have used probes covering the terminal and sub-terminal exons 6 and 7 of the *cox*1 gene in a Southern blot hybridization experiment (Figure 6). Restriction sites for digestion of total genomic *Isoetes* DNA were selected to include the proximal recombination points identified near the upstream (R6) and the downstream part of the *cox*1 gene (R17). With hybridizations using probes for the upstream (Figure 6A) and downstream part of the *cox*1 gene (Figure 6B), two hybridizing restriction fragments were indeed identified in each case, reflecting the co-existing genomic rearrangements exactly as predicted from the genomic map (Figure 2). No further, additional hybridization signals were identified, which

Nucleic Acids Research, 2009, Vol. 37, No. 15 **5101**

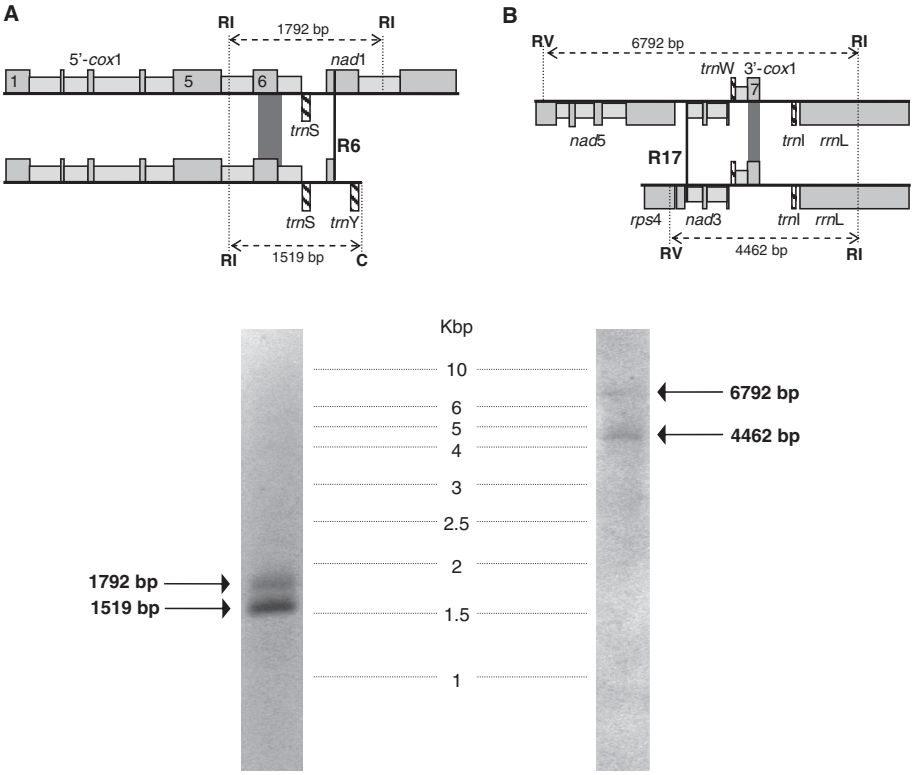


Figure 6. Southern-blot hybridization to verify *cox*1 gene arrangements for the upstream (A) and downstream (B) part of the gene using probes covering *cox*1 exons 6 and 7, respectively (dark grey rectangles). *Isoetes engelmannii* total genomic DNA was digested with *Eco*RI (RI) and *Cfr*9I (C) or with *Eco*RI (RI) and *Eco*RV (RV) to include the nearest identified recombination sites R6 or R17, respectively, in each case. Only two restriction fragments of expected sizes for coexisting genomic arrangements across R6 and R17 were identified by hybridization in each case: the 5′ *cox*1 part followed by *trn*S and either *nad*1 or by Ψ*nad*1-*trn*Y (**A**) and the 3′ *cox*1 part preceded by either *nad*5, Ψ*nad*3 and *trn*W or by *rps*4, *nad*3 and *trn*W (**B**).

could potentially represent a continuous *cox*1 gene copy (either intron-less or with a *cis*-splicing cox1i1305 counterpart) in full accord with the PCR experiments detailed above.

To determine *cox*1 transcript ends we used an approach of RNA circularization by self-ligation, followed by cDNA synthesis and RT–PCR with outward directed primers (oligonucleotides 1 and 2 in Figure 2). This revealed a 3′-UTR (untranslated region) extending 18 bp downstream of the stop codon and a 5′-UTR of 71 bp. The first seven *cox*1 codons are identical to those of the *atp*8 gene provided via recombination event R12 (Figure 2).

DISCUSSION

Lycophytes occupy a crucial position in the phylogeny of land plants (Figure 1), now unequivocally recognized as the sister group to euphyllophytes, which comprise the seed plants and the monilophytes with the latter encompassing the ferns, horsetails and whisk ferns (10,44,45). Comparatively poor in numbers of genera, families and with only three orders (Isoetales, Lycopodiales and Selaginellales) the recent lycophytes represent the most ancient lineage of vascular plants. As such, they could be expected to assume an intermediary position between the non-vascular bryophytes and the evolutionary advanced tracheophytes also with respect to the evolution of complexity in plant mtDNA. However, the *I. engelmannii* mitochondrial genome reported here as a first lycophyte mtDNA rather underlines the notion that plant mitochondria are 'more unique than ever' (46) by providing yet another example of unique pathways of organelle genome evolution.

Two main evolutionary trends have evidently shaped the *I. engelmannii* mtDNA. The gain and rise of recombinational activity seems to be the evolutionary force producing co-existing gene arrangements and the discontinuous group I intron now discovered in the *cox*1 gene. Likewise, highly active DNA recombination may be the ultimate prerequisite for the incorporation of DNA from the nuclear and chloroplast genomes, which has not been observed in bryophyte mtDNAs. After the recent report

of chloroplast DNA inserts in the mtDNA of the gymnosperm *Cycas taitungensis* (47), the first occurrences of such 'promiscuous' inserts of foreign DNA are now pushed back yet way further in plant evolution. The peculiar disposition of plant mtDNA to incorporate foreign genetic material originating from the other two genomes in the plant cell may have evolved with the increase of recombinational activity in the earliest tracheophytes (Figure 1).

The small introns and the small intergenic regions in *Isoetes* mtDNA on the other hand seem to reflect a counter-acting trend for organelle genome compaction. This tendency is also reflected in the *Isoetes* mtDNA gene complement itself, which generally mirrors the observations made for independent nuclear gene transfer in a survey of 280 flowering plant genera (38). Ribosomal protein genes that were found frequently and independently lost from the angiosperm mtDNAs are similarly missing from the *Isoetes* mtDNA, whereas those found to be transferred to the nucleus more rarely are (still) present (notably *rps*2, *rps*3 and *rps*4). Obvious exceptions on the other hand are the *ccm* genes not present in the *Isoetes* mtDNA. The inability to identify *ccm* genes independently via PCR may either indicate the commonly observed significant sequence alteration after nuclear gene transfer or, as a more remote possibility, an evolutionary switch to an alternative pathway of cytochrome *c* maturation (48,49). *Vice versa*, the *sdh*3 gene was here identified in the *Isoetes* mtDNA but is frequently transferred to the nucleus in angiosperms.

Notably, despite high recombinational activity, none of the 27 group II introns in the *I. engelmannii* mtDNA was found in a *trans*-splicing arrangement. On the contrary, four of the conserved *trans*-splicing group II introns of angiosperms find their small orthologues as *cis*-arranged counterparts in *Isoetes* (Table 1). Hence, the *trans*-splicing group I intron reported here to occur in the *cox*1 gene may represent a mere chance product with recombination acting before size reduction towards a minimum ribozyme core had reduced the chances of creating a discontinuous, yet functional, intron. *Trans*-splicing group II introns are known for more than 20 yrs since their discoveries both in chloroplasts (50,51) of algae and land plants and briefly thereafter in plant mitochondria (52–54) and, more recently, also in the mtDNA of an alga (55).

The first example of a *trans*-splicing group I Intron in nature shows that discontinuous molecules exist in yet another class of ribozyme-type RNAs after a discontinuous hammerhead RNA had been reported very recently (56). For mitochondrial genomes, yet another type of gene discontinuity with 'modules' distributed over separate DNA molecules for which the mechanisms of RNA maturation still have to be determined had recently been described for the protist *Diplonema* (57). Other examples for unusual modes of RNA maturation have also been reported outside of mitochondria such as tRNAs encoded in separate genes for 5′ and 3′ halves in *Nanoarchaeum* (58,59) or circularly permuted tRNAs expressed via circular RNA intermediates in the red alga *Cyanidioschyon* (60).

Whereas several of the introns in the *Isoetes* mtDNA have clear homologues at identical positions in bryophyte or seed plant mitochondrial genomes (Table 1), nine of the group II insertion sites are so far unique in the quillwort. It will be highly interesting whether homologues of these introns can be identified in the remaining major land plant clades for which complete mitochondrial genomes are still missing (Figure 1): ferns, horsetails, whisk ferns and hornworts given that their gains and losses could add independent further data relevant to the backbone of land plant phylogeny.

The extreme requirement of RNA editing in *I. engelmannii* mitochondrial RNAs not only to re-establish conserved reading frames in mRNAs but also to reconstitute secondary structures of tRNAs exceeds what has been observed before including the recent estimates of some 1000 editing sites in *Cycas taitungensis* mitochondrial mRNAs (32). The C-to-U editings hitherto observed in tRNAs of plant organelles were considered functionally essential, yet rare, events, e.g. (61–63). Likewise, such events of C-to-U editing have also been reported in animal mitochondria, e.g. (64). Multiple sites of RNA editing in single mitochondrial tRNA species had been observed as different types of editing: the replacement of nucleotides in tRNA acceptor stems, e.g. in the protist *Acanthamoeba castellanii* (65), the chytridomycete fungus *Spizellomyces punctatus* (66), de novo synthesis of 3′ ends in a centipede (67) or the pyrimidine insertional type of editing in slime molds (68).

Given the extraordinary degree of recombination, the presence of chloroplast and nuclear sequence inserts, a *trans*-splicing group I intron, the extraordinary amounts of RNA editing in mRNAs and, most notably, in hitherto unseen amounts also in tRNAs the *I. engelmannii* mtDNA once more demonstrates that 'anything goes' in mitochondrial genome evolution (69).

ACCESSION NUMBERS

FJ010859, FJ536259, FJ390841, FJ176330, FJ628360.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are highly grateful to Dr Jeffrey Palmer and co-workers (Bloomington, IN, USA.) for generously making fresh material of *Isoetes engelmannii* available to us, to Diana Kühn at the Max-Planck-Institut (MPI) für Züchtungsforschung, Cologne, for fosmid clone sorting and filter spotting. We also wish to thank Monika Polsakiewicz for her skillful technical assistance; and Julia Neuwirt, Yesim Kümetepe and Patrick Johner for their early work on lycophyte mtDNA cosmid cloning attempts. Finally, we are very grateful for helpful comments on this manuscript by two anonymous reviewers. Sequences obtained in this study were deposited in

GenBank (accession numbers FJ010859, FJ536259, FJ390841, FJ176330, FJ628360).

FUNDING

Deutsche Forschungsgemeinschaft [DFG Kn411/6-1].

Conflict of interest statement. None declared.

REFERENCES

1. Gray,M.W., Burger,G. and Lang,B.F. (1999) Mitochondrial evolution. *Science*, **283**, 1476–1481.
2. Lavrov,D.V. (2007) Key transitions in animal evolution: a mitochondrial DNA perspective. *Integr. Compar. Biol.*, **47**, 734–743.
3. Gray,M.W., Lang,B.F. and Burger,G. (2004) Mitochondria of protists. *Annu. Rev. Genet.*, **38**, 477–524.
4. Lang,B.F., Burger,G., O'Kelly,C.J., Cedergren,R., Golding,G.B., Lemieux,C., Sankoff,D., Turmel,M. and Gray,M.W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, **387**, 493–497.
5. Vaidya,A.B., Akella,R. and Suplick,K. (1989) Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Mol. Biochem. Parasitol.*, **35**, 97–107.
6. Ward,B.L., Anderson,R.S. and Bendich,A.J. (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, **25**, 793–803.
7. Beagley,C.T., Okada,N.A. and Wolstenholme,D.R. (1996) Two mitochondrial group I introns in a metazoan, the sea anemone Metridium senile: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proc. Natl Acad. Sci. USA*, **93**, 5619–5623.
8. Turmel,M., Otis,C. and Lemieux,C. (2002) The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc. Natl Acad. Sci. USA*, **99**, 11275–11280.
9. Turmel,M., Otis,C. and Lemieux,C. (2003) The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *Plant Cell*, **15**, 1888–1903.
10. Qiu,Y.L., Li,L., Wang,B., Chen,Z., Knoop,V., Groth-Malonek,M., Dombrovska,O., Lee,J., Kent,L., Rest,J. *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl Acad. Sci. USA*, **103**, 15511–15516.
11. Stern,D.B. and Lonsdale,D.M. (1982) Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature*, **299**, 698–702.
12. Knoop,V., Unseld,M., Marienfeld,J., Brandt,P., Sünkel,S., Ullrich,H. and Brennicke,A. (1996) *copia*-, *gypsy*- and LINE-like retrotransposon fragments in the mitochondrial genome of *Arabidopsis thaliana*. *Genetics*, **142**, 579–585.
13. Vaughn,J.C., Mason,M.T., Sper-Whitis,G.L., Kuhlman,P. and Palmer,J.D. (1995) Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of Peperomia. *J. Mol. Evol.*, **41**, 563–572.
14. Cho,Y.R. and Palmer,J.D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol.*, **16**, 1155–1165.
15. Bergthorsson,U., Adams,K.L., Thomason,B. and Palmer,J.D. (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, **424**, 197–201.
16. Richardson,A.O. and Palmer,J.D. (2006) Horizontal gene transfer in plants. *J. Exp. Bot.*, **58**, 1–9.
17. Kubo,T. and Mikami,T. (2007) Organization and variation of angiosperm mitochondrial genome. *Phys. Plant.*, **129**, 6–13.
18. Knoop,V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Curr. Genet.*, **46**, 123–139.
19. Takenaka,M., Verbitskiy,D., van der Merwe,J.A., Zehrmann,A. and Brennicke,A. (2008) The process of RNA editing in plant mitochondria. *Mitochondrion*, **8**, 35–46.
20. Maier,R.M., Zeltz,P., Kössel,H., Bonnard,G., Gualberto,J.M. and Grienenberger,J.M. (1996) RNA editing in plant mitochondria and chloroplasts. *Plant Mol. Biol.*, **32**, 343–365.
21. Steinhauser,S., Beckert,S., Capesius,I., Malek,O. and Knoop,V. (1999) Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? *J. Mol. Evol.*, **48**, 303–312.
22. Oda,K., Yamato,K., Ohta,E., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Kanegae,T., Ogura,Y., Kohchi,T. *et al.* (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J. Mol. Biol.*, **223**, 1–7.
23. Terasawa,K., Odahara,M., Kabeya,Y., Kikugawa,T., Sekine,Y., Fujiwara,M. and Sato,N. (2006) The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol. Biol. Evol.*, **24**, 699–709.
24. Oldenburg,D.J. and Bendich,A.J. (2001) Mitochondrial DNA from the liverwort Marchantia polymorpha: circularly permuted linear molecules, head-to-tail concatemers, and a 5′ protein. *J. Mol. Biol.*, **310**, 549–562.
25. Unseld,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.*, **15**, 57–61.
26. Sugiyama,Y., Watase,Y., Nagase,M., Makita,N., Yagura,S., Hirai,A. and Sugiura,M. (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genom.*, **272**, 603–615.
27. Ogihara,Y., Yamazaki,Y., Murai,K., Kanno,A., Terachi,T., Shiina,T., Miyashita,N., Nasuda,S., Nakamura,C., Mori,N. *et al.* (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res.*, **33**, 6235–6250.
28. Shedge,V., Arrieta-Montiel,M., Christensen,A.C. and Mackenzie,S.A. (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell*, **19**, 1251–1264.
29. Bendich,A.J. (1993) Reaching for the ring: the study of mitochondrial genome structure. *Curr. Genet.*, **24**, 279–290.
30. Malek,O. and Knoop,V. (1998) Trans-splicing group II introns in plant mitochondria: the complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. *RNA*, **4**, 1599–1609.
31. Groth-Malonek,M., Pruchner,D., Grewe,F. and Knoop,V. (2005) Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol. Biol. Evol.*, **22**, 117–125.
32. Chaw,S.M., Chun-Chieh,S.A., Wang,D., Wu,Y.W., Liu,S.M. and Chou,T.Y. (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol. Biol. Evol.*, **25**, 603–615.
33. Lohse,M., Drechsel,O. and Bock,R. (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.*, **52**, 267–274.
34. Kumar,S., Nei,M., Dudley,J. and Tamura,K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.*, **9**, 299–306.
35. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
36. Kempken,F., Bolle,N., Forner,J. and Binder,S. (2008) Transcript end mapping and analysis of RNA editing in plant mitochondria. In Leister,D. and Herrmann,J.M. (eds), *Mitochondria: Practical Protocols,* Vol. 372, Humana Press, Totowa, NJ, pp. 177–192.
37. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
38. Adams,K.L., Qiu,Y.L., Stoutemyer,M. and 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. USA*, **99**, 9905–9912.
39. Dombrovska,E. and Qiu,Y.L. (2004) Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Mol. Phylogenet. Evol.*, **32**, 246–263.
40. Lang,B.F., Laforest,M.J. and Burger,G. (2007) Mitochondrial introns: a critical view. *Trends Genet.*, **23**, 119–125.
41. Saldanha,R., Mohr,G., Belfort,M. and Lambowitz,A.M. (1993) Group I and group II introns. *FASEB J.*, **7**, 15–24.
42. Michel,F. and Westhof,E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.*, **216**, 585–610.
43. Ohta,E., Oda,K., Yamato,K., Nakamura,Y., Takemura,M., Nozato,N., Akashi,K., Ohyama,K. and Michel,F. (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucleic Acids Res.*, **21**, 1297–1305.
44. Pryer,K.M., Schneider,H., Smith,A.R., Cranfill,R., Wolf,P.G., Hunt,J.S. and Sipes,S.D. (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature*, **409**, 618–622.
45. Qiu,Y.L. (2008) Phylogeny and evolution of charophytic algae and land plants. *J. Syst. Evol.*, **46**, 287–306.
46. Rasmusson,A.G., Handa,H. and Moller,I.M. (2008) Plant mitochondria, more unique than ever. *Mitochondrion*, **8**, 1–4.
47. Wang,D., Wu,Y.W., Shih,A.C., Wu,C.S., Wang,Y.N. and Chaw,S.M. (2007) Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Mol. Biol. Evol.*, **24**, 2040–2048.
48. Giegé,P., Grienenberger,J.M. and Bonnard,G. (2008) Cytochrome c biogenesis in mitochondria. *Mitochondrion*, **8**, 61–73.
49. Allen,J.W., Jackson,A.P., Rigden,D.J., Willis,A.C., Ferguson,S.J. and Ginger,M.L. (2008) Order within a mosaic distribution of mitochondrial c-type cytochrome biogenesis systems? *FEBS J.*, **275**, 2385–2402.
50. Koller,B., Fromm,H., Galun,E. and Edelman,M. (1987) Evidence for in vivo trans splicing of pre-mRNAs in tobacco chloroplasts. *Cell*, **48**, 111–119.
51. Choquet,Y., Goldschmidt-Clermont,M., Girard-Bascou,J., Kuck,U., Bennoun,P. and Rochaix,J.D. (1988) Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C. reinhardtii chloroplast. *Cell*, **52**, 903–913.
52. Knoop,V., Schuster,W., Wissinger,B. and Brennicke,A. (1991) *Trans* splicing integrates an exon of 22 nucleotides into the *nad5* mRNA in higher plant mitochondria. *EMBO J.*, **10**, 3483–3493.
53. Wissinger,B., Schuster,W. and Brennicke,A. (1991) *Trans* splicing in Oenothera mitochondria: *nad1* mRNAs are edited in exon and *trans*-splicing group II intron sequences. *Cell*, **65**, 473–482.
54. Chapdelaine,Y. and Bonen,L. (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: a trans-splicing model for this gene-in-pieces. *Cell*, **65**, 465–472.
55. Turmel,M., Otis,C. and Lemieux,C. (2002) The complete mitochondrial DNA sequence of Mesostigma viride identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol. Biol. Evol.*, **19**, 24–38.
56. Martick,M., Horan,L.H., Noller,H.F. and Scott,W.G. (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, **454**, 899–902.
57. Marande,W. and Burger,G. (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science*, **318**, 415.
58. Randau,L., Münch,R., Hohn,M.J., Jahn,D. and Söll,D. (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5′- and 3′-halves. *Nature*, **433**, 537–541.
59. Randau,L. and Soll,D. (2008) Transfer RNA genes in pieces. *EMBO Rep.*, **9**, 623–628.
60. Soma,A., Onodera,A., Sugahara,J., Kanai,A., Yachie,N., Tomita,M., Kawamura,F. and Sekine,Y. (2007) Permuted tRNA genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. *Science*, **318**, 450–453.
61. Binder,S., Marchfelder,A. and Brennicke,A. (1994) RNA editing of $tRNA^{Phe}$ and $tRNA^{Cys}$ in mitochondria of *Oenothera berteriana* is initiated in precursor molecules. *Mol. Gen. Genet.*, **244**, 67–74.
62. 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.
63. Miyata,Y., Sugita,C., Maruyama,K. and Sugita,M. (2008) RNA editing in the anticodon of tRNA(Leu) (CAA) occurs before group I intron splicing in plastids of a moss *Takakia lepidozioides* S. Hatt. & Inoue. *Plant Biol (Stuttg)*, **10**, 250–255.
64. Janke,A. and Pääbo,S. (1993) Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Res.*, **21**, 1523–1525.
65. Lonergan,K.M. and Gray,M.W. (1993) Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science*, **259**, 812–816.
66. Laforest,M.J., Roewer,I. and Lang,B.F. (1997) Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and UAG 'stop' codons recognized as leucine. *Nucleic Acids Res.*, **25**, 626–632.
67. Lavrov,D.V., Brown,W.M. and Boore,J.L. (2000) A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc. Natl Acad. Sci USA*, **97**, 13738–13742.
68. Antes,T., Costandy,H., Mahendran,R., Spottswood,M. and Miller,D. (1998) Insertional editing of mitochondrial tRNAs of *Physarum polycephalum* and *Didymium nigripes*. *Mol. Cell Biol.*, **18**, 7521–7527.
69. Burger,G., Gray,M.W. and Lang,B.F. (2003) Mitochondrial genomes: anything goes. *Trends Genet.*, **19**, 709–716.

4.1.2 Supplementary Data

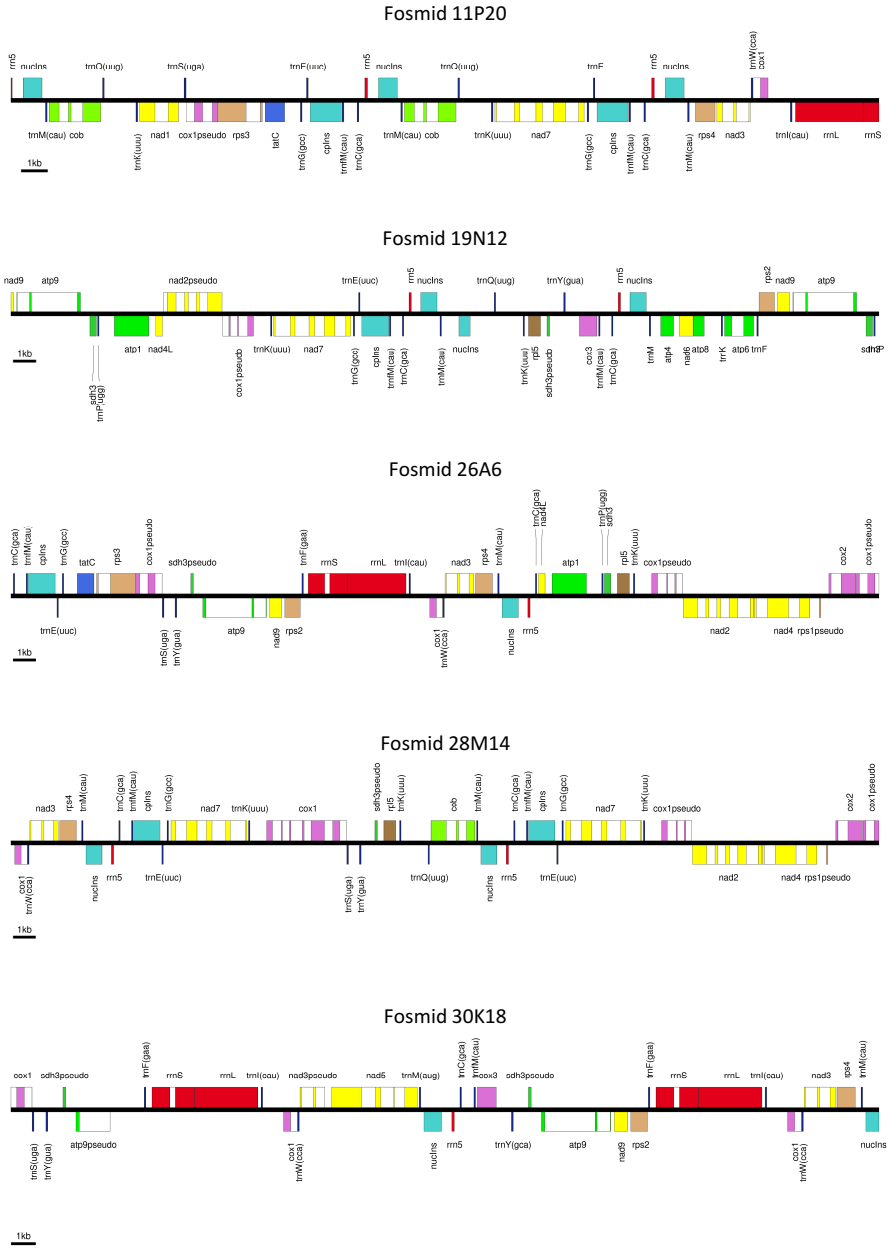


Figure 4.7: (Supplementary Figure 1) Gene maps of five *Isoetes engelmannii* mitochondrial fosmid clones. Protein-coding genes, ribosomal and transfer RNA genes and foreign DNA insertions, shown above or below the lines indicate directions of transcription to the right or to the left, respectively. Maps were generated with OGDRAW v1.1 software.

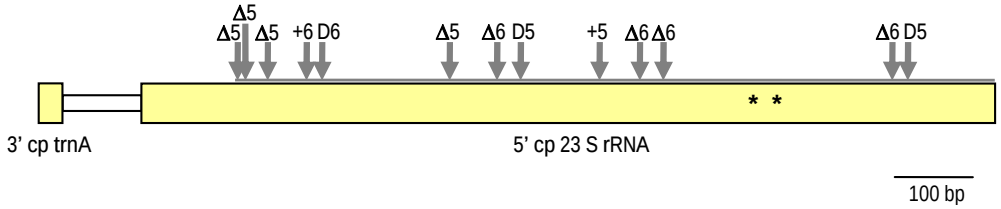


Figure 4.8: (Supplementary Figure 2) A 1208 bp insert of chloroplast (cp) DNA extending over the 3'-end of cp trnA and the 5'-end of cp 23S rRNA is located between *trn*fM and *trn*E in the *Isoetes engelmannii* mtDNA. Drawing is approximately to scale. Comparison to the homologous chloroplast sequence of *I. malinverniana* (database accession DQ629281, grey line) allows to identify mutations accumulating in the transferred chloroplast sequence segment. Only two nucleotide exchanges (asterisks) but several deletions (Δ5/Δ6), insertions (+5/+6) and duplications (D5/D6) of exclusively five or six basepairs, respectively, are observed (grey arrows).

	total (silent)		size	C-to-U codon changes														U-to-C codon changes																		total	start	stop
	C>U	U>C	[bp]	P>L	P>S	P>F	S>L	S>F	L>F	Q>*	R>*	R>C	R>W	A>V	H>Y	T>M	T>I	L>P	S>P	F>P	L>S	F>S	F>L	*>Q	*>R	C>R	W>R	Y>H	V>A	M>T	I>T								
atp1	**105(5)**	**23**	**1545**	**18**	**12**	**3**	**20**	**5**	**1**	**0**	**0**	**5**	**0**	**5**	**9**	**4**	**7**	**0**	**1**	**0**	**1**	**0**	**0**	**12**	**2**	**4**	**0**	**0**	**3**	**0**	**3**	**115**	■	□
atp4	6	1	600	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0		4	□	□
atp6	64	13	762	11	5	6	16	8	2	0	1	3	1	2	3	3	3	1	0	0	1	0	0	5	0	2	0	2	1	0	1	77	■	■
atp8	31	6	489	6	1	1	6	0	0	0	0	1	1	0	3	2	1	0	0	0	0	0	0	3	1	0	0	1	0	0	0	27	■	□
atp9	**27**	**4**	**225**	**4**	**2**	**2**	**5**	**4**	**1**	**1**	**0**	**0**	**0**	**1**	**0**	**2**	**2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**3**	**0**	**0**	**28**	■	■
cob	71	24	1170	15	1	4	14	6	7	0	0	0	6	3	13	0	2	4	3	0	2	0	1	4	1	1	0	2	5	0	1	95	□	□
cox1	**83(8)**	**23(2)**	**1602**	**17**	**4**	**2**	**11**	**11**	**5**	**0**	**0**	**2**	**6**	**0**	**5**	**3**	**2**	**0**	**2**	**0**	**1**	**2**	**1**	**2**	**1**	**2**	**0**	**4**	**5**	**0**	**1**	**89**	■	□
cox2	11	5	758	1	0	0	1	0	2	0	0	1	2	0	1	1	2	0	1	0	0	0	0	2	0	1	0	0	1	0	0	16	■	□
cox3	70	5	792	14	6	11	5	8	1	1	0	1	9	2	4	5	3	0	0	0	0	1	0	1	0	1	0	1	1	0	0	75	■	■
nad1	31	10	987	5	6	2	6	1	1	0	0	0	2	2	1	4	1	1	3	0	0	1	0	3	1	1	0	0	0	0	0	41	■	□
nad2	81	13	1464	13	8	2	11	12	8	1	0	6	1	0	10	3	6	1	3	0	1	0	2	2	0	0	0	0	2	1	1	94	□	■
nad3	**53**	**4**	**357**	**6**	**2**	**2**	**9**	**10**	**0**	**1**	**0**	**2**	**1**	**0**	**2**	**0**	**3**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**40**	□	■
nad4	94	16	1488	22	5	11	17	13	0	1	0	3	5	1	8	6	2	2	4	0	2	1	1	2	2	0	0	2	0	0	0	110	■	■
nad4L	24	1	303	2	3	0	6	3	0	1	0	0	0	0	0	4	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	22	■	■
nad5	113	12	2001	13	11	7	25	17	6	0	0	7	8	2	7	7	3	0	0	0	2	0	0	3	3	0	0	1	1	0	2	125	□	□
nad6	56	5	627	11	2	4	8	1	1	0	0	1	3	2	1	5	2	0	1	0	1	1	0	1	0	0	0	1	0	0	0	46	■	□
nad7	**83(4)**	**17(1)**	**1190**	**16**	**3**	**2**	**15**	**7**	**0**	**1**	**0**	**4**	**2**	**5**	**7**	**8**	**4**	**2**	**0**	**0**	**0**	**1**	**1**	**5**	**5**	**1**	**0**	**4**	**0**	**0**	**2**	**95**	□	■
nad9	37	10	558	2	2	3	8	4	1	0	0	1	5	1	1	3	1	2	0	0	2	0	0	3	2	0	0	0	1	0	0	42	■	□
rpl5	20	5	549	3	1	0	3	1	0	0	0	1	0	2	0	1	1	1	0	0	1	1	0	0	1	0	0	0	1	0	0	18	□	□
rps2	14	4	699	2	1	0	6	0	1	0	0	3	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	2	18	□	□
rps3	44	7	1200	5	4	1	6	6	0	0	0	3	2	2	3	0	1	0	0	0	0	0	0	4	2	0	0	0	0	0	0	39	□	□
rps4	43	9	765	13	5	1	4	3	0	0	0	1	1	2	4	0	3	0	0	0	0	0	0	3	2	0	0	0	0	0	1	43	□	□
sdh3	**19(5)**	**2**	**297**	**3**	**3**	**1**	**2**	**1**	**0**	**1**	**0**	**1**	**0**	**0**	**1**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**16**	□	■
tatC	46	6	756	8	7	2	3	9	4	0	0	4	3	0	4	0	2	1	1	0	1	0	0	2	0	1	0	0	0	0	0	52	□	□
trnC(gca)	4																																					
trnE(uuc)	0																																					
trnF(gaa)	3																																					
trnG(gcc)	1																																					
trnI(cau)	3																																					
trnK(uuu)	4																																					
trnM(cau)	2																																					
trnMf(cau)	1																																					
trnP(ugg)	**18**																																					
trnQ(uug)	6																																					
trnS(uga)	0																																					
trnW(cca)	5																																					
trnY(gua)	4																																					
total	1299	228																																				

Table 4.2: (Supplementary Table 1) RNA editing positions predicted to correct codon identities in *Isoetes engelmannii* mitochondrial mRNAs and to re-establish base pairing in tRNAs. All predicted editing sites have exemplarily been confirmed through cDNA analyses (bold) for *atp*9, *nad*7 and *trn*P (see Figure 3) and for *atp*1, *cox*1 and *sdh*3 (not shown).

4.2 Das mitochondriale Transkriptom von *Isoetes engelmannii*

Mit der vollständigen Sequenzierung und Analyse der mtDNA von *I. engelmannii* wurden viele neuen Eigenheiten des mitochondrialen Evolutionsweges dieses Lycophyten erkannt. Aber nicht nur eine außergewöhnliche mtDNA wurde sichtbar, sondern auch besondere post-trankriptionelle Modifikationen, wie das *trans*-Spleißen eines Gruppe I Introns und ein hohes Vorkommen von *RNA Editing*. Insgesamt wurden mehr als 1420 *RNA Editing* Positionen vorhergesagt und in ausgewählten Genen verifiziert. Auch eine Änderung von 18 Basen in der tRNA-P, dem sogenannten *tRNA-Hyperediting*, wurde nachgewiesen [Grewe *et al.* 2009].

Wie bereits in Abschnitt 2.7 dargestellt, wurden bisher nur wenige mitochondriale Transkriptome vollständig analysiert, darunter keines aus der Gruppe der Lycophyten. Deshalb sollte die vollständige Sequenzierung und Analyse des Transkriptoms des Lycophyten *I. engelmannii* eine weitere phylogenetische Lücke füllen und Informationen über die post-transkriptionellen Besonderheiten dieser Gruppe liefern. Die Transkripte von *I. engelmannii* werden über die besondere Form des U-zu-C *Editings* geändert und unterscheiden dieses Transkriptom damit von allen bisher analysierten.

Durch eine Sequenzierung vieler cDNA-Klone jedes Transkripts wurden im Rahmen einer begleitenden Diplomarbeit auch partielle Editing Positionen von *I. engelmannii* präzise erfasst [Stefan Herres, Universität Bonn, Diplomarbeit]. Neben der höchsten bisher gesehenen Gesamtzahl von 1782 *RNA Editing* Positionen in proteinkodierenen Genen, wurden die Transkriptomanalysen zudem auf alle tRNA-Transkripte und alle Introns des *nad*7-Gens ausgeweitet, wodurch weitere Charakteristika der Transkript-Modifikationen nachgewiesen werden konnten. Die nachfolgende Arbeit *A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte* veröffentlicht in *Nucleic Acids Research*, zeigt – aufbauend auf den Besonderheiten der ersten mtDNA Sequenzierung eines Lycophyten – die Eigenheiten dessen post-transkriptionaler Modifikationen.

Nucleic Acids Research Advance Access published December 7, 2010

Nucleic Acids Research, 2010, ***1–13***
doi:10.1093/nar/gkq1227

A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*

Felix Grewe[1,*], Stefan Herres[1], Prisca Viehöver[2], Monika Polsakiewicz[1], Bernd Weisshaar[2] and Volker Knoop[1]

[1]IZMB (Institut für Zelluläre und Molekulare Botanik), Universität Bonn, Kirschallee 1, 53115 Bonn and [2]IGS (Institut für Genomforschung und Systembiologie), Universität Bielefeld, Universitätsstraße 25, 33594 Bielefeld, Germany

Received September 22, 2010; Revised October 30, 2010; Accepted November 11, 2010

ABSTRACT

The analysis of the mitochondrial DNA of *Isoetes engelmannii* as a first representative of the lycophytes recently revealed very small introns and indications for extremely frequent RNA editing. To analyze functionality of intron splicing and the extent of RNA editing in *I. engelmannii*, we performed a comprehensive analysis of its mitochondrial transcriptome. All 30 groups I and II introns were found to be correctly removed, showing that intron size reduction does not impede splicing. We find that mRNA editing affects 1782 sites, which lead to a total of 1406 changes in codon meanings. This includes the removal of stop codons from 23 of the 25 mitochondrial protein encoding genes. Comprehensive sequence analysis of multiple cDNAs per locus allowed classification of partially edited sites as either inefficiently edited but relevant or as non-specifically edited at mostly low frequencies. Abundant RNA editing was also found to affect tRNAs in hitherto unseen frequency, taking place at 41 positions in tRNA-precursors, including the first identification of U-to-C exchanges in two tRNA species. We finally investigated the four group II introns of the *nad*7 gene and could identify 27 sites of editing, most of which improve base pairing for proper secondary structure formation.

INTRODUCTION

Even 20 years after the original discovery of RNA editing that exchanges cytidine and uridine nucleotides in plant mitochondrial and chloroplast transcripts, many aspects of the phenomenon remain enigmatic (1–4). The reasons why RNA editing came into being during land plant evolution remain unclear. No obvious functional gain or evolutionary adaptation can be connected with RNA editing in plant organelles. Similarly, strong evidence that plant RNA editing may functionally modulate gene activity is lacking, although several publications have reported variability of RNA editing among different environments, ecotypes or plant tissues (5–13). RNA editing in plant organelles largely seems to act as a correction mechanism which reinstalls codons conserved during evolution for proper protein function. In other words: RNA editing mainly re-establishes sequences on RNA level that could directly be encoded as such in the DNA (and in fact often are so in related species). Significant progress, however, has come lately with the identification of specific RNA-binding PPR (pentatricopeptide repeat) proteins providing sequence recognition specificities to determine nucleotide positions for editing in organelle transcripts (14–17).

It seems well supported that RNA editing, which is absent in algae, arose with the emergence of the earliest land plants. Now it is universally distributed among plants with the unique exception of the marchantiid liverworts where it appears to be secondarily lost (18,19). Correspondingly, RNA editing varies widely in appearance and frequency, ranging from zero sites in the marchantiid liverworts, over only 11 in the mitochondrial transcriptome of the model moss *Physcomitrella patens* (20), to some 200–500 sites in flowering plant mitochondria (21–27).

A further mystery in the evolution of RNA editing during 500 million years of plant diversification concerns the direction of pyrimidine exchanges. The initially discovered type of cytidine-to-uridine editing, most likely a simple deamination (28–30), is the dominating or even exclusive form of editing in angiosperms, mosses and also in the jungermanniid liverworts (which show editing

*To whom correspondence should be addressed. Tel: +49 228 73 5507; Fax: +49 228 73 6467; Email: mail@felixgrewe.de

in contrast to their marchantiid sister group). RNA editing in the reverse direction, which converts uridines into cytidines, has been discovered early (31,32) but is obviously very rare among the flowering plants. Other plant clades such as the hornworts and the ferns, however, show substantial additional 'reverse' U-to-C editing, incompatible with a simple biochemical deamination step (18,33–36).

We have recently analyzed the mitochondrial DNA sequence of a lycophyte, the quillwort *Isoetes engelmannii* (37). Extant lycophytes represent the most ancient lineage of vascular plants and are the sister group to all other tracheophytes. Among other peculiarities, the *I. engelmannii* mtDNA sequences seemed to require a substantial amount of RNA editing, possibly even exceeding 1500 sites, to correct sequences of its 24 encoded proteins (plus one intron-encoded maturase). First cDNA sequence analyses supported this hypothesis. In particular, *I. engelmannii* mt sequences seemed to require substantial amounts of reverse U-to-C exchanges in addition to numerous C-to-U edits. Furthermore, editing appeared also to be required for generating intact tRNAs. Additionally, the *I. engelmannii* mtDNA is characterized by particularly small intron sequences.

Here, we describe the results of an exhaustive analysis of mitochondrial transcripts in *I. engelmannii* which shows the hitherto most extensive degree of RNA editing observed among plants affecting mRNAs, tRNAs and intron sequences. Moreover, the data strongly support that (i) RNA editing mainly acts before other forms of RNA maturation such as splicing or tRNA processing; (ii) partial editing at moderate to high-frequency reflects inefficient editing; and (iii) partial editing at particularly low frequency indicates mis-operation of the RNA editing machinery. Also, we verifed splicing of the particularly small introns in the *I. engelmannii* mtDNA and discuss the concomitant secondary loss of introns and neighboring RNA editing sites by retro-processing.

MATERIALS AND METHODS

Plant material and RNA isolation

Isoetes engelmannii plant material originally collected in South Central Indiana (USA) by Jerry Gastony, and subsequently greenhouse cultivated, was kindly made available through Jeff Palmer and Erin Badenhop (Bloomington, IN, USA) and further cultivated at the Botanical Garden of the University of Bonn. Total RNA was isolated from whole tissue using the TRI-Reagent (Sigma-Aldrich, Steinheim, Germany) or the NucleoSpin RNA Plant Kit (Macherey Nagel, Düren, Germany) and treated with DNase I (Fermentas, Burlington, ON, USA) for 30 min at 37°C to remove contaminating genomic DNA.

Molecular work

cDNA was synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas) in the presence of random hexamer primers as specified by the manufacturer. Oligonucleotide pairs (all sequence information available from the authors upon request) were designed to anneal to 5′- and 3′-untranslated gene regions (UTRs). Gene specific cDNA products were amplified by PCR according to the GoTaq protocol (Promega, Madison, WI, USA). A GeneAmp PCR System 2700 (Applied Biosystems, Foster City, CA, USA) with annealing temperatures between 50°C and 55°C was used. Amplicons were recovered from agarose gels using the NucleBond Xtra Midi EF Kit (Macherey Nagel) and cloned into pGEM T Easy vector (Promega). For analysis of processed tRNA species, a transcript end mapping protocol was used (38). Total *I. engelmannii* RNA was ligated by T4 RNA ligase (New England Biolabs, Ipswich, MA, USA) and cDNAs were synthesized with RevertAid First Strand cDNA Synthesis Kit (Fermentas), in presence of 200 pmol of a reverse oriented primer. RT–PCR across the RNA-ligation site was done with the same reverse primer in combination with an upstream forward-oriented primer according to the BD Advantage 2 protocol (BD Bioscience, Franklin Lakes, NJ, USA) with annealing temperatures at 47°C.

Sequence handling and analyses

On average 16 independent cDNA clones were sequenced for each locus to estimate the degree of partial editing. Partial editing was considered authentic when found in at least two independent cDNA clones. For cross-validation, selected RT–PCR amplicons were sequenced directly (Supplementary Figure S1). All cDNA sequences were deposited in the database under accession numbers HQ616410-HQ616434 with editing sites annotated using the recently proposed nomenclature (20,39). Sequence handling and sequence alignment were essentially done using the alignment explorer of the MEGA software (40). Prediction, analysis and graphic display of RNA editing was done with the PREPACT software (39) and display and shading of sequence alignments was done with GeneDoc (http://www.nrbsc.org/gfx/genedoc/).

RESULTS

Massive mRNA editing characterizes the *I. engelmannii* mitochondrial transcriptome

To allow amplification of complete coding sequences, RT–PCR primers were designed to anneal in the respective 5′- and 3′-flanking UTRs. This has the added benefit of reducing the risk of introducing a bias for differentially edited transcripts by accidentally targeting edited sequence regions (25). To exclude potential DNA contamination, we strived for amplification across introns to select for spliced RNAs whenever possible, and performed control PCR assays without reverse transcriptase.

RNA editing was identified for all *I. engelmannii* mitochondrial protein-coding genes without exception (Table 1). In total, 1782 sites of pyrimidine exchanges in both directions were identified in messenger RNAs, which change 1406 codon meanings—the highest number of editing sites reported so far for a plant mitochondrial or chloroplast transcriptome. About 1/7th of these events (222 of 1782) are U-to-C changes.

Table 1. Summary of editing sites detected in the mRNAs of the 25 protein encoding genes of the *I. engelmannii* chondrome

Gene	Size [bp]	Codon changes by C > U editing														C-to-U		Codon changes by U > C editing														U-to-C		Total AA changes	Total sites	Creation of	
		A>V	H>Y	L>F	P>F	P>L	P>S	Q>*	R>*	R>C	R>W	S>F	S>L	T>I	T>M	Total	Silent	V>A	Y>H	F>L	F>P	L>P	S>P	*>Q	*>R	C>R	W>R	F>S	L>S	I>T	M>T	Total	silent			Start	Stop
***atp*1**	**1545**	5	9	1	3	18	12	0	0	5	0	5	20	8	4	104	10	3	0	0	0	0	1	12	3	5	0	0	1	2	0	27	0	117	131	Yes	No
***atp*4**	**588**	0	0	0	0	0	0	1	0	0	0	1	0	0	0	6	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6	No	No
***atp*6**	**762**	3	4	1	7	12	5	0	1	3	1	9	15	3	3	83	9	1	1	0	0	1	0	3	2	2	0	0	1	1	0	12	0	79	95	Yes	Yes
***atp*8**	**489**	0	3	0	1	6	1	0	0	1	1	0	6	1	1	31	9	0	0	0	0	0	0	3	1	0	0	0	0	0	0	4	0	25	35	Yes	No
***atp*9**	**225**	1	0	1	2	4	1	1	0	0	0	4	4	2	2	30	6	3	0	0	0	0	0	0	1	0	0	0	0	0	0	4	0	26	34	Yes	Yes
cob	**1170**	4	11	6	4	17	1	0	0	1	6	8	13	2	1	99	20	3	3	0	0	4	3	4	1	1	0	0	2	1	0	22	0	96	121	No	No
***cox*1**	**1632**	0	5	5	2	16	4	0	0	2	6	11	11	2	3	88	18	4	4	1	0	0	2	2	1	2	0	2	1	1	0	22	1	87	110	Yes	No
***cox*2**	**753**	0	0	2	0	1	0	0	0	1	2	0	1	0	1	10	2	0	0	0	0	0	0	3	0	1	0	0	0	0	0	4	0	12	14	Yes	No
***cox*3**	**792**	3	3	1	8	17	6	1	0	1	9	10	5	2	6	98	18	0	1	0	0	0	0	1	0	1	0	0	0	0	0	3	0	75	101	Yes	Yes
***nad*1**	**987**	1	1	1	2	8	5	1	0	1	3	1	7	1	5	48	9	0	0	0	0	1	2	3	1	1	0	1	0	0	0	9	0	46	57	Yes	No
***nad*2**	**1464**	0	8	5	2	13	8	1	0	8	1	8	12	1	4	94	20	0	0	0	0	1	2	2	0	0	0	0	1	0	0	6	0	77	100	No	Yes
***nad*3**	**357**	0	2	0	1	6	3	1	0	2	1	12	8	3	0	53	13	0	0	1	0	0	0	0	0	0	0	1	0	1	0	4	0	42	57	No	Yes
***nad*4**	**1488**	2	8	1	9	23	6	1	0	5	5	13	17	3	6	140	30	0	2	1	0	2	3	2	2	0	0	1	2	0	0	16	0	114	156	Yes	Yes
***nad*4L**	**303**	0	0	0	0	2	4	1	0	0	0	3	6	2	4	26	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	23	27	Yes	Yes
***nad*5**	**1986**	2	7	6	7	15	13	0	0	8	8	19	25	4	7	152	22	1	1	0	0	0	0	3	3	0	0	0	3	2	0	14	1	134	166	No	No
***nad*6**	**627**	2	1	1	4	11	2	0	0	1	3	1	8	2	5	58	12	0	1	0	0	0	1	1	0	0	0	1	1	0	0	6	1	46	64	Yes	No
***nad*7**	**1191**	5	7	0	2	16	3	1	0	5	2	8	13	5	8	95	18	0	4	1	0	2	0	5	5	1	0	0	0	1	0	20	0	94	115	No	Yes
***nad*9**	**558**	1	1	1	3	2	2	0	0	1	5	4	8	1	3	40	5	1	0	0	0	2	0	3	2	0	0	0	2	0	0	10	0	42	50	Yes	No
***rpl*5**	**549**	2	0	0	0	3	1	0	0	1	0	1	3	1	1	23	10	1	0	0	0	1	0	0	1	0	0	1	1	0	0	7	2	18	30	No	No
***rps*2**	**699**	0	0	0	0	2	2	0	0	3	0	0	8	0	0	23	8	0	0	0	0	1	1	1	0	0	0	1	0	1	0	5	0	20	28	No	No
***rps*3**	**1200**	2	4	0	1	5	4	0	0	3	2	6	8	1	0	56	19	0	0	0	0	0	0	4	2	0	0	0	0	0	0	6	0	42	62	No	No
***rps*4**	**765**	2	4	0	0	12	7	0	0	1	1	5	5	3	0	48	8	0	0	0	0	0	0	3	2	0	0	0	0	1	0	7	1	46	55	No	No
***sdh*3**	**297**	0	2	1	0	4	4	1	0	1	0	1	2	1	0	21	4	0	0	0	0	0	0	1	0	0	1	0	0	0	0	2	0	19	23	No	Yes
***tat*C**	**756**	0	4	3	2	11	8	0	0	4	3	8	3	4	0	62	10	0	0	0	0	1	1	2	0	1	0	0	1	0	0	6	0	56	68	No	No
mat*	**2019**	1	8	0	0	24	8	0	0	3	3	3	6	5	2	72	9	1	0	0	0	0	0	3	1	0	0	0	0	0	0	5	0	68	77	Yes	No
Partial		1	7	4	0	16	9	2	0	3	1	12	15	10	3	282	198	2	0	2	0	1	0	16	6	0	0	1	2	2	0	38	6				
Full		35	85	32	60	232	101	8	1	58	61	129	199	47	63	1278	99	16	17	2	0	16	16	45	22	15	1	7	14	9	0	184	0				
Total		36	92	36	60	248	110	10	1	61	62	141	214	57	66	1560	297	18	17	4	0	17	16	61	28	15	1	8	16	11	0	222	6	1406	1782	12	9

Codon conversions resulting from C-to-U and U-to-C editing that were observed in the complete coding sequences are listed separately from the respective silent editing events. Additionally indicated are editing events causing the creation of start and stop codons, and the fraction of sites for which either full or partial editing was observed in the respective cDNA pools. The total editing counts contain number of amino acid changes, silent editing and multiple editings affecting single codons (S>L [3], L>S [9], L>P [1], P>F [60]). The atp9i87 maturase (mat) is translated in frame with the two upstream *atp*9 exons and the asterisk indicates that editing sites in these exons were not included here but only counted once for the line labelled *atp*9.

Interestingly, the fraction of silent RNA editing events leaving the encoded amino acid unchanged is far lower for the reverse U-to-C edits (six of 222, 2.7%) when compared to silent events among the C-to-U editing events (297 of 1560, 19%). A detailed listing of all editing sites following a recently proposed nomenclature (20,39) is given in Supplementary Table S1. This nomenclature is composed of the name of the respective gene, followed by an 'e' (for editing), the respective nucleotide introduced by the editing event (U or C), the nucleotide position in the transcript (with position 1 corresponding to the A of the AUG start codon) and finally the resulting amino acid change e.g. nad5eU22PS. Further qualifiers may be added such as 'p' for partial editing as we will discuss below.

All predictable cases for reconstitution of appropriate AUG start codons from ACG threonine codons (12 cases) and of stop codons introduced through conversion of glutamine or arginine codons (nine cases) were confirmed (Table 1). Both ends of the reading frame are actually introduced by editing in five genes: *atp*6, *atp*9, *cox*1, *nad*4 and *nad*4L. Conversely, a total of 89 U-to-C editings are necessary to remove genomically encoded stop-codons within reading frames to recover 61 conserved glutamine and 28 arginine codons, respectively (Figure 1). In fact, only three of the 25 protein coding sequences—the small *atp*4, *nad*3 and *nad*4L genes—are free of in-frame stop codons whereas between one and up to 15 (in *atp*1) have to be removed at the transcript level in all others (Table 1). These conversions constitute a major share (42%) of the U-to-C type of editing events. In contrast, the recreation of leucine codons is the major effect of C-to-U editing with 462 codon changes (38.7%) altering either genomic proline or serine codons (Figure 1).

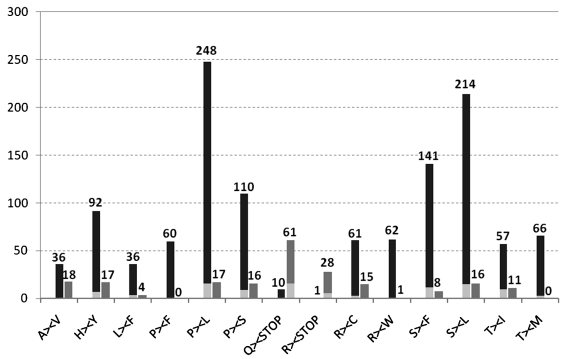


Figure 1. Bar chart displaying the numbers of observed codon changes introduced through RNA editing in the *I. engelmannii* mitochondrial transcriptome. Dark gray bars display C-to-U and light gray bars represent the 'reverse' U-to-C exchanges. The numbers of conversions in the C-to-U direction is significantly higher for the 12 possible codon sense changes introduced by pyrimidine transitions with ratios ranging from 2:1 for A–V to 62:1 for R–W codon sense changes. Reverse U-to-C editing in contrast dominates in conversions involving stop codons. Lighter colours indicate the fraction of codon conversions in each type, for which partial editing was observed.

The numbers of editing sites per particular transcript varies widely. Most affected by RNA editing is the (large) *nad*5 mRNA, where 152 pyrimidine changes effect 133 codon changes (Figure 2A). The highest density of editing sites, however, is reached in the (small) *nad*3 mRNA, in which 57 out of 357 nucleotides are edited (Figure 2B). On the other end of the scale, only six edited sites were identified in *atp*4 mRNAs (Figure 2C). Among these, only one site (atp4eU62SF) is efficiently edited, four sites are partially edited silent sites (see below) and one partial editing event erroneously introduces a stop codon (atp4eU235Q*). This exceptionally deviant pattern of RNA editing in *atp*4 may suggest that the mitochondrial gene copy is evolving into a pseudogene to be replaced by a functional nuclear copy.

RNA editing of the single intron-encoded ORF in *Isoetes* mtDNA

Introns in the *I. engelmannii* mtDNA are generally very small, irrespective of the class they belong to (groups I or II) and do not carry intron-encoded open reading frames (ORFs) for endonucleases or maturases, respectively. The unique exception is group II intron atp9i87g2, which carries maturase sequence similarities with its *Marchantia polymorpha* homolog. Two overlapping intron regions were RT–PCR-amplified to gain complete sequence insights on potential functional significance of this maturase homology on cDNA level. Indeed, a conserved reading frame with a size of 2019 bp (672 amino acids) was found to be reconstituted by RNA editing. Altogether, 81 codons in the maturase sequence are affected, including conversion of four stop codons into one arginine and three glutamine codons (Figure 2D). The majority of editing events in the *I. engelmannii* atp9i87g2 maturase is located in conserved maturase domains and significantly increases sequence similarity with the *Marchantia* homolog (Figure 3). Similar to the situation in *Marchantia*, the maturase ORF in atp9i87g2 is translated in-frame with upstream *atp*9 exon(s). Hence, apart from RNA editing, splicing of the upstream *atp*9 intron atp9i21g2 not present in *Marchantia* (Figures 2D and 3) would be an additional prerequisite for proper maturase translation in *I. engelmannii*. In addition to the editing sites within the atp9i87g2 maturase reading frame, we could identify two more editing sites in the downstream domains V and VI of the group II intron secondary structure (see below).

Partial RNA editing at 320 sites in the *I. engelmannii* mitochondrial transcriptome

Although operating very efficiently at most sites in plant organelles, RNA editing is not a yes/no process. The analyses of independent cDNA clones or of sequence data generated by direct sequencing of RT–PCR amplicons can lead to the discovery of partially edited transcripts. To account for recognized partial editing in the recently proposed RNA editing nomenclature (39) we have introduced the additional qualifier 'p'. With sufficient cDNA data available the 'p' qualifier may be followed by the percentage of cDNA clones reflecting a

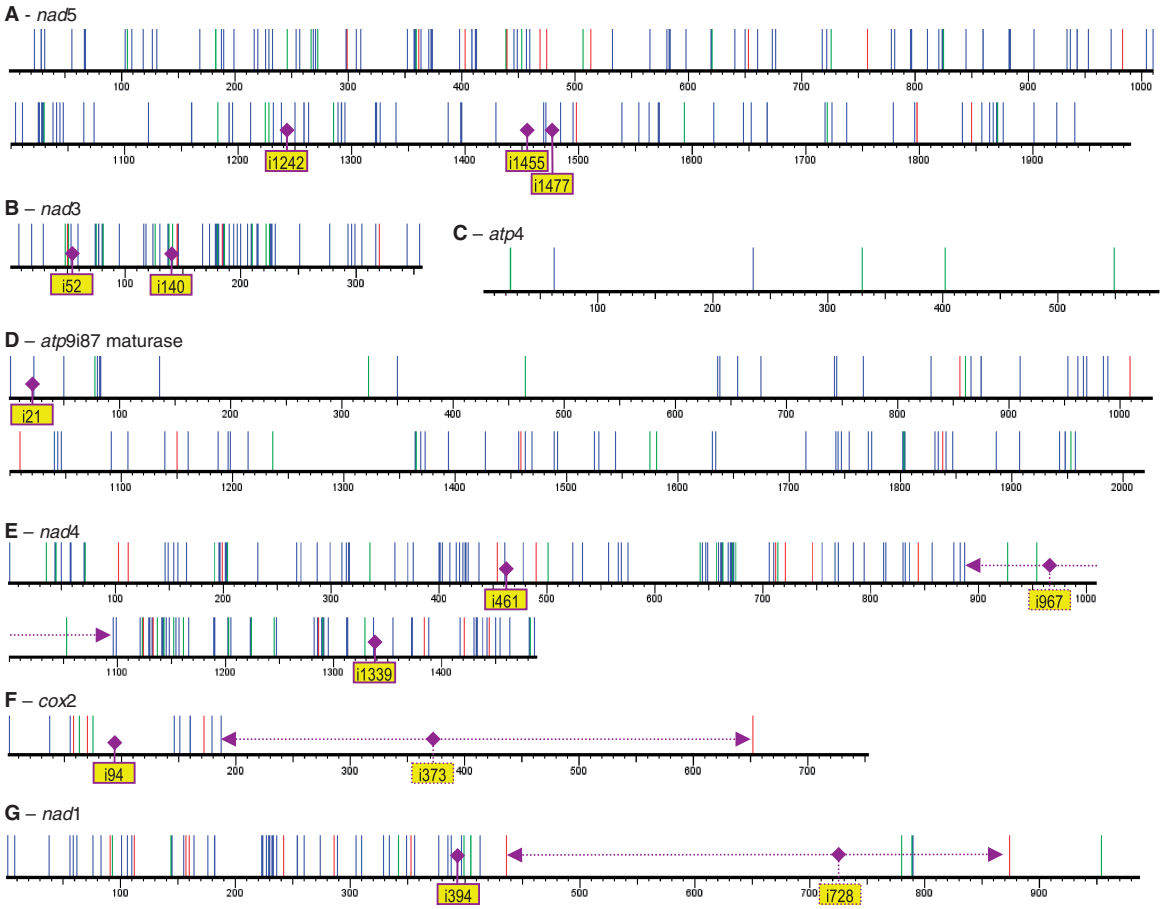


Figure 2. RNA editing sites in *I. engelmannii* mitochondrial genes displayed with the default graphic tool options of PREPACT (39). Non-silent C-to-U and U-to-C editings are indicated by blue or red lines, respectively, and green lines indicate silent editings. RNA-editing patterns are exemplarily shown for selected genes *nad*5 (**A**), *nad*3 (**B**), *atp*4 (**C**), the maturase in atp9i87g2 (**D**), *nad*4 (**E**), *cox*2 (**F**) and *nad*1 (**G**). Group II intron insertion sites are indicated, those shown with stippled lines are absent in *Isoetes*, coinciding with regions lacking RNA-editing sites (horizontal lines with arrowheads). The atp9i87 maturase reading frame is in frame with the first 29 codons of *atp*9 which contain the upstream group II intron atp9i21.

given editing event. To estimate the extent of partial editing we generally sequenced several independent cDNA clones (on average 16 per locus). Partial editing sites were considered as verified only when deviant pyrimidines were determined at least twice independently in the respective cDNA population (i.e. above the threshold of occasional PCR-derived sequence errors). Independent cDNA clone sequencing versus RT–PCR product bulk sequencing has the benefits of higher sequence qualities and, more importantly, allows to detect editing site interdependence among the diverse cDNA-editing patterns (as discussed below for the *nad*4 case). However, independent cDNA clone sequencing comes at the risk of cloning bias. For comparison, we sequenced several selected RT–PCR amplicons (for *nad*1, *nad*3, *nad*4 and *nad*7) directly without observing significant discrepancies in comparison to populations of cDNA clones, as exemplarily outlined for *nad*3 (Supplementary Figure S1). All events of full (100%) editing were equally detected as such both in direct RT–PCR product and cDNA-clone sequencing. Similarly, no further evidence for any editing was found by direct amplicon sequencing that was not equally reflected in cDNA pool sequencing. In contrast, however, only some of the partial editing events of particularly low (<10%) or high (>90%) efficiency identified in the cDNA populations were adequately reflected in direct RT–PCR sequences, whereas others would have been missed in the latter approach. Given that such sites are equally detected in spliced and unspliced cDNA clones (see below for comparison of mature versus immature *nad*7 transcripts), we consider the cDNA clone sequences to faithfully and better reflect the RNA-editing status of transcripts. Certainly the *P*-value percentages may suggest higher precision of determining partial editing efficiencies than actually given before some 100 independent cDNA clones are sequenced. As an alternative for such

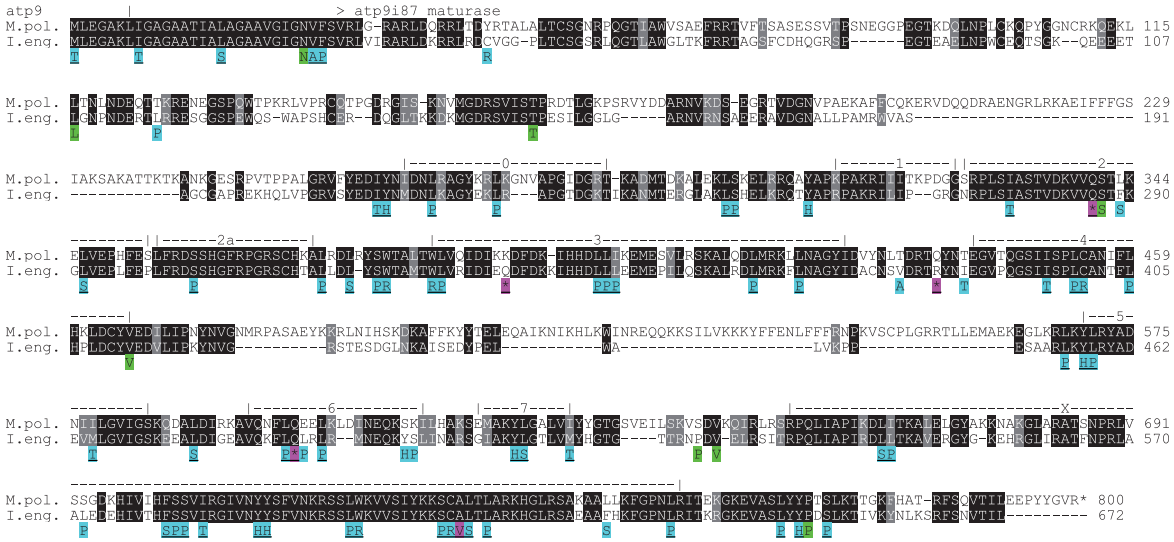


Figure 3. Sequence alignment of the *I. engelmannii* maturase protein sequence in intron atp9i87g2 and its homolog from *M. polymorpha*. The *Isoetes* cDNA-derived protein sequence results from RNA editing of 81 DNA-encoded codons as shown below the alignment. Background shading colours indicate C-to-U (cyan), U-to-C (magenta) and silent edits (green). Codon changes increasing sequence similarity with *Marchantia* are underlined and are mostly located within conserved maturase domains 0–7 and 'X' (56) as indicated. The atp9i87 maturase is translated in frame with the small upstream exon(s) of *atp*9 across the splice donor site (arrowhead). Upstream intron atp9i21g2 (vertical line) is present in *Isoetes*, but not in *Marchantia*.

pseudo-precise values five stages of editing efficiency could alternatively be distinguished, e.g. pVH, very high (>80%); pHI, high (>60%); pME, medium (40–60%); pLO, low (<40%) and pVL, very low (<20%).

In total, 320 out of the 1782 editing sites in mRNAs were identified as partially edited. A full 63.8% of these partial editing events (204/320) affect silent editing sites, leaving codon identities unchanged (Figure 4). In contrast, only 6.8% of the fully edited sites (99/1462) affect silent positions. Partial editing at silent sites is particularly ineffective: 111 of the 206 partially edited silent sites are edited in <25% of sequenced cDNA clones (Supplementary Table S1).

The remaining, non-silent partial editing sites affect 115 codon changes (Figures 1 and 4). On average, partial editing affects 8.7% of codon conversions of a particular type. Individual exceptions with higher degrees of partial editing are the C-to-U type threonine into isoleucine codon conversions (10/57 = 18%) and the U-to-C editings converting phenylalanine into leucine (2/4 = 50%), stop into glutamine (16/61 = 26%) and stop into arginine (6/28 = 21%) codons (Figure 1). In some instances partial editing may (to generally low degrees) lead to mis-conversions of evolutionarily conserved codons on DNA level or erroneously introduce stop codons: atp6eU182SFp13, atp6eU430HYp13, nad1e U790Q*p13, atp4eU235Q*p32. The latter case, in particular may indicate an ongoing degeneration of the *I. engelmannii atp*4 gene into a pseudogene as mentioned above. Such events with low efficiencies of partial editing and the high proportion of silent sites among the partial editing events support the idea that a majority of those

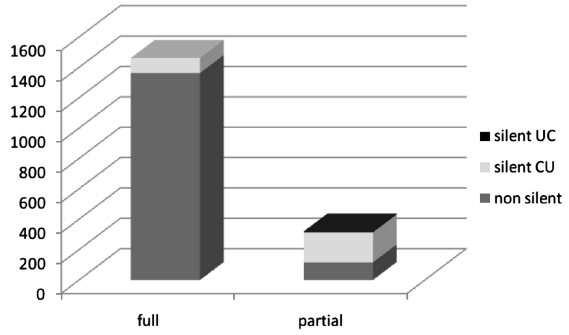


Figure 4. Bar chart displaying the fractions of silent editings for both directions of pyrimidine conversions among fully (left) versus partially (right) edited sites. All six silent U-to-C editing sites are partially edited (right).

reflect collateral misfiring of the editing machinery targeting other sites. In other cases, however, partial editing of relevant sites may simply reflect inefficiencies of the particular editing factors.

Different editing patterns among independent *nad*4 cDNA clones

We present the *nad*4 gene as a typical model case for frequent RNA editing including numerous reverse, silent and partial RNA-editing events. With only five pairs of identical sequences, 14 different editing patterns were revealed in a pool of 19 *nad*4 cDNA clones (Figure 5). In the case of the *nad*4 gene, 140 C-to-U and 16 U-to-C

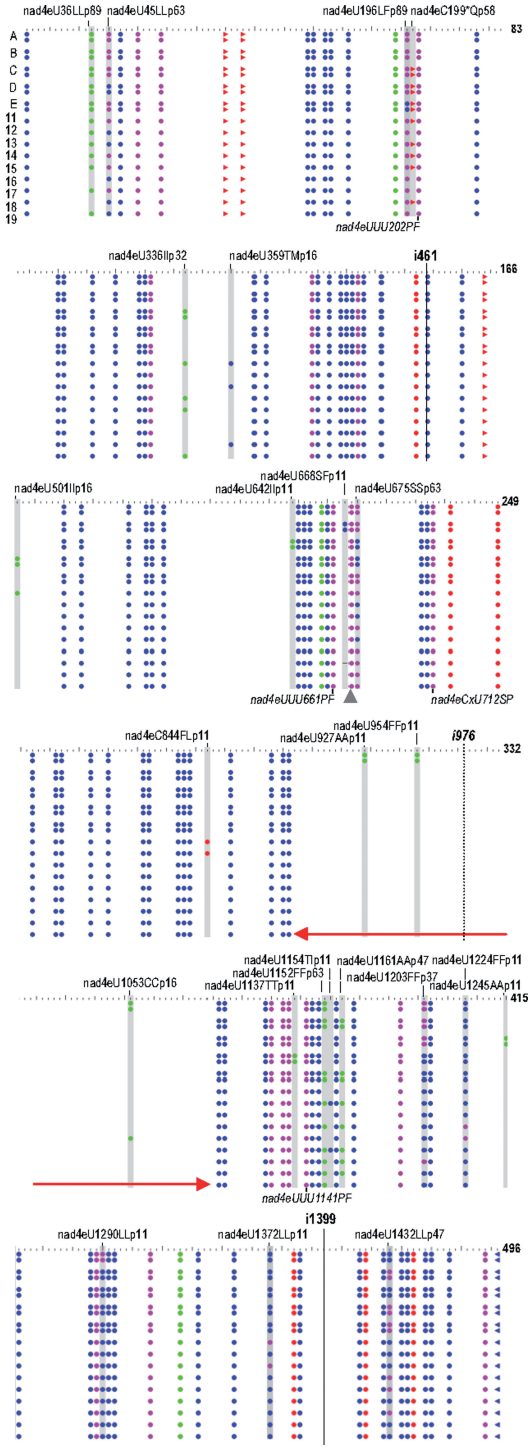


Figure 5. RNA-editing status in a population of 19 *nad*4 cDNA clones. The graphic is based on the cDNA variants overview of PREPACT (39) with default settings: blue and red circles indicate codon sense changes derived from C-to-U or U-to-C editings, respectively, green circles indicate silent codon changes and purple circles reflect more than one editing in a codon. Forward arrows indicate stop codon removal in codons 35, 38, 67, 164 and 496 while reverse arrows indicate stop-codon creation at the end of the reading frame. Positions of partial editing are highlighted with gray columns and the labeling for the respective editing events are indicated. With five pairs of identical cDNA sequences (A–E) each showing identical partial editing patterns, a total of 14 different editing patterns is recognized. Group II intron insertion sites nad4i461 and nad4i1399 (solid lines) in *I. engelmannii* mtDNA are indicated, intron nad4i976 (dotted line) is absent but conserved in other taxa. The intron loss coincides with absence of editing sites over an extended *nad*4 region (red arrows) which may be the result of cDNA retroprocessing.

editing sites lead to 114 codon changes (Figure 1E). Both the start and the stop codon have to be introduced by RNA editing to define the correct *nad*4 reading frame ends and four genomically encoded stop-codons have to be eliminated from the reading frame to avoid early termination of translation (Table 1). Additionally, 30 silent editing sites are located at first or third triplet positions not influencing the encoded protein (Supplementary Table S1). The majority of partial editing sites (19 of 25) affects silent sites to varying degrees (between 2 and 17 out of 19 clones): nad4eU36LLp89, nad4eU45LLp63, nad4eU675SSp63, nad4eU1152p63, nad4eU672FFp58, nad4eU1161AAp47, nad4eU1432LLp47, nad4eU1203FFp37, nad4eU336IIp32, nad4eU501IIp16, nad4eU1053CCp16, nad4eU642IIp11, nad4eU927AAp11, nad4eU954FFp11, nad4eU1137TTp11, nad4eU1224FFp11, nad4eU1245AAp11, nad4eU1290LLp11 and nad4eU1372LLp11 in descending order of editing frequencies. Three additional partial editing events are expected as they reinstate evolutionarily conserved amino acid codons: nad4eU196LFp89, nad4eC199*Qp58 and nad4eU359TMp16. Finally, three further sites of partial editing are unexpected as they introduce evolutionarily non-conserved amino acid codons. Notably, all three occur at low frequencies in only two out of 19 clones each: nad4eU668SFp11, nad4eC844FLp11 and nad4eU1154TIp11. Considering the above, only two of 19 clones (cDNAs #13 and #18) actually reflect 'proper' editing as predicted with editing at the three expected but not at the three unexpected partial sites (Figure 5). These cDNAs are also non-edited at most silent partial editing sites. On the other side of the spectrum, cDNA pair #A lacks several important codon conversions but in contrast shows silent editing at the infrequently, partially edited positions (Figure 5).

Concomitant loss of introns and RNA-editing sites

A total of 27 group II introns and three group I introns were predicted on the basis of the *I. engelmannii* mitochondrial DNA sequence (37). The cDNA analyses performed here showed that all 30 introns were correctly spliced as predicted in spite of their generally small sizes. As common for plant mitochondrial introns, many of them are shared with other land plant groups whereas some *Isoetes* introns were novel discoveries. In contrast, some ancient introns conserved in other plant clades (including

non-liverwort bryophytes and angiosperms) were absent in *I. engelmannii*: cox2i373, nad1i728 and nad4i967.

A lack of editing positions (except for the few partially edited silent sites) which are otherwise densely packed in *I. engelmannii* is immediately apparent in *nad*4 between amino acid positions 297 and 365, which corresponds to the gene region surrounding intron nad4i967g2 present in other plant species (Figures 2E and 5). In full congruence, two other extended gene regions in *cox*2 (Figure 2F) and *nad*1 (Figure 2G) lacking RNA-editing sites also perfectly coincide with the absence of introns in *I. engelmannii* which are present in other taxa: cox2i373 and nad1i728. Only one single U-to-C editing site (cox2eC652*Q) essential for the correction of a genomically encoded stop codon was identified in the *cox*2 3′-region, located >300-nt downstream of the intron cox2i373 insertion site in other species. Likewise, the closest upstream editing event, cox2eU187LF, is located in a distance of nearly 200 nt (Figure 2F). Similarly, editing sites are absent within some 300-nt upstream of the nad1i728g2 insertion site in other taxa up to editing site nad1eC436*Qp88 (Figure 2G). Interestingly, all sites downstream of the intron insertion site are irrelevant and/or weakly edited: nad1eU780SSp13, nad1eU789FFp13, nad1eU790Q*p13 and nad1eU954SSp13 with the unique exception of nad1eC874*R. Congruently only three silent sites are partially and rarely edited (nad4eU927AAp11, nad4e U954FFp11 and nad4eU1053CCp16) in proximity to the lost intron nad4i967 within the region ranging from base 888 to 1095 (Figure 2E). We conclude that simultaneous absence of RNA-editing sites and introns in these three gene regions in *cox*2, *nad*1 and *nad*4 is a result of recent partial retro-processing of mature mRNAs.

Editing in tRNAs

Modeling of the cloverleaf secondary structures of mitochondrially encoded tRNAs in *I. engelmannii* had revealed that RNA editing may also affect tRNAs to a much higher degree than previously observed in land plants (37). Altogether, 43 sites of C-to-U editings can be predicted in several crucial positions, which could reconstitute conserved uridine residues in the tRNA-consensus structure or improve base pairings in stem regions. The initial cDNA-sequence analysis of the tRNA for proline (*trn*P) had revealed that such editing sites could indeed be confirmed and that RNA editing takes place already in precursor transcripts with *trn*P still connected to the downstream *sdh*3 gene (37). Now, we first investigated whether the tRNA-editing status may differ between precursor versus processed tRNAs after 5′- and 3′-trimming and the addition of the CCA tail. To target the matured tRNAs, we used self-ligation across the tRNA-ends, followed by overlapping RT–PCRs. We succeeded in obtaining such fully matured, CCA-tailed tRNA sequences for tRNA–Q(UUG) and tRNA–W(CCA) in parallel to the cDNA sequences for the corresponding tRNA precursors with primers targeting sequences flanking the tRNAs (Figure 6A and B). No differences in RNA-editing status were observed, confirming that the RNA-editing machinery indeed targets tRNA precursor sequences before processing, as already suggested by the initial analysis of *trn*P.

The analysis of the editing status of *trn*Q revealed five editing sites exactly as predicted: four of these re-establish A–U pairings in the acceptor stem and in the pseudouridine stem and one re-establishes the conserved uridine residue in position 8 (Figure 6A). A further potential site of editing (position 21), however, remained unedited in the *trn*Q dihydrouridine stem. Similarly, four of five predicted editing sites were identified in *trn*W (Figure 6B). Base pairing mismatches are corrected in the acceptor, pseudouridine and anticodon stems and the conserved uridine in position 33 is re-established. However, a proximal A–C mismatch in the anticodon stem remained uncorrected. Interestingly *trn*Q and *trn*W feature unmatched bases in these positions in other taxa as well (e.g. A–A in the liverwort *M. polymorpha*). These mispairings may in fact be relevant to tRNA functionality such as appropriate amino acid charging. Analogous observations were made for the other tRNA species as well: three predicted sites were completely edited in *trn*K but the corresponding base mismatch in the proximal anticodon stem position remains unaltered (Figure 6C) and again, this is a base mismatch position also present in *Marchantia*. Similarly, anticodon position 36 of *trn*M is found edited as expected whereas the proximal acceptor stem base mismatch remains unchanged (Figure 6D) and positions in the dihydrouridine and the pseudouridine stem but not in the acceptor stem are edited in *trn*F (Figure 6E). In two further cases of tRNAs with one potential, predicted editing each (*trn*fM, *trn*G), we did not confirm the editing events in cDNAs. This is not astonishing for *trn*fM where the distal acceptor stem mismatch again is also present in *Marchantia* (Figure 6F), but all the more in the latter case, given that the conserved GUUC motif of the pseudouridine loop is absent in *trn*G (Figure 6G). Moreover, the corresponding position (55 in the tRNA-consensus structure) was found to be edited both in the previously analyzed *trn*P and in the now investigated *trn*C (Figure 6H). Investigation of editing in *trn*C also revealed the first-documented case of reverse U-to-C editing in a tRNA. Here, the weak U–G pair in the proximal position of the pseudouridine stem is converted into a stronger C–G base pair. Two further C-to-U edits efficiently reconstitute base pairings in the acceptor and anticodon stem of *trn*C as predicted (Figure 6H). Interestingly, three more positions of the *trn*C molecule reveal inefficient partial editing and two of these are also of the reverse U-to-C type. Of these three partial editing events, two are 'reasonable' in the sense that they could further improve base pairing in the pseudouridine stem. However, the event affecting position 56 would destroy the GUUC consensus motif. In spite of numerous different attempts we could obtain only partial cDNA information for *trn*Y (Figure 6I) and none for *trn*I (Figure 6J). Given the novel insights on U-to-C editing in *trn*C, we wondered whether similar events could take place in the *trn*Y pseudouridine stem, which is particularly rich in G–U pairs. We could indeed confirm a further and efficient reverse editing event in position 62 (Figure 6I), but were unable to retrieve cDNA covering the remaining four

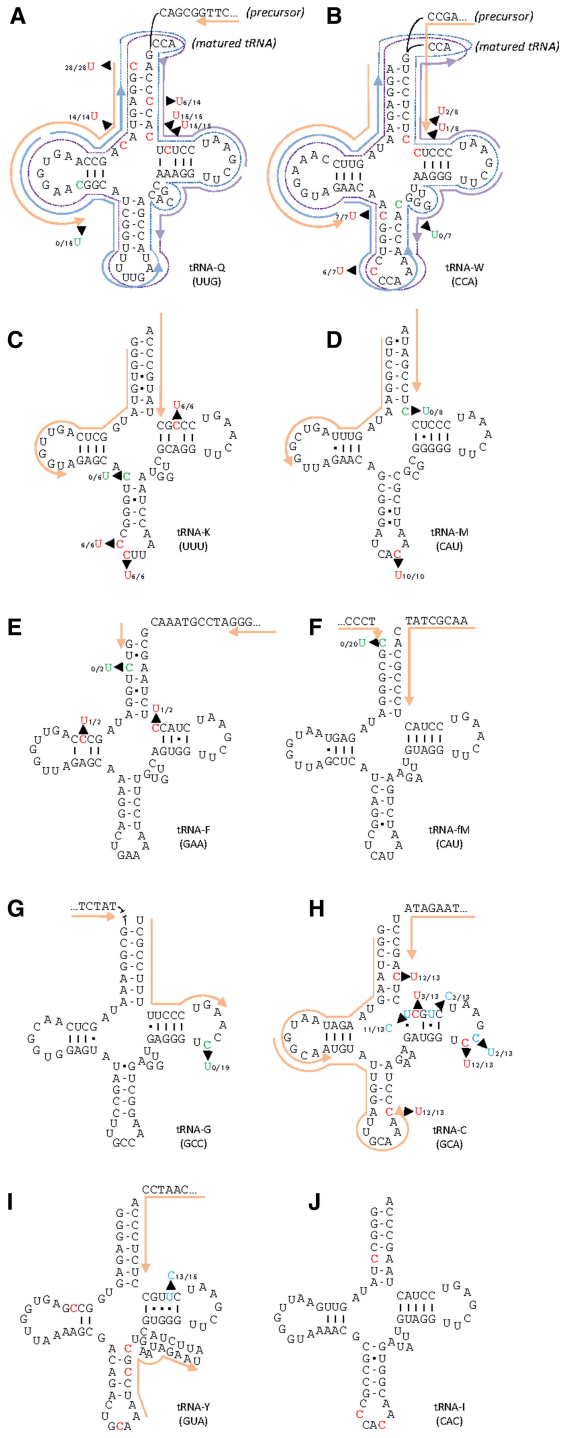


Figure 6. Outline of editing in 10 tRNAs. Nucleotides highlighted in red were found to be edited from C to U (**A–J**), those highlighted in blue were identified as edited from U to C in cDNA analyses with the numbers of edited and total cDNAs indicated for each site before and after the slash, respectively. Nucleotides shown in green were initially predicted to be editing sites but were found unchanged in all studied cDNAs. Primers successfully used in RT–PCR of precursor transcripts are shown in orange. No cDNA sequences could be retrieved for *trn*I (**J**) and the 5′-part of *trn*Y (**I**). Overlapping primer pairs used in the parallel amplification of self-ligated, circularized *trn*Q (**A**) und *trn*W (**B**) are indicated in blue and purple with the respective amplified regions shown with stippled lines.

predicted positions in *trn*Y or the three predicted position in *trn*I.

In summary, out of 43 predicted sites of tRNA editing, cDNA could not be retrieved for seven. Of the remaining 36 candidate sites, 29 were confirmed to reveal C-to-U editing. Additionally, four positions of U-to-C editing were discovered of which two are partially edited to a low degree. This suggests that partial/inefficient tRNA-editing mirrors the results observed for silent or mis-editing of mRNAs.

Mitochondrial intron editing in *nad7*

The finding of frequent tRNA editing as well as two RNA-editing events in intron atp9i87g2 downstream of its maturase prompted us to investigate more mitochondrial intron sequences systematically on cDNA level. To this end, we explored the most intron-rich *nad7* gene in *I. engelmannii* mtDNA with its four group II introns nad7i209, nad7i676, nad7i917 and nad7i1113. Here, RT–PCR approaches could be designed to ideally select for partially matured transcripts that remained unspliced for the respective intron under investigation but were spliced for other introns of the *nad7* gene. Overlapping intron amplicons were designed with one primer binding in the respective intron and one in a flanking or distant exon. The sequencing of such partially matured transcript cDNAs revealed editing sites in the flanking exons that had previously been determined in mature cDNAs as well as several editing sites in the intron regions. In total, we found 27 sites of RNA editing in the four *nad7* introns reliably determined as they were identified in more than one cDNA clone each. As observed in coding regions and tRNAs, the preferential direction of editing is C-to-U, counting 26 sites. Most intron editing sites were partially edited and only seven of the 27 editing sites were edited in all sequenced cDNA clones. Mapping the sites of RNA editing onto secondary structure models suggest that many of the editing events may actually improve RNA base pairings and might be a prerequisite for splicing as exemplarily shown for nad7i676 (Figure 7). RNA editing converts five A–C mispairings in domain I of nad7i676g2 into canonical A–U base pairs. However, one obvious A–C base mismatch in domain VI is not subject to editing. Interestingly, an editing site 9-nt downstream of the 5′-splicing site has recently been found to be mandatory for splicing (41) and the corresponding position is subject to RNA editing in nad7i676g2 in addition to three more sites involved in stem formation of domain I. At the basal stem of domain I, one G–U wobble base pair is modified to a more stable G–C Watson–Crick base pair by U-to-C editing—to our knowledge the first U-to-C intron editing

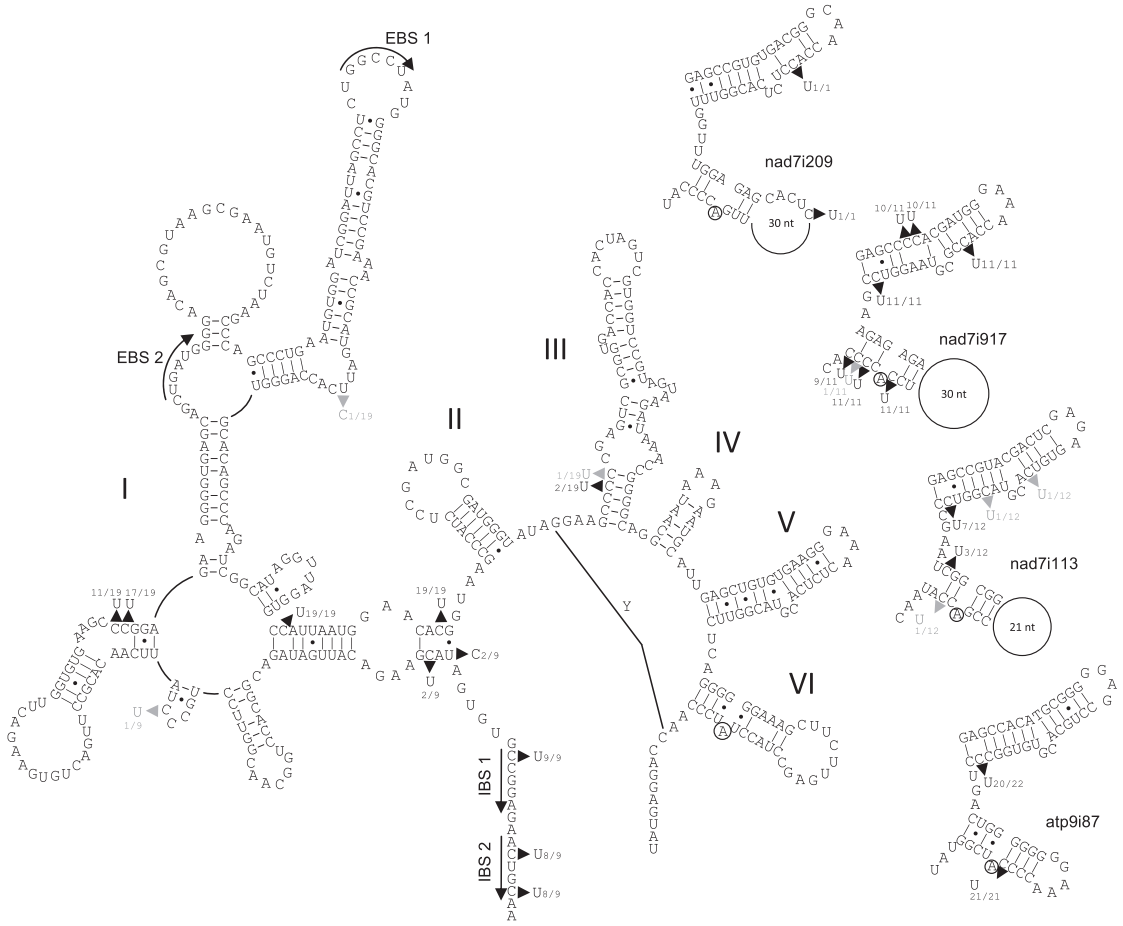


Figure 7. Left: complete secondary structure model of *I. engelmannii* group II intron nad7i676g2 following the proposed consensus structure (57). Despite its reduced size the essential characteristic features in the six conserved group II intron domains I to VI are present and classify nad7i676g2 as a member of subgroup IIB. The conserved bulged adenosine for lariat formation during splicing is encircled. Selected tertiary interactions EBS-IBS and γ–γ' are indicated. Arrow heads indicate editing sites identified in unspliced *nad7* cDNAs with numbers before and after the slash indicating the amounts of edited and total sequenced cDNA clones for each site. Right: conserved domains V and VI of introns nad7i209, nad7i917, nad7i1113 and atp9i87 respectively. Editing is indicated as described above.

observed. In contrast, stem stability of domain IV is loosened by RNA editing introducing a weaker G–U wobble base pairing. No editing was observed at exon binding sites (EBS). However, editing at three positions in intron binding sites (IBS 1 and IBS 2) in the upstream *nad7* exon introduce weaker wobble pairings. No editing events were seen in conserved intron domains V and VI of nad7i676. However, ten C-to-U editing sites located in intron stems and necessary to remove A–C mispairings were disclosed in the other *nad7* introns nad7i209, nad7i917 and nad7i1113 and this includes several sites in their highly conserved domain V and VI structures (Figure 7). Likewise the two editing sites in atp9i87 outside of its maturase are also located in these two domains. Only a single additional editing site discovered in the loop of domain VI of intron nad7i209 is obviously not involved in intron secondary structure stability (Figure 7).

The *nad7* dataset of immature, partially spliced cDNAs allows for a comprehensive comparison with the previously determined editing data from fully spliced cDNAs (Figure 8). The data clearly show that many non-silent sites are already edited fully or to a considerable extent in unspliced pre-mRNAs. Notable exceptions are nad7eU277RC and the stop codon removals like nad7eC82*R, which is significantly less edited in unspliced versus spliced cDNAs. Likewise, many silent edits are identified less frequently in unspliced versus spliced transcripts. This finding supports the idea that such silent editings may represent results of mis-targeting that become more apparent with transcript age in the mitochondria.

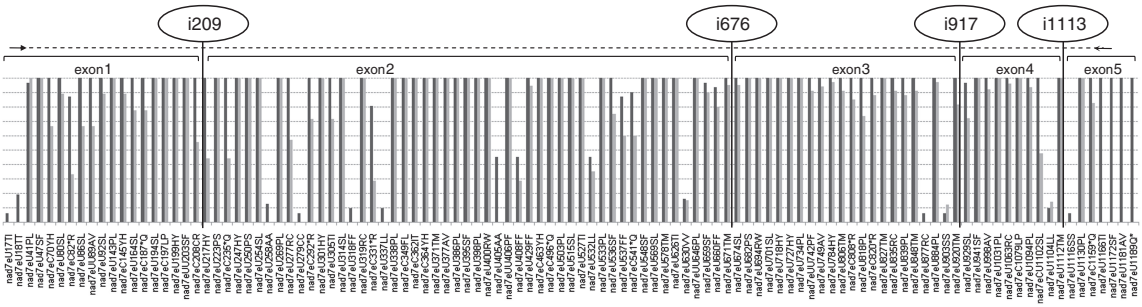


Figure 8. Comparison of RNA-editing status for *nad7* exon sites with bars indicating percentage of editing among spliced (left, dark gray) and partially spliced immature cDNAs (right, light gray). Intron-insertion sites and the overall extension of the exon region covered for partially matured transcripts (dotted line) are indicated.

DISCUSSION

Isoetes engelmannii has now surpassed all other plant taxa in frequency of reported RNA editing, including the gymnosperm *Cycas taitungensis*, for which close to 1000 sites were predicted (39,42) and nearly 600 confirmed among 25 of its 39 genes that were recently analyzed (43). Why and how plant organelle RNA editing has emerged in the first place still remains mysterious. General speculations on benefits of editing include a gain in gene expression variability or a compensating mechanism for mutations. However, in our opinion, strong conclusive support has never been found for any of these possible explanations. In any case, they are challenged by the striking variability of editing patterns and frequencies among land plants, which suggest an (occasionally very fast) coming and going of editing sites through the course of evolution. With this regard, we consider the editing data for the bizarre mitochondrial transcriptome of *I. engelmannii* a strong case in point. The maintenance of genetic machinery specifically recognizing more than 1800 editing positions, the overwhelming majority of which restores proper mRNA, tRNA and intron sequences, seems unlikely to be of regulatory benefit. It is hardly conceivable that more than 100 editing sites in a single mRNA offer important new dimensions in modulation of protein activity.

Among all hypotheses on evolutionary origins and pathways that have been put forward (44–47), we consider those most likely which simply assume merely neutral evolution of a vast molecular rococo in the genetic playground of endosymbiotic organelles. An important recent contribution in this regard highlights the convergent pathways in the evolution of peculiar phenomena including RNA editing in Euglenozoa and Dinoflagellates (48).

Much progress has recently been made to understand the underlying mechanisms of plant organelle RNA editing, notably with the identification of several protein factors specifically targeting editing sites. After the seminal discovery of a particular pentatricopetide repeat (PPR) protein necessary for an editing event in the *Arabidopsis thaliana* chloroplast transcriptome (14), a similar factor was discovered for mitochondrial editing (15) among several others in both organelles (16,49).

The vast amount of editing and the screening of multiple independent cDNA clones for *I. engelmannii* gives a conclusive picture on the issue of partial, silent, irrelevant or non-beneficial editing: while some (efficient) partial editings suggest minor inefficiency of the underlying mechanisms, other (inefficiently) partially edited sites suggest unspecific binding of PPR-proteins actually targeting other important editing sites. In one case, we assume the RNA-editing pattern to reflect an emerging pseudogene after a likely gene transfer to the nucleus. The *I. engelmannii atp*4 gene is characterized by low-level editing with silent sites dominating and editing event atp4eU235Q*p32 in fact de-functionalizing the gene's transcript. Most interestingly, *atp*4 appears to be actually missing from the mitochondrial gene complement of *Isoetes*' sister genus *Selaginella* (J. Hecht *et al*. unpublished data).

Obviously, editing sites can be gained but they can also be lost subsequently. The here reported cases of larger scale editing site losses in the environment of lost introns in three genes strongly support the idea of retro-processed mature mRNAs via reverse transcriptase mechanisms. Interestingly, the three group II introns in question which are lost from the *I. engelmannii* mtDNA (cox2i373, nad1i728 and nad4i967) are known to be lost independently among angiosperms, too. Yet more noteworthy, only one single maturase (in atp9i87) could possibly provide the necessary reverse transcriptase activity for retroprocessing in *Isoetes* mitochondria with its otherwise tiny introns. The much larger number of intron-encoded maturases in *M. polymorpha* (50) may actually be the cause of complete absence of RNA editing in this liverwort by providing more extensive retroprocessing.

A further enigma concerns the occurrence of reverse U-to-C editing accompanying the classic C-to-U editing. After the early discoveries of rare U-to-C edits (31,32), similar events have been reported very rarely in flowering plants and none have been confirmed for mosses and liverworts. In contrast, significant amounts of reverse U-to-C editings can be identified in hornworts and ferns

(34–36,51). Given the now well-corroborated sister-group relationship of hornworts and vascular plants (52,53), the occurrence and rise in frequency of reverse U-to-C editing in plant evolution may actually be connected to the common ancestor of hornworts and early tracheophytes after the split from liverworts and mosses. In any case, base deamination can certainly not be the biochemical mechanism behind U-to-C editing and (trans-)amination mechanisms rather need to be looked for. Hence, it will be highly exciting to see the first protein factor relevant for a U-to-C type of editing event identified. The so far 'non-model' taxa such as the hornworts, ferns or *Isoetes* will be the obvious organisms for investigation.

Finally and as conclusively shown above for *I. engelmannii*, it is interesting to find the same biochemical constraints defining the ratios of U-to-C versus C-to-U editing and the fractions of partial or irrelevant editings reflected in mRNAs, tRNAs and intron sequences. Similar to editings in mRNAs, which can be predicted as they reconstitute conserved codon identities, editing in the structured RNAs can be predicted as they reconstitute base pairings. In those instances where tRNA editing could have been assumed to take place in order to create base pairings but was not observed, the unedited state is most likely of functional relevance. In fact, examples have been reported, where RNA editing creates U–U mispairings in the *trn*C of dicots (54,55). Interestingly, we could not identify any editing sites in the mitochondrial rRNAs of *I. engelmannii*. In the light of ample editing seen in precursor tRNAs and unspliced mRNAs, this suggests that the RNA-editing machinery very preferentially acts on immature and single-stranded RNAs and that folding and processing of rRNAs, at least in *I. engelmannii* mitochondria, proceeds too fast for RNA-editing evolution to take hold in this type of RNAs as well.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to Dr Jeffrey Palmer and colleagues (Bloomington, IN, USA) for generously making fresh material of *I. engelmannii* available to us and to Dr Jeffrey Mower for comments on the *atp*9 maturase in *Isoetes*. We also wish to thank Henning Lenz for his work on the PREPACT tool and Julia Hecht for proofreading the article. Sequences obtained in this study were deposited in GenBank (accession numbers HQ616410-HQ616434).

FUNDING

Deutsche Forschungsgemeinschaft [DFG Kn411/6-1].

REFERENCES

1. Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.H. and Grienenberger,J.M. (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature*, **341**, 660–662.
2. Covello,P.S. and Gray,M.W. (1989) RNA editing in plant mitochondria. *Nature*, **341**, 662–666.
3. Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) RNA editing in plant mitochondria. *Science*, **246**, 1632–1634.
4. 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.
5. Karcher,D. and Bock,R. (2002) Temperature sensitivity of RNA editing and intron splicing reactions in the plastid *ndhB* transcript. *Curr. Genet.*, **41**, 48–52.
6. Bentolila,S., Chateigner-Boutin,A.L. and Hanson,M.R. (2005) Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in Arabidopsis. *Plant Phys.*, **139**, 2006–2016.
7. Zehrmann,A., van der Merwe,J.A., Verbitskiy,D., Brennicke,A. and Takenaka,M. (2008) Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of Arabidopsis thaliana. *Mitochondrion*, **8**, 319–327.
8. Bock,R., Hagemann,R., Kössel,H. and Kudla,J. (1993) Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids - a new regulatory mechanism? *Mol. Gen.Genet.*, **240**, 238–244.
9. Grosskopf,D. and Mulligan,R.M. (1996) Developmental- and tissue-specificity of RNA editing in mitochondria of suspension-cultured maize cells and seedlings. *Curr. Genet.*, **29**, 556–563.
10. Howad,W. and Kempken,F. (1997) Cell type-specific loss of *atp6* RNA editing in cytoplasmic male sterile *Sorghum bicolor*. *Proc. Natl Acad. Sci. USA*, **94**, 11090–11095.
11. Miyata,Y. and Sugita,M. (2004) Tissue- and stage-specific RNA editing of *rps14* transcripts in moss (*Physcomitrella patens*) chloroplasts. *J. Plant Physiol.*, **161**, 113–115.
12. Ruf,S. and Kössel,H. (1997) Tissue-specific and differential editing of the two *ycf3* editing sites in maize plastids. *Curr. Genet.*, **32**, 19–23.
13. Karcher,D. and Bock,R. (2002) The amino acid sequence of a plastid protein is developmentally regulated by RNA editing. *J. Biol. Chem.*, **277**, 5570–5574.
14. Kotera,E., Tasaka,M. and Shikanai,T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, **433**, 326–330.
15. Zehrmann,A., Verbitskiy,D., van der Merwe,J.A., Brennicke,A. and Takenaka,M. (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. *Plant Cell*, **21**, 558–567.
16. Chateigner-Boutin,A.L. and Small,I. (2010) Plant RNA editing. *RNA Biol.*, **7**, 213–219.
17. Tasaki,E., Hattori,M. and Sugita,M. (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial *ccmFc* transcript. *Plant J.*, **62**, 560–570.
18. Steinhauser,S., Beckert,S., Capesius,I., Malek,O. and Knoop,V. (1999) Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? *J. Mol. Evol.*, **48**, 303–312.
19. Groth-Malonek,M., Wahrmund,U., Polsakiewicz,M. and Knoop,V. (2007) Evolution of a pseudogene: Exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol. Biol. Evol.*, **24**, 1068–1074.
20. Rüdinger,M., Funk,H.T., Rensing,S.A., Maier,U.G. and Knoop,V. (2009) RNA editing: 11 sites only in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Mol. Genet. Genom.*, **281**, 473–481.
21. Giegé,P. and Brennicke,A. (1999) RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc. Natl Acad. Sci. USA*, **96**, 15324–15329.

22. Notsu,Y., Masood,S., Nishikawa,T., Kubo,N., Akiduki,G., Nakazono,M., Hirai,A. and Kadowaki,K. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genom.*, **268**, 434–445.
23. Bentolila,S., Elliott,L.E. and Hanson,M.R. (2008) Genetic architecture of mitochondrial editing in *Arabidopsis thaliana*. *Genetics*, **178**, 1693–1708.
24. Handa,H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Res.*, **31**, 5907–5916.
25. Mower,J.P. and Palmer,J.D. (2006) Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. *Mol. Genet. Genom.*, **276**, 285–293.
26. Picardi,E., Horner,D.S., Chiara,M., Schiavon,R., Valle,G. and Pesole,G. (2010) Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Res.*, **38**, 4755–4767.
27. Sloan,D.B., MacQueen,A.H., Alverson,A.J., Palmer,J.D. and Taylor,D.R. (2010) Extensive loss of RNA editing sites in rapidly evolving silene mitochondrial genomes: selection vs. retroprocessing as the driving force. *Genetics*, **185**, 1369–1380.
28. Rajasekhar,V.K. and Mulligan,R.M. (1993) RNA editing in plant mitochondria: α-phosphate is retained during C-to-U conversion in mRNAs. *Plant Cell*, **5**, 1843–1852.
29. Blanc,V., Litvak,S. and Araya,A. (1995) RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Lett.*, **373**, 56–60.
30. Yu,W. and Schuster,W. (1995) Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *J. Biol. Chem.*, **270**, 18227–18233.
31. Gualberto,J.M., Weil,J.H. and Grienenberger,J.M. (1990) Editing of the wheat coxIII transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. *Nucleic Acids Res.*, **18**, 3771–3776.
32. Schuster,W., Hiesel,R., Wissinger,B. and Brennicke,A. (1990) RNA editing in the cytochrome *b* locus of the higher plant *Oenothera berteriana* includes a U-to-C transition. *Mol. Cell Biol.*, **10**, 2428–2431.
33. Yoshinaga,K., Iinuma,H., Masuzawa,T. and Uedal,K. (1996) Extensive RNA editing of U to C in addition to C to U substitution in the *rbcL* transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Res.*, **24**, 1008–1014.
34. Kugita,M., Yamamoto,Y., Fujikawa,T., Matsumoto,T. and Yoshinaga,K. (2003) RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Res.*, **31**, 2417–2423.
35. Wolf,P.G., Rowe,C.A. and Hasebe,M. (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene*, **339**, 89–97.
36. Vangerow,S., Teerkorn,T. and Knoop,V. (1999) Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. *Plant Biol.*, **1**, 235–243.
37. Grewe,F., Viehoever,P., Weisshaar,B. and Knoop,V. (2009) A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res.*, **37**, 5093–5104.
38. Kempken,F., Bolle,N., Forner,J. and Binder,S. (2007) Transcript end mapping and analysis of RNA editing in plant mitochondria. *Methods Mol. Biol.*, **372**, 177–192.
39. Lenz,H., Rüdinger,M., Volkmar,U., Fischer,S., Herres,S., Grewe,F. and Knoop,V. (2009) Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Curr. Genet.*, **56**, 189–201.
40. Tamura,K., Dudley,J., Nei,M. and Kumar,S. (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, **24**, 1596–1599.
41. Castandet,B., Choury,D., Begu,D., Jordana,X. and Araya,A. (2010) Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Res*, **38**, 7112–7121.
42. Chaw,S.M., Chun-Chieh,S.A., Wang,D., Wu,Y.W., Liu,S.M. and Chou,T.Y. (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Mol. Biol. Evol.*, **25**, 603–615.
43. Salmans,M.L., Chaw,S.M., Lin,C.P., Shih,A.C., Wu,Y.W. and Mulligan,R.M. (2010) Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. *Curr. Genet.*, **56**, 439–446.
44. Covello,P.S. and Gray,M.W. (1993) On the evolution of RNA editing. *Trends Genet.*, **9**, 265–268.
45. Tillich,M., Lehwark,P., Morton,B.R. and Maier,U.G. (2006) The evolution of chloroplast RNA editing. *Mol. Biol. Evol.*, **23**, 1912–1921.
46. Jobson,R.W. and Qiu,Y.L. (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biol. Direct.*, **3**, 43.
47. Maier,U.G., Bozarth,A., Funk,H.T., Zauner,S., Rensing,S.A., Schmitz-Linneweber,C., Börner,T. and Tillich,M. (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biol.*, **6**, 36.
48. Lukeš,J., Leander,B.S. and Keeling,P.J. (2009) Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proc. Natl Acad. Sci. USA*, **106(Suppl. 1)**, 9963–9970.
49. Knoop,V. (2010) When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci.*, doi:10.1007/s00018-010-0538-9 (12.10.2010, last date accessed).
50. 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, pp. 115–129.
51. Malek,O., Lättig,K., Hiesel,R., Brennicke,A. and Knoop,V. (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.*, **15**, 1403–1411.
52. Groth-Malonek,M., Pruchner,D., Grewe,F. and Knoop,V. (2005) Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Mol. Biol. Evol.*, **22**, 117–125.
53. Qiu,Y.L., Li,L., Wang,B., Chen,Z., Knoop,V., Groth-Malonek,M., Dombrovska,O., Lee,J., Kent,L., Rest,J. *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl Acad. Sci. USA*, **103**, 15511–15516.
54. Binder,S., Marchfelder,A. and Brennicke,A. (1994) RNA editing of $tRNA^{Phe}$ and $tRNA^{Cys}$ in mitochondria of *Oenothera berteriana* is initiated in precursor molecules. *Mol. Gen. Genet.*, **244**, 67–74.
55. Fey,J., Tomita,K., Bergdoll,M. and Maréchal-Drouard,L. (2000) Evolutionary and functional aspects of C-to-U editing at position 28 of tRNA(Cys)(GCA) in plant mitochondria. *RNA*, **6**, 470–474.
56. Zimmerly,S., Hausner,G. and Wu,X. (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res.*, **29**, 1238–1250.
57. Michel,F. and Ferat,J.L. (1995) Structure and Activities of Group II Introns. *Annu. Rev. Biochem.*, **64**, 435–461.

4.2.2 Supplementary Data

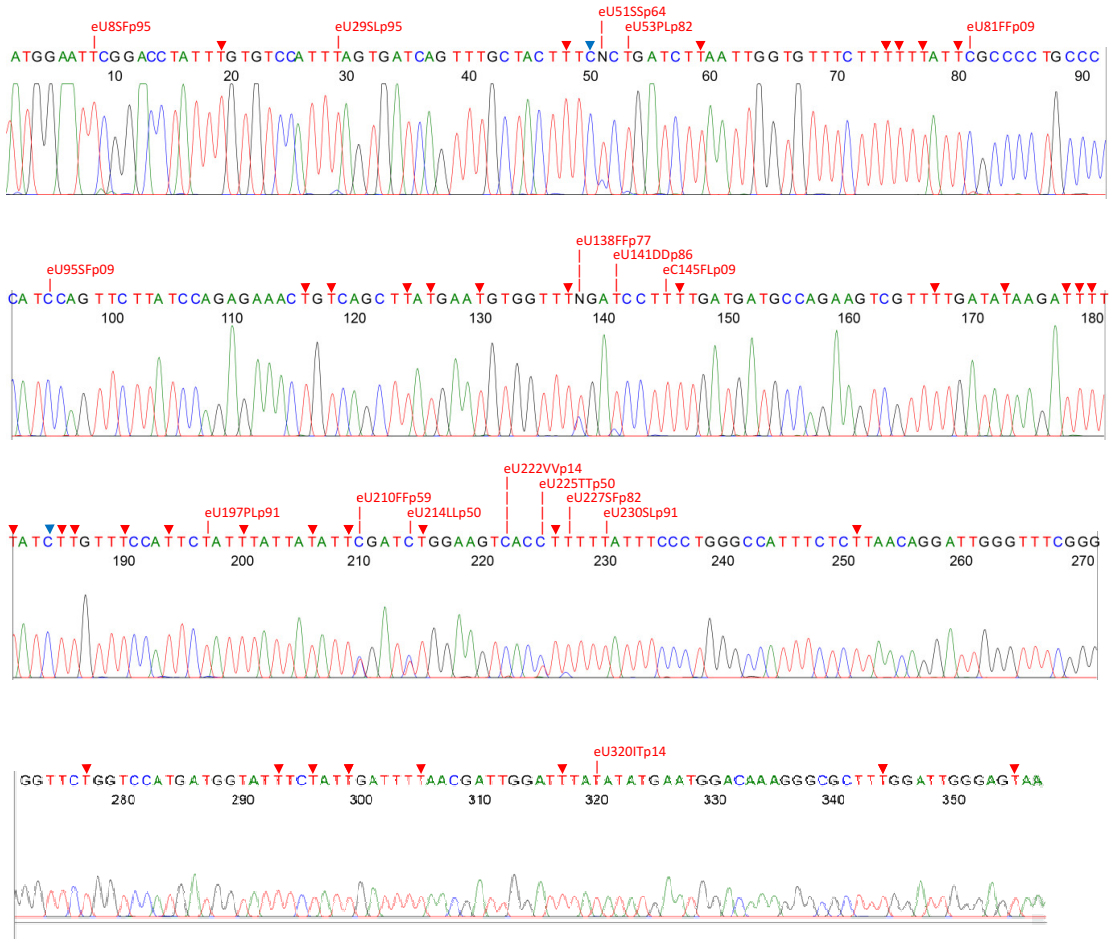


Figure 4.17: (Supplementary Figure 1) Direct sequencing of RT-PCR products for comparison with results from independent sequencing of 22 cDNA clones. Electropherograms reflect exclusively the edited nucleotide for all sites identified as fully edited (red triangles: C-to-U edits, blue triangles: U-to-C edits). Partial editing quantified from independent clone sequencing is confirmed by overlapping C and T peaks in the direct sequencing approach in most cases (e.g. nad3eU29SLp95, nad3eU138FFp77, nad3eU141DDp86) but not evident in the electropherograms for some other sites of particularly low (e.g. nad3eU95SFp09) or high (e.g. nad3eU8SFp95) partial editing.

4.3 Die mitochondriale DNA und das mitochondriale Transkriptom von *Selaginella moellendorffii*

Im Rahmen der ersten Fosmid-Klonierungen, die unter anderem zur vollständigen mtDNA Sequenzierung von *I. engelmannii* geführt haben, wurden auch Fosmid-Klonierungsversuche mit der Gesamt-DNA einer weiteren Lycophytenart der Gattung *Selaginella* unternommen. Die Ergebnisse dieser Klonierungen waren jedoch stets wenig effizient, sodass auf eine weitere Bearbeitung bis zur vollständigen Sequenz vorerst verzichtet wurde.

Aufgrund ihres kleinsten bekannten Pflanzengenoms wurde 2008 die Art *Selaginella moellendorffii* für eine vollständige Sequenzierung durch das Energy Joint Genome Institut (JGI) ausgewählt. Mit der kompletten Genomsequenz von *Selaginella*, in Verbindung der Gesamtsequenz von *P. patens* [Rensing *et al.* 2008], sollen weitere Erkenntisse über die Evolution der frühen Landpflanzen gewonnen werden [Banks 2009]. Die Rohsequenzen dieses Projekts waren in Form einzelner *scaffolds* vorab über das JGI zugänglich. Unter Hinzunahme dieser Daten wurde die vollständige *S. moellendorffii* cpDNA-Sequenz bereits veröffentlicht [Smith 2009].

Durch BLAST-Suchen stellten sich zwei *scaffolds* dieses Sequenzierprojekts als mitochondrial heraus und bildeten die Grundlage einer umfassenden Untersuchung des Chondroms von *S. moellendorffii*. Auf diesen *scaffolds* wurden bereits viele mitochondriale Gene identifiziert, deren besonders große Introns, die in vier Fällen sogar *trans*-gespleißt vorliegen, bereits eine äußerst interessante mtDNA andeuteten. Dass die identifizierten *scaffolds* allerdings nicht die vollständige mtDNA von *S. moellendorffii* repräsentieren, wurde beispielsweise durch eine fehlende 26S rRNA deutlich.

Um die Lücken in der Gesamtsequenz der mtDNA von *S. moellendorffii* zu füllen, wurde nun doch in einer begleitenden Masterarbeit die ineffiziente Fosmid-Klonierung angewandt [Julia Hecht, Universität Bonn, Masterarbeit]. Das Pflanzenmaterial wurde dazu kommerziell erworben und im Labor sowie im Botanischen Garten Bonn vermehrt. Mit der Fosmidbank von *S. moellendorffii* bestätigte sich die Annahme, dass ein kleines nukleäres Genom auch in einer aus Gesamt-DNA hergestellten Fosmidbank unterrepräsentiert vorliegt und deshalb mitochondriale Fosmide selbst in einer solchen kleinen Fosmidbank gefunden werden können. In

einer sortierten Bank, bestehend aus nur 5000 Fosmid-Klonen, wurden vier mitochondriale Fosmide identifiziert. Die anschließende Sequenzierung eines Fosmids vervollständigte die fehlenden Teile der Gesamtsequenz mit dem 26S-rRNA kodierenden Gen. In der zweiten vollständig sequenzierten mtDNA eines Lycophyten wurden viele bisher ungesehene Besonderheiten offenbar und innerhalb der Arbeit *Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes* bei *Genome Biology and Evolution* eingereicht. Das Manuskript befindet sich zur Zeit im Revisionsprozess.

Darüber hinaus wird im Manuskript über die Ergebnisse weiterführender cDNA Sequenzierungen berichtet, wodurch unter anderem das (*trans*-)spleißen aller Introns sowie das *RNA Editing* analysiert wurde. Es konnten vier *trans*-spleißende Introns verifiziert werden, von denen die Introns atpi21g2T und cobi787g2T erstmalig in *S. moellendorffii trans*-konfiguriert vorliegen. Mit der zweiten umfassenden Analyse eines Lycophyten-Transkriptoms wurde erneut eine *RNA Editing* Rekordsumme von 2139 Positionen in Transkripten proteinkodierender Gene gezählt, die in diesem Falle sämtlich in Richtung C-zu-U *editiert* werden. Zudem wurde erstmals in *S. moellendorffii* ein *RNA Editing* auch in ribosomalen RNAs nachgewiesen.

Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of *Selaginella moellendorffii* mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes

Julia Hecht[1], Felix Grewe[1] and Volker Knoop*

IZMB – Institut für Zelluläre und Molekulare Botanik
Abteilung Molekulare Evolution
Universität Bonn
Kirschallee 1, D-53115 Bonn, Germany

[1]The first two authors contributed equally to this work.

*Corresponding author: E-mail: volker.knoop@uni-bonn.de.
Tel. +49 228 73 5507; Fax +49 228 73 6467.

Data deposition: The *Selaginella moellendorffii* mitochondrial genome sequences have been deposited in GenBank under accession numbers JF338143-JF338147 (parts 1-5). Accompanying cDNA sequences have been deposited under accession numbers JF276233-JF276250.

Key words:
Lycophytes, RNA editing, *trans*-splicing, microsatellites, endosymbiotic gene transfer

Abstract

Using an independent fosmid cloning approach and comprehensive transcriptome analysis to complement data from the *Selaginella moellendorffii* genome project, we determined the complete mitochondrial genome structure of this spikemoss. Numerous recombination events mediated mainly via long sequence repeats extending up to 7 kbp result in a complex mtDNA network structure. Peculiar features associated with the repeat sequences are more than 80 different microsatellite sites (predominantly trinucleotide motifs). The *S. moellendorffii* mtDNA encodes a plant-typical core set of a twin-arginine translocase (*tat*C), 17 respiratory chain subunits and two rRNAs but lacks *atp*4 and any tRNA genes. As a further novelty among plant chondromes, the *nad*4L gene is encoded within an intron of the *nad*1 gene. A total of 37 introns occupying the 20 mitochondrial genes (four of which are disrupted into *trans*-splicing arrangements including two novel instances of *trans*-splicing introns) make the *S. moellendorffii* chondrome the intron-richest and gene-poorest plant mtDNA known. Our parallel transcriptome analyses demonstrates functional splicing of all 37 introns and reveals a new record amount of plant organelle RNA editing with a total of 2,139 sites in mRNAs and 13 sites in the two rRNAs, all of which are exclusively of the C-to-U type.

Introduction

Plant mitochondrial genomes have become legendary for the ever growing list of molecular peculiarities discovered over the last three decades. The mitochondrial genomes (chondromes) of plants exceed the ones of animals in size by two orders of magnitude and exceptionally even more, occasionally resulting in plant chondromes of bacterial genome sizes (Alverson et al., 2010; Ward et al., 1981). Most of the "extra" DNA in plant chondromes is intergenic, in parts derived from foreign sequences acquired through transfer from the nucleus or the chloroplast or even horizontally from mtDNA of other species (Bergthorsson et al., 2003). Other factors increasing plant mitochondrial genome sizes are additional genes not present in the animal lineage and organellar introns belonging to the group I or group II classes (Knoop et al., 2010). Moreover, frequent DNA recombination shuffling gene orders continuously and producing co-existing alternative genomic arrangements are typical of flowering plant genomes. Such recombination events have also led to rearrangements not only between but also within genes and have produced *trans*-splicing introns in the course of plant evolution (Malek and Knoop, 1998). Finally, the plant organelle type of RNA editing exchanges the pyrimidine nucleotides cytidine and uridine at specific transcript sites and thus alters the encoded information at the RNA level (Knoop, 2010).

Many of the peculiar plant chondrome features – including RNA editing, genomic recombination, *trans*-splicing and insertions of foreign DNA – are absent, however, in green algae and early branching land plant lineages. For example, neither of these four mtDNA features exists in the mtDNA of the liverwort *Marchantia polymorpha* (Oda et al., 1992). RNA editing, however, is a very early gain in land plant evolution, present in other bryophyte clades and apparently only secondarily lost in the marchantiid liverworts (Groth-Malonek et al., 2007; Malek et al., 1996;

Steinhauser et al., 1999). No evidence for insertions of chloroplast, nuclear or foreign DNA, for active DNA recombination or for *trans*-splicing introns has so far been identified, however, in any bryophyte chondrome including the completely determined mtDNA sequences of a moss (Terasawa et al., 2006), a jungermanniid liverwort (Wang et al., 2009) and two hornworts (Li et al., 2009; Xue et al., 2010). In contrast, all of these genomic peculiarities were recently shown to exist in the mitochondrial DNA of a lycophyte, the quillwort *Isoetes engelmannii* (Grewe et al., 2009).

Lycophytes are the sister clade to all other vascular plants (tracheophytes), i.e. the euphyllophytes, which comprise the spermatophytes (seed plants) and the monilophytes (true ferns, whisk ferns and horsetails). Hence, lycophytes represent the most ancient surviving lineage of tracheophytes that dominated the earth's flora since Devonian times, i.e. much more than 200 million years before the rise of flowering plants in the cretaceous some 140 million years ago. Highly active chondrome DNA recombination, *trans*-splicing introns and the incorporation of foreign DNA all seem to arise with the earliest tracheophytes as reflected with the existence of all these phenomena in the mtDNA of *Isoetes*.

Extant lycophytes comprise three well-defined orders, the Isoetales (quillworts), the Selaginellales (spike mosses) and the Lycopodiales (club mosses). Given their pivotal position in the evolution of land plants and in the light of the chondrome peculiarities in *Isoetes engelmannii* (Grewe et al., 2009), we wished to elucidate the mtDNA structure of a sister lycophyte. To this end, we made initial use of available mitochondrial sequence data from the *Selaginella moellendorffii* genome sequencing project. With support from an independent fosmid cloning approach, we ultimately determined the *S. moellendorffii* mitochondrial genome structure, determined its complete mtDNA sequence and complemented this by exhaustive cDNA analyses.

To our surprise we found that the *Selaginella* mtDNA reveals many differences to its sister lycophyte *Isoetes* showing that quite different pathways have been followed in mitochondrial genome evolution in the two ancient tracheophyte lineages. Both taxa are characterized by abundant RNA editing and, with more than 2,000 sites of C-to-U RNA editing, *Selaginella* actually even breaks the record of editing abundance hitherto set by *Isoetes* (Grewe et al., 2010). However, whereas lots of U-to-C editing sites were found in *Isoetes,* no single such event was detected in *Selaginella.* Both taxa are characterized by frequent DNA recombination but as a novelty, the sequence repeats involved in recombination are associated with microsatellite repeats in *Selaginella*, which have no counterparts in *Isoetes*. Introns in *Isoetes* mtDNA are tiny whereas they are huge in *Selaginella* and while promiscuous DNA of nuclear and chloroplast origin had been identified in the *Isoetes* mtDNA no such sequences exist in the *Selaginella* mtDNA. Furthermore, and as a complete novelty for a plant chondrome, the *Selaginella moellendorffii* mitochondrial DNA is devoid of any tRNA encoding gene.

Material and Methods

Molecular cloning

Selaginella moellendorfii plant material was obtained commercially (Shady Plants Fern Nursery, Coolbooa, Clashmore, Co. Waterford, Ireland) and further cultivated in the lab. Total genomic DNA was isolated using a CTAB protocol. After size-fractionation into approximately 38 kbp fragments, DNA was blunt-ended and cloned into the fosmid vector pCC1FOS using the CopyControl Fosmid Library Production Kit (EPICENTRE, Madison, Wisconsin). A library of 5,000 fosmid clones was obtained, of which 1,500 were manually sorted, filter-spotted and hybridized with P^{32}-labelled gene probes (RT-PCR-derived for *nad*2, *nad*9, *cob*, *cox*2 and *atp*9). Identity of four positive fosmid clones was verified through terminal insert sequencing, showing homology with two available *S. moellendorffii* scaffold sequences (M162 and 213). Fosmid clone 3 was used for sub-library production after revealing a PCR product for the *S. moellendorffii* mitochondrial 26S rRNA gene present in database entry GQ246802 but lacking from scaffold sequences M162 and 213. Fosmid DNA was isolated using NucleoBond Xtra Midi EF Kit (Macherey Nagel, Düren, Germany), sheared by Nebulizers (Invitrogen, Carlsbad, California), blunted using an End-It DNA End-Repair Kit (EPICENTRE, Madison, Wisconsin), A-tailed with Taq-Polymerase (Genaxxon, Biberach, Germany), and fractionated by preparative electrophoresis in 0.8 % agarose. Fragments of 2-2.5 kb in size were recovered using the NucleoSpin Extract II Kit (Macherey Nagel, Düren, Germany) and cloned into pGEM-T Easy vector (Promega, Madison, Wisconsin). 96 plasmid subclones were end-sequenced and assembled and primer walking was used to complete the sequence gap including the 26 S rRNA.

Total *S. moellendorffii* RNA was isolated using the NucleoSpin RNA Plant Kit (Macherey Nagel, Düren, Germany), cDNA was synthesized with RevertAid First

Strand cDNA Synthesis Kit (Fermentas, Burlington, Ontario) in the presence of random hexamer primers as recommended by the manufacturer. Oligonucleotides for RT-PCR were designed to ideally target 5'- and 3'-flanking UTRs to allow for amplification of full coding regions of all genes, which was successful except for the 5'-UTRs of *atp*6 and *atp*9 and the 3'-UTR of *cox*1. Oligonucleotide pairs (all sequence information available from the authors upon request) were used for PCR amplification according to the standard protocol of GoTaq DNA Polymerase (Promega, Madison, Wisconsin) in a GeneAmp PCR System 2700 (Applied Biosystems, Foster City, California) with annealing temperatures between 50-55 ℃. Amplicons were recovered from agarose gels as described and cloned into pGEM T Easy vector (Promega, Madison, Wisconsin).

Sequence analysis

Mitochondrial protein coding regions were initially identified with sensitive (minimum word size) TBLASTN and BLASTN similarity searches (Altschul et al., 1997) using an artificial concatenation of encoded proteins in the *Marchantia polymorpha* and *Chara vulgaris* mitochondrial genomes and of structural RNA genes (tRNAs + rRNAs) of *Physcomitrella patens* mtDNA against the available sequence data of the *Selaginella moellendorffii* genome project at http://genome.jgi-psf.org/Selmo1/. The tRNAscan-SE server was additionally used to scan for tRNA genes and pseudogenes (Lowe and Eddy, 1997). Further sequence handling was essentially done using the alignment explorer of the MEGA software (Kumar et al., 2008). Identification and refinement of gene structures for annotation was aided by parallel cDNA analysis. Repeat sequences and microsatellites were analyzed using REPUTER at http://bibiserv.techfak.uni-bielefeld.de/reputer/ and MSATFINDER at http://www.genomics.ceh.ac.uk/msatfinder/. Identification of RNA candidate sites in

DNA and analyses and annotation of RNA editing on cDNA sequences was done with the help of PREPACT (Lenz et al., 2009). Assignment of group I and group II introns to their respective classes was based on inspection for conserved sequence signatures and identification of orthologues in the respective insertion sites in other plant taxa. Splice sites were determined by comparison with cDNA sequences.

Sequence assembly, annotation and submission

The mitochondrial nature of *Selaginella moellendorffii* JGI (Joint Genome Institute) genome scaffolds M162 and 213 was verified by cDNA analyses of candidate gene-coding regions showing sequence co-linearity except for *cis*- and *trans*-splicing events and perfect sequence identity except for numerous sites of C-to-U RNA editing. Likewise, terminal sequences of four fosmid clones independently retrieved from cloning DNA from a different biological source of *S. moellendorffii* and the insert sequence from one of those (fosmid #3) were identical to the two scaffold sequences except for microsatellite repeat numbers as outlined under results. Despite an overall complex *S. moellendorffii* mitochondrial genome structure (Fig. 1) resulting from 10 recombining sequence repeats R1-R10, the fosmid cloning suggested that (i) mtDNA molecules in a size range of at least 40 kbp should indeed physically exist in nature and (ii) that scaffold assemblies 213 and M162 were correct, not considering two extended gaps in the contig assembly of the latter. Ultimately, mtDNA sequences were submitted and annotated as five separate database entries (excluding M162 assembly gaps) each ending in active repeat sequences to limit sequence redundancies and to clearly suggest alternative co-existing mtDNA arrangements in nature (see Figs. 1-3) while avoiding any suggestion of hitherto unproven large sequence continuities in a hypothetical “master circle”: Part 1: R9-R6a-*atp*8-*nad*6-*nad*3-R1-*cox*3-R2-*nad*2-R4 (42,231 bp, acc. no. JF338143, from scaffold 213), Part

2: R4-*atp*9e2-R8-*nad*4-R3a-R10-*cox*1-R1-*cox*2e12-*nad*9-R5-*rrn*S-R6-R9-*nad*1-R9-*cob*e3-*tat*C-R4-*nad*5-*atp*1-*cob*e12-R3 (143,606 bp, acc. no. JF338144, from upstream sequence continuity in scaffold M162), Part 3: R6a-*nad*7-R7b- *atp*9e34-R8-*atp*9e2-R4b-*atp*6-R7 (35,649 bp, acc. no. JF338145, from central sequence continuity in scaffold M162), Part 4: R6-*cox*2e34-R5 (20,004 bp, acc. no. JF338146 from terminal sequence continuity in scaffold M162) and Part 5: R5-*rrn*L-R10 (acc. no. JF338147, 19,723 bp from sequence insert in fosmid #3). Notably, the mitochondrial sequence gap closed with fosmid #3 insert (part 5) comes from a different biological source of *S. moellendorffii* and makes the total mtDNA complexity a “hybrid” in a strict sense, reflected by the recognizable microsatellite repeat variability (see results). Database entry annotations include features on coding sequences, introns, microsatellites and repeat regions. At present, scaffold 213 is also deposited in GenBank as entry NW_003314473, however, unidentified as mitochondrial in nature and suggesting several hypothetical but very unlikely gene models. Seven separate mtDNA regions had earlier been submitted to the database without detailed annotation (accessions GQ264802-08) as *nad*9-like, *atp*1-like, *cox*1-like, *nad*2-like, *nad*7-like, *atp*9-like and *nad*4-like respectively, as a result of a previous *S. moellendorffii* study focusing on chloroplast DNA (Smith, 2009). All cDNA sequences retrieved in our study, which confirm splice sites and reveal RNA editing positions are submitted separately for each gene (database accessions JF276233-JF276250).

Results

Elucidating the Selaginella chondrome complexity

We started our experimental approach to determine the *Selaginella moellendorffii* mitochondrial genome by creating an artificial protein sequence concatenation which

comprises all known plant mitochondrial protein sequences (essentially the *Marchantia polymorpha* mitochondrial gene complement plus *nad*7 and *rpl*14 missing there that were taken from *Physcomitrella patens*). Using this artificial concatenated query sequence we scanned the publicly available *Selaginella moellendorffii* genome sequence data by TBLASTN and identified scaffolds “213” of 46.7 kbp and “M162” of 227 kbp, which carried significant mitochondrial sequence homologies. Given that (i) scaffold M162 contained two gaps in the contig assembly (of 1,951 bp and 3,568 bp, respectively), that (ii) no physical linkage between the two scaffolds was evident and that (iii) no evidence for a mitochondrial 26S rRNA was found in the sequence data (although obviously existing, as documented by database accession GQ246802, Smith, 2009), we decided to complement the sequence analysis with an independent fosmid cloning approach of *S. moellendorffii* DNA. A small fosmid library of ca. 5,000 fosmid clones was obtained, of which 1,536 were manually sorted and screened with a mix of mitochondrial gene probes. Initial RT-PCR analyses performed in parallel had verified transcription (and RNA editing) in initial *bona fide* mitochondrial gene sequence candidates and cDNA probes of five mitochondrial genes (*nad*2, *nad*9, *cob*, *cox*2 and *atp*9) were used for library screening and revealed four positive fosmid clones. Their eight end sequences perfectly mapped to the initial scaffold sequences with the exception of one end sequence of fosmid 2. End sequences of the other fosmids in fact mapped to both scaffolds and hence suggested that physical connections should exist in at least some recombined mtDNA arrangements. Continuous, full-length co-linearity of fosmid inserts (around ca. 39 kbp each) and the scaffold assemblies, however, was obviously absent in all four cases. This indicated frequent recombination events and suggested that the initial scaffold assemblies would reflect only one of several possible co-existing mitochondrial genome arrangements.

Attempts to identify 26S rRNA gene homologies on the fosmids via PCR revealed products of expected size in two of them and one (fosmid 3) was selected for sequencing. Fosmid 3 end sequences (36.3 kbp insert size) were co-linear with scaffold M162 but ran in opposite directions up to two recombination sites. Recombination points embedded 7.5 kbp of novel sequence in fosmid 3 not represented in the *S. moellendorffii* genome scaffold sequences and this included both the so far elusive 26S rRNA gene and the missing insert terminal sequences of fosmid 2.

We concluded (i) that sequence assembly of *S. moellendorffii* scaffolds was impeded by frequent recombination events in its mtDNA, resulting in apparent gaps in the M162 scaffold and (ii) that unique sequences in the two separate scaffolds were in fact physically linked through recombination across sequence repeats in at least some co-existing topological alternatives.

Repeats and microsatellites

Inspecting all available sequence data we found several large sequence repeats flanking different single copy sequence regions. By combining the existing scaffold sequences and the new fosmid sequences we were ultimately able to assemble a recombinational mitochondrial genome map (Fig. 1). It turned out that the *Selaginella moellendorffii* mitochondrial DNA complexity is essentially explained by ten large sequence repeats, labelled R1 through R10 (Fig. 2), which provide alternative pathways through the flanking single copy genomic regions (Fig. 3). These large repeated sequences in *S. moellendorffii* are mostly intergenic (R1, R4, R5 and R6) or located in introns (R2, R3 and R8). The exceptions are R7 affecting exons 6 and 7 of the *cox*1 gene, R9 extending into *nad*1 exon 1 and R10 including exon 1 of the *atp*1 gene. Four of the large repeated sequences (R3, R4, R6 and R7) feature additional

internal recombination breakpoints and are hence in fact composed of two recombinationally active parts (designated as a/b in Figs. 1 and 2).

The long sequence repeats immediately remind of those previously found in flowering plant mitochondria, which are involved in chondrome rearrangements and can result in stoichiometric shifting of subgenomic mtDNA topologies, occasionally beyond detection level for one (or more) of alternative DNA topologies (Arrieta-Montiel et al., 2001). In the cases of R1 and R9 no evidence for recombinational activity was found. For example, the pathways *nad*3-R1-*cox*3 and *cox*1-R1-*cox*2 were identified for R1 in the available sequences, but not the reciprocal exchanges (Fig. 2), also leaving open the possibility that some repeated sequences may be recombinationally inactive. In three cases (R2, R8 and R10), recombination breakpoints were only identified for one end leaving the actual extension of these repeated sequences undetermined. Alternatively, these may be regarded as part of a composite repeat with neighbouring repeats, which additionally include coding regions, i.e.: [R2-nad4e2-nad4e1-R3], [R8-atp9e2-R4] and [R10-R7].

Frequent recombination in *S. moellendorffii* mtDNA left only a single continuity of three genes undisturbed by recombination: *atp*8-*nad*6-*nad*3 (Figs. 1 and 3). In one additional case of three linked, non-recombining coding sequences, *cob* exons 1 and 2 are located upstream of *atp*1 and *nad*5, but the missing *cob* exon 3 is added by a novel *trans*-splicing event now discovered (Fig. 3), which will be discussed below. Interestingly though, directions of transcription are retained not only for these two gene continuities but very frequently also across recombination events. In fact, we could easily arrange the orientations of repeated sequences according to the dominating direction of transcription of flanking gene sequences (Fig. 1) and only five cases are found, where individual genes run against the dominating trend for transcriptional orientation (Fig. 1): *nad*9 at R5, the 18S rRNA at R6, *cob* exon 3 at

R9, *tat*C at R4, and finally the first seven *cox*1 exons running in opposite direction to the other genes flanking R7. Notably, *cox*1 exon 8 is located upstream of exon 1 in inverted orientation but in the same transcriptional direction as the other three genes flanking R1 (*nad*3, *cox*3 and *cox*2). Exon 8 of *cox*1 is added via a *trans*-splicing event during transcript maturation (see below).

Striking features of the repeated sequence units in the *Selaginella moellendorffii* mtDNA are numerous short sequence repeat (SSR) microsatellite motifs (Fig. 2). We observed a total of 82 such microsatellites with at least 5 repeat units. Dominating in abundance are trinucleotide motif repeats (57), followed by fewer tetranucleotide (11), dinucleotide (10) and pentanucleotide (4) microsatellites. In comparison, only 16 such microsatellites occur in the single copy genomic regions (Fig. 3). Here, they are completely absent from coding sequences and mainly occur in intron sequences, most significant examples being AGC_{17} and ACC_{9} in nad4i976 (Fig. 3j), GCA_{12} in nad5i1455 (Fig. 3o) and ACC_{10} in nad7i209 (Fig. 3q).

The largest among all microsatellites is an $AAAGG_{15}$ microsatellite motif at the beginning of R9 (Fig. 2). Occasional minor variability in microsatellite repeat numbers are the only sequence differences between repeat copies in different sequence environments (Fig. 2). For statistical significance we initially only considered microsatellites with repeat numbers of minimally five. However, upon closer inspection of repeats we found that microsatellites with lower and variable repeat numbers were present close to recombinational breakpoints, e.g. $AGCC_{1/2}$ and $CCG_{3/4}$ close to the termini of R3a (Fig. 2). Most surprising, however, were insights from comparison of homologous sequences in the genomic scaffolds with those derived from our fosmid cloning approach: Numerous sites of additional low copy number repeat variability (mostly 1/2, occasionally 2/3, 0/1 or 0/2) were seen also outside of repeat regions, with trinucleotide motifs again strongly dominating in

abundance. One such example of increased repeat variability is shown for the upstream terminus of R5 (Fig. 4). These differences of yet higher microsatellite variability most likely reflect the different *Selaginella moellendorffii* isolates used for DNA preparations in our independent fosmid cloning approach versus the previous genomic sequencing approach. We conclude that the peculiar variability of very low copy number motifs (mostly trinucleotide) reflect the tendency of *S. moellendorffii* mtDNA towards evolving novel microsatellite sites, which subsequently pave the way to evolve recombinationally active sequences.

The gene and intron complement

The *Selaginella moellendorffii* mtDNA encodes a full, plant-typical core set of genes for subunits of the respiratory chain complexes I (*nad*1-9), III (*cob*), IV (*cox*1-3) and V (*atp*1, 6, 8 and 9) and the *tat*C gene encoding a twin-arginine translocase (Tab. 1). Absent from the chondrome, however, are an *atp*4 gene and genes for subunits of complex II (*sdh*), cytochrome c biogenesis and maturation (*ccm*) and a 5S rRNA. Highly striking and a novelty among the complete plant chondrome sequences determined so far is the absence of any genes encoding tRNAs. Whereas *ccm* genes are absent from the *Isoetes engelmannii* chondrome as well, *atp*4, *sdh*3, four ribosomal proteins, a 5S rRNA and 13 tRNAs are encoded there (Tab. 1).

A contrasting picture is observed for the complement of 35 group II (g2) and two group I (g1) introns: Seven group II introns having homologues in seed plants are present in *Selaginella* but absent in *Isoetes* (nad1i477g2, nad1i669g2, nad1i728g2, nad4i976g2, nad7i140g2, cox2i373g2 and cox2i691g2). Reciprocally, introns nad7i1113g2 and cox1i395g1 conserved in liverworts are present in *Isoetes* but not in *Selaginella*. The intron-rich *cox*1 gene in particular shows further variability with two more introns in each taxon not present in the respective other (Tab. 1). The

recently discovered *trans*-splicing group I intron cox1i1305g1 in *Isoetes*, however, is conserved in *Selaginella* with the same physical discontinuity.

Most interestingly, all introns known as *trans*-splicing in seed plants (nad1i394g2, nad1i669g2, nad1i728g2, nad2i542g2, nad5i1455g2 and nad51477g2) have *cis*-splicing counterparts in *Selaginella moellendorffii* without exception whereas the spikemoss features *trans*-splicing group II introns in three other locations (Tab. 1). The case of a disrupted, *trans*-splicing cox2i373g2T intron has very recently also been reported for *Allium* (Kim and Yoon, 2010), whereas cobi787g2T and atp9i21g2T represent two novel cases of *trans*-splicing introns in nature.

In comparison to other plant mtDNAs the *Selaginella moellendorffii* mtDNA features an extraordinarily high average GC content of 68.1% that exceeds by far those of other land plant mtDNAs hitherto sequenced (in all of which GC content is below 50%). It is interesting to see that the chloroplast DNA of *S. moellendorffii* similarly has an elevated GC content of 51% (Smith, 2009).

Intergenic regions are comparatively large in *S. moellendorffii* mtDNA with two exceptions: *cob-tat*C with only 50 bp and *nad*2-*nad*4, which actually overlap by 17 bp. In contrast to the *Isoetes engelmannii* mtDNA (Grewe et al., 2009) no insertions of chloroplast or nuclear DNA are discernible. Finally, one further peculiarity previously not observed in other plant mtDNAs (not considering intron-borne maturases or endonucleases) is the insertion of one gene in the intron of another: The *S. moellendorffii nad*4L gene is located within group II intron nad1i728g2 of the *nad*1 gene in the same direction of transcription (Fig. 3n).

RNA editing in mRNAs and rRNAs

In parallel to determining the complete *Selaginella moellendorffii* mtDNA we performed an exhaustive complementary cDNA analysis covering all coding regions

(18 protein coding and 2 rRNAs). We found that all 37 intron sequences (Tab. 1) are correctly spliced as can be predicted from their respective group I or group II secondary structures. Moreover, the cDNA analyses showed that RNA editing in *Selaginella moellendorffii* is in fact yet more abundant than in the sister lycophyte *Isoetes engelmannii*. In total, we observed 2139 RNA editing sites in the 18 mRNAs (Table 1). Of these, 424 (i.e. ca. 20%) are silent whereas the others introduce 1488 codon changes. A complete list of RNA editing positions discovered in our study using the recently proposed nomenclature (Lenz et al., 2009; Rüdinger et al., 2009) is given in supplementary table 1. In two cases, more than 200 RNA editing events affect single mRNAs: 249 sites in *nad*2 and 237 sites in *nad*4. In fact, these two genes and *cox*1 in *S. moellendorffii* now even exceed the *Isoetes engelmannii nad*5 gene previously setting the record for editing site abundance among land plant mitochondrial genes (Tab. 2).

The highest density of editing sites, however, was found in the small *nad*3 reading frame (Fig. 5). Here, 77 edits change more than every 5th nucleotide in the only 369 nucleotide long *nad*3 reading frame and introduce 45 codon changes. This includes the alteration of seven codons in a row by 14 edits changing the genomically encoded PTSPTTP protein sequence into SILFIIF. This sequence of codons is located at the end of a stretch of 45 nucleotides, 22 of which are edited (Fig. 5). The alteration of even eight codons in a row (PPTPPTPT to LFILLMSI) in the *nad*4L gene is another example of extreme RNA editing density (Suppl. Tab. 1). Only one of the 18 protein encoding genes (*nad*7) has start and stop codons on DNA level whereas either the one or the other or even both (in *atp*1, cox3 and *nad*4) have to be introduced by RNA editing (Suppl. Tab. 1).

Most strikingly, all events of RNA editing in *S. moellendorffii* are exclusively C-to-U exchanges, a very surprising finding given that RNA editing frequently also

operates in reverse in hornworts, ferns and lycophytes including *Selaginella*'s sister genus *Isoetes* (Grewe et al., 2010). Likewise striking is a particular bias in the codon changes observed. The 392 proline (CCN) to leucine (CUN) codon changes exceed all others by far, followed by 140 proline (CCY) to phenylalanine (UUY) changes which require two simultaneous editings per codon. Surprisingly, these are followed in abundance by 137 alanine (GCN) to valine (GUN) changes. This latter type of RNA editing is generally rare in other plants due to the guanidine nucleotide preceding the editing position but in the case of *Selaginella* they even exceed the proline (CCN) to serine (UCN) edits normally observed in very high proportion in other plant mtDNAs.

Finally, we could identify several editing sites in the two rRNAs. Three C-to-U editing events were found in the 26S rRNA, ten in the 18S rRNA. In the latter case, all edits are located in the first 18S rRNA exon where nine of them cluster in an upstream sequence stretch of only 200 nucleotides (Fig. 6A). With one single exception (editing site rrnSeU295), all editing events re-introduce uridines conserved in rRNAs of other taxa as exemplarily shown in the alignment with the *Pinus* homologue (Fig. 6A). Notably, the 10th editing identified in the *Selaginella* 18S rRNA affects the last nucleotide of the 18S rRNA 5'-exon and may directly influence splicing of group I intron rrn18i839g1, as it likely influences the base pairing needed for the conserved paired region P1 (Fig. 6B). A further potential editing event may be located 6 nt. downstream in the intron region contributing to the P1 pairing and could similarly be a prerequisite for splicing of rrn18i839g1.

Discussion

The number of features distinguishing the mtDNA of *Selaginella moellendorffii* from the one of the sister lycophyte *Isoetes engelmannii,* which has previously been

determined (Grewe et al., 2009), is astounding (Tab. 3). A yet higher number and density of RNA editing sites than previously observed in *Isoetes* is impressive as such, but yet more significant is the complete absence of U-to-C editing in the presence of 2139 sites of C-to-U editing in *Selaginella* contrasting the corresponding numbers of 222 and 1560 in *Isoetes*. The concentration of editing sites (Fig. 5a) actually reminds of the extreme "pan-editing" of the uridine insertion-type occasionally observed in trypanosomes (Feagin et al., 1988; for a recent review see Knoop, 2010). The one obvious explanation for the high degree of C-to-U editing observed, is the unprecedented high GC content of 68.1% in a plant mtDNA now found in *Selaginella moellendorffii* (Tab. 3), which even exceeds the recently reported case of 57% GC in the alga *Polytomella capuana* by far (Smith and Lee, 2008). This observation in fact revitalizes an early suggestion that RNA editing frequencies may shift in response to GC content drift in organelle genomes (Malek et al., 1996).

The extension of introns to sizes of more than 6 kb in *Selaginella moellendorffii* likewise stands in complete contrast to the tiny introns in the mtDNA of *Isoetes engelmannii* (Tab. 3). Orthologous introns occupying identical sites may be up to 8-fold (rrnSi839g1) or even 9-fold (nad2i156g2) larger in *Selaginella* than in *Isoetes*. Consequently, it is all the more difficult to explain why several typical plant mitochondrial genes and promiscuous DNA sequences originating from the chloroplast or nucleus are present in the compact and gene-dense chondrome of *Isoetes* but are absent in *Selaginella* (Tab. 3).

Along the same lines, the complete absence of any tRNA-encoding genes in *S. moellendorffii* mtDNA is a first example for plants. It could be envisaged that tRNA genes may be overlooked when subject to massive RNA editing. Testing the tRNAscan-SE program (Lowe and Eddy, 1997) on the *Isoetes engelmannii* mtDNA, however, gave evidence that even the heavily edited tRNAs in this taxon (with up to

18 sites in a single tRNA) are faithfully identified, at least as tRNA-pseudo-gene candidates.

Such massive loss of mitochondrial tRNA genes has previously been reported for animal mtDNA, e.g. in the phylum Cnidaria (Haen et al., 2010) and very recently also for the angiosperm *Silene latifolia* where only six native mitochondrial tRNAs are retained in the chondrome (Sloan et al., 2010a). In contrast to endosymbiotic gene transfer (EGT) of protein genes which acquire mitochondrial targeting signals in their new nuclear location, tRNA genes lost from mitochondrial DNA are known to be functionally replaced by cytosolic tRNA counterparts (Duchene et al., 2009; Duchene et al., 2011; Glover et al., 2001). Hence, we could expectedly not identify any *bona fide* mitochondrial tRNA genes in the nuclear *Selaginella moellendorffii* assemblies.

Genes for 5S rRNAs are a generally conserved feature in land plant mitochondrial genomes and the absence of a 5S rRNA gene in the *Selaginella moellendorffii* mtDNA is surprising. Outside of the plant lineage, however, 5S rRNA genes are frequently absent from mitochondrial genomes (Lang et al., 1996) or so highly diverged that their identification on DNA level alone was impossible (Bullerwell et al., 2003). An ultimate answer for *Selaginella moellendorffii* mitochondria relies on purification and analyses of mitochondrial ribosomes or rRNAs, respectively.

The absence of any ribosomal protein (*rpl, rps*) genes and of *sdh* genes encoding subunits of complex II in *Selaginella moellendorffii* mtDNA seems to evolutionarily anticipate the frequent, independent EGT of these genes into the nucleus, which has been demonstrated among angiosperms (Adams et al., 2002). Although EGT was not observed for *ccm* genes encoding proteins of cytochrome-c-biogenesis in that angiosperm survey, their absence from *Selaginella* mtDNA is also unsurprising given that they also lack from *Isoetes engelmannii* and hornwort mtDNAs (Grewe et al., 2009; Li et al., 2009; Xue et al., 2010). However, these

observations postulate at least two independent losses of *ccm* genes from mitochondria in the lycophyte and hornwort stem lineages. A novelty with respect to gene complement is the loss of a mitochondrially encoded *atp*4 gene from the mtDNA of *S. moellendorffii*, which has previously not been reported for the land plant lineage. Interestingly, the *atp*4 gene although retained in the *Isoetes engelmannii* mtDNA is mis-edited on transcript level to introduce a stop codon (Grewe et al., 2010), possibly indicating an emerging pseudogene which may be functionally replaced by a functional nuclear copy. Moreover, an absence of *atp*4 but simultaneous retention of *atp*1, *atp*6 and *atp*9 (as well as *sdh*3 and several *rpl* and *rps* genes) in mtDNA has previously been described for the cryptophyte *Hemiselmis andersenii* (Kim et al., 2008). Given the high degree of recombinational activity in the mtDNAs of *Selaginella* and *Isoetes*, it is not surprising that nearly no gene arrangements are conserved between the two taxa. In fact, *atp*8-*nad*6 and *nad*4-*nad*2 are the only two conserved gene continuities.

The actual physical structures of plant mitochondrial genomes have been debated since long. Populations of smaller, overlapping and/or linear molecules may actually be stoichiometrically dominant over circular DNAs of genome size, if existing at all, even in the absence of recombination repeats in the circular-mapping genomes of the bryophytes (Bendich, 1993; Manchekar et al., 2006). The situation is significantly complicated with the larger and recombinationally active flowering plant mtDNAs, in which subgenomic molecules may shift in stoichiometry (Abdelnoor et al., 2003; Arrieta-Montiel and Mackenzie, 2011) and which may differ significantly in appearance even between closely related isolates of the same species (e.g. Allen et al., 2007; Ullrich et al., 1997). Obviously this is also true for *Selaginella* mtDNA where elucidating the true physical structure of mtDNA molecule populations will necessarily

rely on purification of large amounts of mtDNA to be analyzed by methods such as pulsed-field gel electrophoresis.

Being the extant representatives of the most ancient surviving lineage of vascular plants, lycophytes are a crucial clade to understand plant evolution. The complete sequence of the *Selaginella moellendorffii* nuclear genome will be a major milestone in this regard (Banks, 2009). As evolutionarily "early" vascular plants, lycophytes may be considered a transitionally evolutionary state in the development from a bryophyte-type lifestyle to the more advanced tracheophytes, the euphyllophytes. On the other hand it is noteworthy that common characters may also evolve surprisingly convergently and independently and the recent studies on evolution of syringyl lignin biosynthesis in tracheophytes are an important case in point (Weng et al., 2010).

So far, it appears that *Isoetes* has had a tendency to retain more ancient mtDNA features from the bryophyte-like ancestors (reverse U-to-C editing shared with hornworts, introns shared with liverworts) whereas *Selaginella* features more common characteristics with seed plants (introns, extended intergenic repeat regions active in recombination). However, given the unequivocal monophyly of lycophytes as a whole it must be assumed that all these characteristics were jointly present in their last common ancestor with euphyllophytes. It will be highly interesting in this regard to have the mitochondrial genome sequence of a member of the remaining third order of lycophytes, the Lycopodiales (genus *Huperzia*, Dr. Y.L. Qiu, Ann Arbor, MI, USA, pers. comm.), available soon.

The largest benefit to understand the peculiarities of the *Selaginella moellendorffii* mtDNA now outlined here will come from the completed nuclear genome analysis (Banks, 2009). One highly interesting aspect will be the investigation of homologues for nuclear factors recently determined to take part in

mitochondrial genome rearrangements in flowering plants (Abdelnoor et al., 2003; Shedge et al., 2007). Possibly even more exciting will be the investigation of nuclear-encoded factors for recognizing mitochondrial RNA editing positions, among which the large plant-specific gene families of RNA-binding pentatricopeptide repeat (PPR) proteins with PPR repeats of variable length (PLS-type) hold a very crucial position (Knoop, 2010; Knoop and Rüdinger, 2010; Lurin et al., 2004; Salone et al., 2007; Tasaki and Sugita, 2010; Zehrmann et al., 2009). The extreme degree of editing in its mitochondrial transcriptome described here obviously correlates well with a particularly large PPR gene family in *Selaginella moellendorffii* (http://wiki.genomics.purdue.edu/index.php/PPR_gene_family) and this could make the lycophyte an interesting object of study to understand the RNA sequence-recognition code of PLS-type PPR proteins.

Acknowledgements

The *Selaginella moellendorffii* complete genome sequencing initiative coordinated by Dr. Jo Ann Banks (Purdue University West Lafayette, IN) and communication on this project with Dr. Mitsuyasu Hasebe (National Institute for Basic Biology, Okazaki, Japan) is very gratefully acknowledged. We also wish to thank Monika Polsakiewicz for excellent technical assistance and to acknowledge earlier work on lycophyte mtDNAs based on cosmid cloning attempts by Karolin Kalmbach, Yesim Kümetepe, Julia Neuwirt and Patrick Johner in Ulm, Germany. We are very grateful for grant Kn411/6-1 by the Deutsche Forschungsgemeinschaft (DFG) supporting the work on early vascular plant mtDNAs in the authors' laboratory.

Figure legends

Table 1. Gene and intron complement in the mtDNA of *Selaginella moellendorffii* (Smoe, this work) in comparison to the ones of the liverwort *Marchantia polymorpha* (Mpol, Oda et al., 1992), the moss *Physcomitrella patens* (Ppat, Terasawa et al., 2006), the hornwort *Megaceros aenigmaticus* (Maen, Li et al., 2009), the lycophyte *Isoetes engelmannii* (Ieng, Grewe et al., 2009) and the gymnosperm *Cycas taitungensis* (Ctai, Chaw et al., 2008). Omitted from the list were chloroplast-derived tRNA genes *trn*H, *trn*M and *trn*V exclusively present in the *Cycas* mtDNA. Intron nomenclature relies on labelling according to the preceding nucleotide in the mature RNA (Dombrovska and Qiu, 2004) amended by the labels g1/g2 for group I and group II introns (Knoop, 2004), respectively. Double plus signs indicate presence of intron-borne ORFs, Ψ indicates recognizable pseudogene remnants and "*trans*" indicates *trans*-splicing intron configurations.

Table 2. RNA editing sites discovered in the *Selaginella moellendorffii* (Smoe) transcriptome are listed for comparison with data from similarly comprehensive transcriptome studies in the moss *Physcomitrella patens* (Ppat, Rüdinger et al., 2009), the lycophyte *Isoetes engelmannii* (Ieng, Grewe et al., 2010), the flowering plants *Beta vulgaris* (Bvul, Mower and Palmer, 2006), *Silene noctiflora* and *Silene latifolia* (Snoc and Slat, Sloan et al., 2010b), *Arabidopsis thaliana* (Atha, Giegé and Brennicke, 1999), *Brassica napus* (Bnap, Handa, 2003) and *Oryza sativa* (Osat, Notsu et al., 2002). Dark grey shading indicates that a given gene is absent from the respective mtDNA, light shading indicates gene overlaps where editing has been counted once only (*rpl*16, *rps*3 in *Arabidopsis*) or pseudo-genes (*sdh*4 in Beta, *rps*11 and *rps*14 in Oryza), *mat*R is the nad1i728g2 intron-borne maturase present in seed

plants (n. i.= not investigated). The four top-scoring genes with more than 160 editing sites each in the respective mRNAs are indicated in bold.

Table 3. Comparative overview about mtDNA features deviating in the mitochondrial genomes of *Isoetes engelmannii* (Grewe et al., 2009) and *Selaginella moellendorffii* (this work).

Supplementary Table 1. Complete list of RNA editing sites identified in the *Selaginella moellendorffii* mitochondrial transcriptome study. Editing sites are listed using the recently established nomenclature (Lenz et al., 2009; Rüdinger et al., 2009). The corresponding cDNA sequences were submitted to the database under accession numbers JF276233-JF276250.

Supplementary Table 2. Overview about codon changes introduced by RNA editing in the 18 protein-encoding genes encoded in the mitochondrial genome of *Selaginella moellendorffii*. Additionally, the total numbers of codon changes per coding region and the respective introductions of start or stop codons for translation are given. In the case of the predicted editings for the start codon in *atp*6 and *atp*9 and the stop in *cox*1, we were unable to confirm these with RT-PCR primers binding in the flanking UTRs.

Figure 1. The *Selaginella moellendorffii* mitochondrial genome structure is determined by ten recombinationally active sequence repeats (R1-R10, see Fig. 2) linking the single copy regions (a-r, see Fig. 3) through a multitude of recombinational pathways. Repeats R3, R4, R6 and R7 possess additional recombinational breakpoints (a/b), further increasing mtDNA complexity. Repeated sequences are

mostly intergenic with the exception of R7 extending over exons 6 and 7 of the *cox*1 gene, R10 containing exon 1 of the *atp*1 gene and R9 extending into the first exon of *nad*1. Exon (e) numbers are indicated only where gene continuities are disrupted through rearrangements, introns are also present in *atp*6, *nad*3, *nad*5, *nad*7 and the 18SrRNA. Arrows indicate the directions of transcription for coding regions, which are mostly conserved across sequence repeats with the exception of *nad*9/R5 *cob*e3/R9, *tat*C/R4 and *cox*1/R7.

Figure 2. The ten large repeated sequences R1-R10 in *Selaginella moellendorffii* mtDNA extend to up to 7.3 kb (R7). Repeats R3, R4, R6 and R7 are composite repeats (a/b) with internal sites of recombination. The exact sizes of R2, R8 and R10 are undetermined given that only one terminating recombination point is identified in these three cases. All repeats are characterized by numerous microsatellite sequences which often vary in repeat number as indicated.

Figure 3. A total of 18 single-copy sequence islands (a-r), flanked by sequence repeats R1-R10 (see Fig. 2), mainly most of the coding regions of the *Selaginella moellendorffii* mtDNA. To display full gene complement, repeats R7 and R9 are additionally included, which contain parts of *cox*1 and *atp*9, respectively. Gene displays were created using OGDRAW (Lohse et al., 2007). Color coding of exons indicates gene for complex I (yellow), III (light green), IV (purple) V (dark green), ribosomal RNAs (red) and tatC (blue). Dotted lines connect exons linked via *trans*-splicing introns cox1i1305g1T (c), cox2i373g2T (d-e), atp9i21g2T (R10-h) and cobi787g2T (o-r). The *atp*8-*nad*6-*nad*3 arrangement (a) is the only case of a complete three-gene-continuity unaffected by recombination in *S. moellendorffii* mtDNA.

Figure 4. An example of microsatellite variability at the beginning of repeat R9, proximal to the *nad*9 gene. The upper sequence comes from *Selaginella moellendorffii* scaffold M162 and continues into the 18S rRNA, the lower sequence comes from fosmid 3 identified in an independent cloning approach and continues towards the 26S rRNA.

Figure 5. RNA editing sites in the *nad*3 gene are highlighted with underlining where cytidines are converted into uridines in the mRNA. Resulting codon changes are highlighted by shading.

Figure 6. Ten events of RNA editing were identified in the mitochondrial 18S rRNA of *Selaginella moellendorffii*. A: Nine sites of RNA editing converting cytidines into uridines are located in the region between nucleotides 294 and 488 of the upstream 18S rRNA exon. All but the second of these events (editing site rrnSeU295) introduce uridines conserved in the corresponding 18S rRNA sequence of the gymnosperm *Pinus* shown for comparison. B: A 10th editing event affects the nucleotide upstream of the rrnSi869g1 splice site (arrow) and may be crucial to the splicing event by influencing base-pair formation in P1.

Genes / Introns	*Mpol*	*Ppat*	*Maen*	*leng*	*Smoe*	*Ctai*
atp1	+	+	+	+	+	+
atp1i805g2			+			
atp1i989g2	++					
atp1i1019g2			++			
atp1i1050g2	++		++			
atp1i1128g2		+				
atp4	+	+	+	+		+
atp6	+	+	+	+	+	+
atp6i80g2		+	+			
atp6i439g2			+	+	+	
atp8	+	+	Ψ	+	+	+
atp9	+	+	+	+	+	+
atp9i21g2		+		+	*trans*	
atp9i87g2	++	+		++	+	
atp9i95g2		+	++		+	
ccmB	+	+				+
ccmC	+	+				+
ccmF	+	+	Ψ			+
ccmFCi829g2		+	+			+
cob	+	+	+	+	+	+
cobi372g2	+					
cobi420g1		+				
cobi693g2				+	+	
cobi783g2	+					
cobi787g2			++	+	*trans*	
cobi824g2	++					
cobi838g2			++			
cox1	+	+	+	+	+	+
cox1i44g2	++		++			
cox1i150g2			++			
cox1i178g2	++					
cox1i227g2				+	+	
cox1i266g2				+	+	
cox1i323g2				+		
cox1i375g1	+					
cox1i395g1	++			+		
cox1i511g2	++	+			+	
cox1i624g1	+	+				
cox1i730g1	++					
cox1i732g2		++				
cox1i876g1					+	
cox1i995g2				+	+	
cox1i1064g2		+				
cox1i1116g1	++					
cox1i1149g2					+	
cox1i1298g2			++			
cox1i1305g1	+			*trans*	*trans*	
cox2	+	+	+	+	+	+
cox2i94g2				+	+	
cox2i97g2	+					
cox2i104g2		+				
cox2i250g2	++					
cox2i281g2			++			
cox2i373g2		+	++		*trans*	+
cox2i691g2		+				+
cox3	+	+	+	+	+	+
cox3i171g2	+					
cox3i506g2		+				
cox3i625g2	+					
nad1	+	+	+	+	+	+
nad1i287g2		+	+			
nad1i348g2			+			
nad1i394g2				+	+	*trans*
nad1i477g2					+	+
nad1i669g2					+	*trans*
nad1i728g2		+	++		+	++
nad2	+	+	+	+	+	+
nad2i156g2		+		+	+	+
nad2i542g2				+	+	*trans*
nad2i709g2	+		+	+	+	+
nad2i830g2				+	+	
nad2i1282g2			+			+
nad3	+	+	+	+	+	+
nad3i52g2			++	+	+	
nad3i140g2	+		++	+	+	
nad4	+	+	+	+	+	+
nad4i461g2		+	++	+	+	+
nad4i548g2	+					
nad4i976g2			++		+	+
nad4i1399g2				+	+	+
nad4L	+	+	+	+	+	+
nad4Li100g2	+					
nad4Li283g2	+	+				
nad5	+	+	+	+	+	+
nad5i230g2		+	+			+
nad5i753g1	+	+				
nad5i1242g2				+	+	
nad5i1455g2		++	++	+	+	*trans*
nad5i1477g2			++	+	+	*trans*
nad5i1872g2						+
nad6	+	+	+	+	+	+
nad6i444g2			+			
nad7	Ψ	+		+	+	+
nad7i140g2		+			+	+
nad7i209g2		+		+	+	+
nad7i336g2	+					
nad7i676g2				+	+	+
nad7i917g2				+	+	+
nad7i1113g2	+			+		
nad9	+	+	+	+	+	+
nad9i246g2			+			
nad9i283g2		+				
nad9i502g2			+			
rpl2	+	+				+
rpl2i28g2	+					
rpl2i917g2						+
rpl5	+	+	Ψ	+		+
rpl6	+	+	Ψ			
rpl10	+	+	+			
rpl16	+	+				+
rps1	+	+	Ψ	Ψ		+
rps2	+	+		+		+
rps3	+	+		+		+
rps3i74g2				+		+
rps3i257g2						++
rps4	+	+	Ψ	+		+
rps7	+	+	Ψ			+
rps8	+		Ψ			
rps10	+					+
rps10i235g2						+
rps11	+	+	Ψ			+
rps12	+	+	Ψ			+
rps13	+	+	+			+
rps14	+	+	+			+
rps14i114g2	+					
rps19	+	+				+
rrn5	+	+	+	+		+
rrnL	+	+	+	+	+	+
rrnLi827g2	+					
rrnS	+	+	+	+	+	+
rrnSi839g1				+	+	
rrnSi1065g2	++					
sdh	+	+	Ψ	+		+
sdh3i100g2		+	+			
sdh4	+	+	+			
tatC	+	+	+	+	+	+
trnA(ugc)	+	+	+			
trnC(gca)	+	+	+	+		+
trnD(guc)	+	+	+			+
trnE(uuc)	+	+	+	+		+
trnF(gaa)	+	+	+	+		+
trnG(gcc)	+	+	+	+		+
trnG(ucc)	+	+				
trnH(gug)	+	+	+			
trnI(cau)	+	+	+	+		+
trnK(uuu)	+	+	+	+		+
trnL(caa)	+	+	+			+
trnL(uag)	+	+		Ψ		+
trnL(uaa)	+	+	+			
trnM(cau)	+	+	+	+		+
trnMf(cau)	+	+	+	+		+
trnN(guu)	+					+
trnP(ugg)	+	+	+	+		+
trnQ(uug)	+	+	+	+		+
trnR(acg)	+	+				
trnR(ucg)	+					
trnR(ucu)	+	+				+
trnS(gcu)	+					+
trnSi43g2	+					
trnS(uga)	+	+		+		+
trnS(gga) cp						+
trnT(ggu)	+	+	+			
trnV(uac)	+	+				
trnW(cca)	+	+	+	+		+
trnY(gua)	+	+	+	+		+
∑ genes	**70**	**66**	**41**	**40**	**20**	**62**
∑ introns	**32**	**27**	**30**	**30**	**37**	**27**

Table 1

Table 2

Genes	*Ppat*	*Ieng*	*Smoe*	*Bvul*	*Slat*	*Snoc*	*Bnap*	*Atha*	Osat
	numbers of editing sites								
***nad*1**	0	57	137	20	19	11	23	24	23
***nad*2**	0	100	**249**	24	21	18	25	32	30
***nad*3**	1	57	77	12	8	5	10	12	15
***nad*4**	1	156	**237**	19	16	11	35	32	20
***nad*4L**	0	27	47	10	9	6	9	9	10
***nad*5**	2	**166**	143	17	18	15	29	27	11
***nad*6**	0	64	89	11	10	6	11	10	18
***nad*7**	0	115	103	20	19	9	28	27	32
***nad*9**	0	50	72	5	5	1	8	7	12
***sdh*3**	0	23							
***sdh*4**	0			4					
cob	0	121	122	13	9	6	8	7	19
***cox*1**	1	110	**181**	0	0	0	1	0	4
***cox*2**	1	14	97	9	3	2	13	15	19
***cox*3**	1	101	133	4	1	1	7	8	1
***atp*1**	0	131	160	3	3	0	5	5	5
***atp*4**	0	6		12	11	6	8	8	9
***atp*6**	0	95	80	12	11	7	1	1	17
***atp*8**	0	35	35	2	2	2	3	0	4
***atp*9**	1	34	44	5	4	1	4	4	8
***ccm*B**	0			30	27	19	39	39	35
***ccm*C**	0			28	23	15	25	28	36
***ccm*FN(1)**	0			23	22	16	15	22	31
***ccm*FN(2)**							10	12	
***ccm*FC**	2			13	12	8	13	16	27
***rpl*2**	0						2	1	1
***rpl*5**	0	30		5	6	4	9	10	1
***rpl*6**	0								
***rpl*10**	0				n.i.	n.i.			
***rpl*16**	0						6	5	12
***rps*1**	0								3
***rps*2**	0	28							10
***rps*3**	0	62		6	4	3	8	13	10
***rps*4**	0	55		11			19	15	15
***rps*7**	0			3			1	0	2
rps8	0								
***rps*10**									
***rps*11**	0								4
***rps*12**	0			6			7	8	0
***rps*13**	0			2	1	0			8
***rps*14**	1						0		0
rps19	0								6
***tat*C**	0	68	133	19	15	11	27	24	33
***mat*R**				9	8	6	8	9	n.i.
total	**11**	**1705**	**2139**	**357**	**287**	**189**	**417**	**430**	**491**

Table 3

Chondrome features	***Isoetes engelmannii***	***Selaginella moellendorffii***
RNA editing		
mRNA, C-to-U sites	1560	2139
mRNA, U-to-C sites	222	0
in rRNAs	0	13
Genes and introns		
proteins	24	18
tRNAs	13	0
rRNAs	3	2
group I introns (*trans*)	2 (1)	2 (1)
group II introns (*trans*)	27 (0)	34 (3)
intron size range	327-2072	516-6932
average intron size[1]	526	2455
Sequence characteristics		
GC content	48.7%	68.1%
cp inserts	1	0
nuc inserts	2	0
microsatellites[2]	6	90

[1] Considering *cis*-splicing introns only
[2] With motif repeats >4

Figure 1

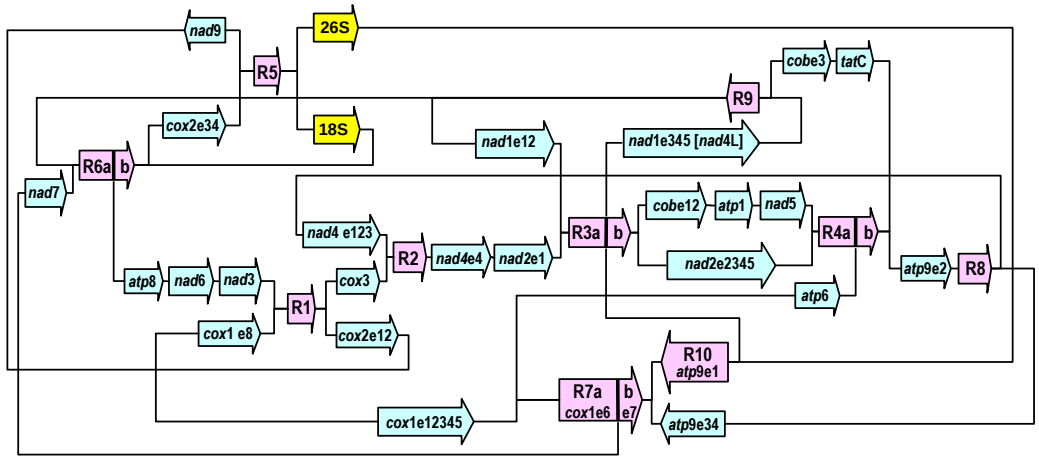

nad9
26S
R5
cobe3
tatC
R9
cox2e34
18S
nad1e12
nad1e345 [nad4L]
R6a
b
nad7
cobe12
atp1
nad5
nad4 e123
R3a
b
R4a
b
R2
nad4e4
nad2e1
nad2e2345
atp9e2
R8
atp8
nad6
nad3
cox3
R1
atp6
cox1 e8
cox2e12
R10
atp9e1
R7a
cox1e6
b
e7
cox1e12345
atp9e34

Figure 2

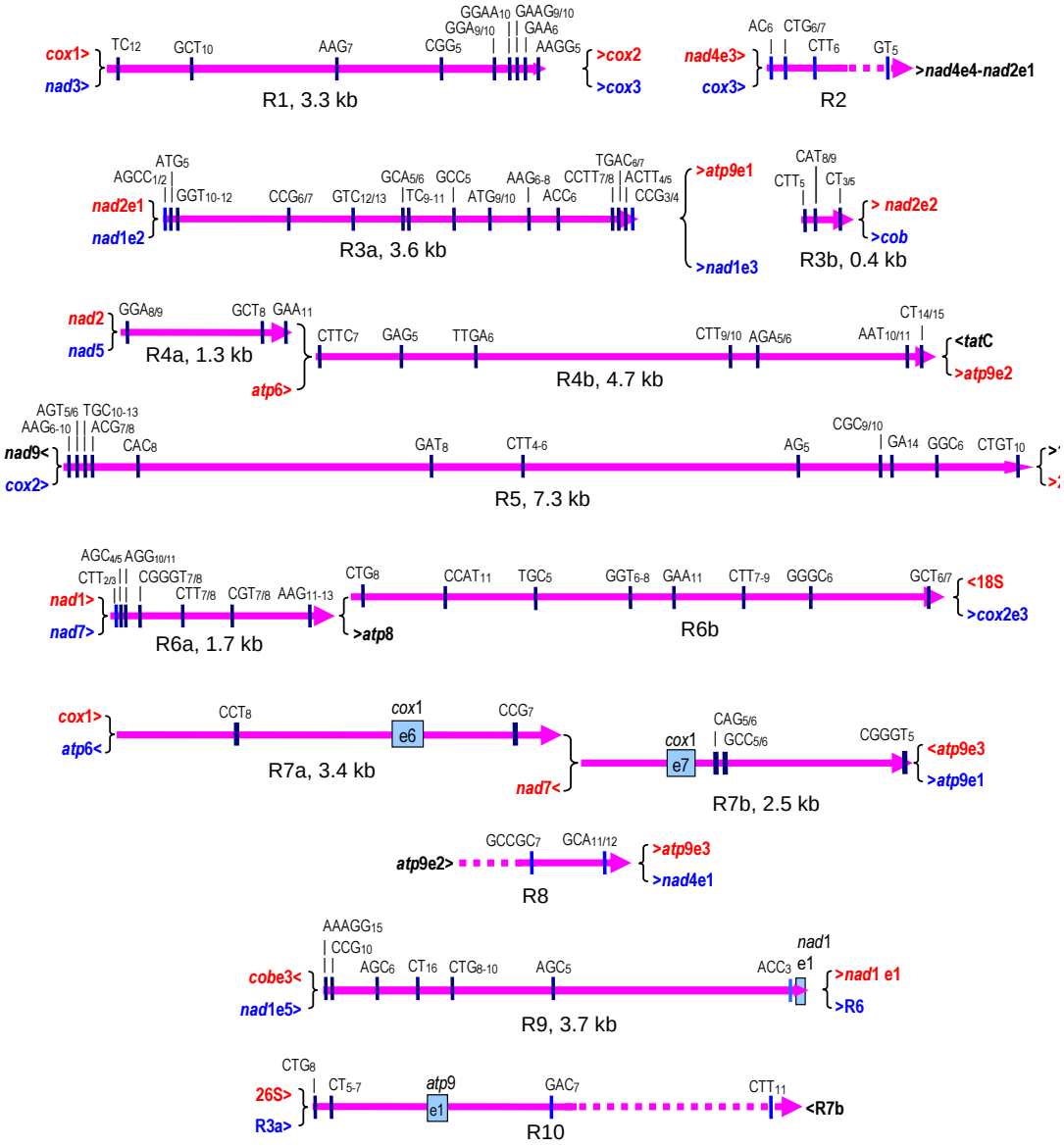

cox1>
nad3>
TC12
GCT10
AAG7
CGG5
GGAA10 GAAG9/10
GGA9/10 GAA6
AAGG5
>cox2
>cox3
R1, 3.3 kb
nad4e3>
cox3>
AC6 CTG6/7
CTT6
GT5
>nad4e4-nad2e1
R2
nad2e1
nad1e2
ATG5
AGCC1/2
GGT10-12
CCG6/7
GTC12/13
GCA5/6 GCC5
TC9-11 ATG9/10
AAG6-8 ACC6
TGAC6/7
CCTT7/8 ACTT4/5
CCG3/4
>atp9e1
>nad1e3
R3a, 3.6 kb
CAT8/9
CTT5 CT3/5
> nad2e2
>cob
R3b, 0.4 kb
nad2
nad5
GGA8/9
GCT8 GAA11
atp6>
CTTC7 GAG5
TTGA6
R4a, 1.3 kb
CTT9/10 AGA5/6
AAT10/11
CT14/15
<tatC
>atp9e2
R4b, 4.7 kb
nad9<
cox2>
AGT5/6 TGC10-13
AAG6-10 ACG7/8
CAC8
GAT8
CTT4-6
AG5
CGC9/10
GA14 GGC6 CTGT10
R5, 7.3 kb
nad1>
nad7>
AGC4/5 AGG10/11
CTT2/3 CGGGT7/8
CTT7/8 CGT7/8 AAG11-13
CTG8
CCAT11 TGC5
GGT6-8 GAA11
CTT7-9 GGGC6
GCT6/7
<18S
>cox2e3
>atp8
R6a, 1.7 kb
R6b
cox1>
atp6<
CCT8
cox1
e6
CCG7
nad7<
cox1
e7
CAG5/6
GCC5/6
CGGGT5
<atp9e3
>atp9e1
R7a, 3.4 kb
R7b, 2.5 kb
atp9e2>
GCCGC7
GCA11/12
>atp9e3
>nad4e1
R8
cobe3<
nad1e5>
AAAGG15
CCG10
AGC6
CT16
CTG8-10
AGC5
nad1
ACC3 e1
>nad1 e1
>R6
R9, 3.7 kb
26S>
R3a>
CTG8
CT5-7
atp9
e1
GAC7
CTT11
<R7b
R10

Figure 3

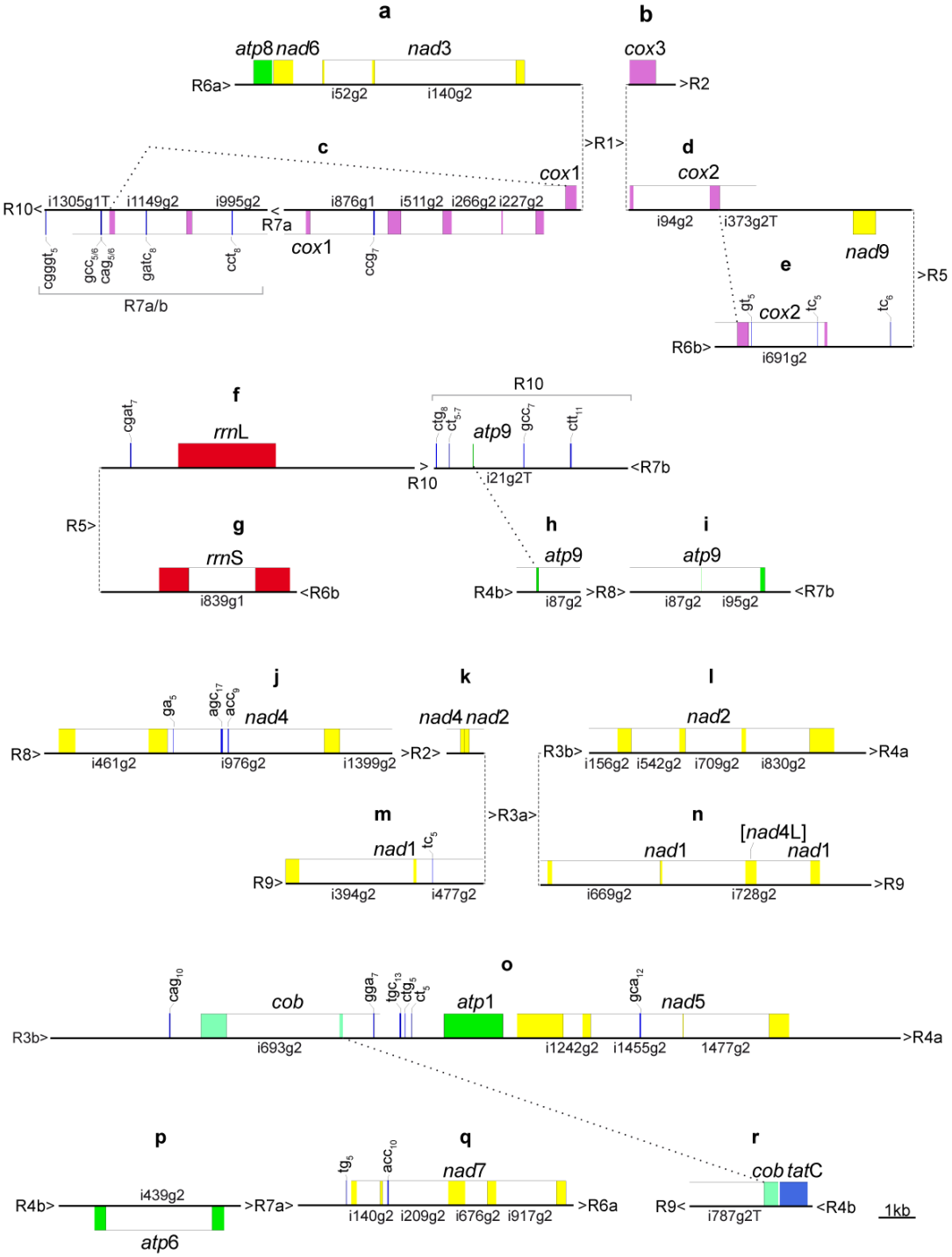

a
atp8 nad6
nad3
R6a>
i52g2
i140g2
b
cox3
>R2
>R1>
c
cox1
R10<
i1305g1T
i1149g2
i995g2
R7a
i876g1
i511g2
i266g2
i227g2
cox1
cggt
gcc
cag
gatc
cct
ccg
R7a/b
d
cox2
i94g2
i373g2T
nad9
>R5
e
gt
cox2
tc
tc
R6b>
i691g2
f
cgat
rrnL
R10
ctg
ct
atp9
gcc
ctt
<R7b
i21g2T
R10
R5>
g
rrnS
i839g1
<R6b
h
atp9
R4b>
>R8>
i87g2
i
atp9
i87g2
i95g2
<R7b
j
ga
agc
acc
nad4
R8>
i461g2
i976g2
i1399g2
>R2>
k
nad4 nad2
l
nad2
R3b>
i156g2
i542g2
i709g2
i830g2
>R4a
m
nad1
tc
R9>
i394g2
i477g2
>R3a>
n
[nad4L]
nad1
nad1
i669g2
i728g2
>R9
o
cag
cob
gga
tgc
ctg
ct
atp1
gca
nad5
R3b>
i693g2
i1242g2
i1455g2
1477g2
>R4a
p
i439g2
R4b>
>R7a>
atp6
q
tg
acc
nad7
i140g2
i209g2
i676g2
i917g2
>R6a
r
cob tatC
R9<
i787g2T
<R4b
1kb

Figure 4

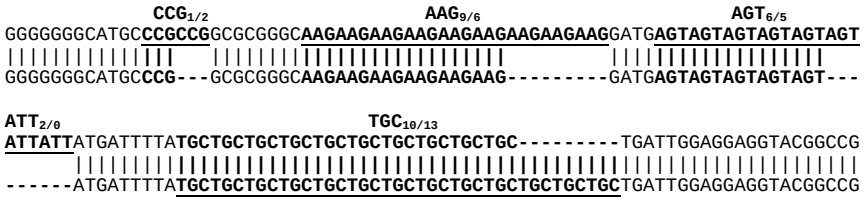

CCG1/2
AAG9/6
AGT6/5
GGGGGGGCATGCCCGCCGGCGCGCGGGCAAGAAGAAGAAGAAGAAGAAGAAGGATGAGTAGTAGTAGTAGTAGT
GGGGGGGCATGCCCG---GCGCGGGCAAGAAGAAGAAGAAGAAG---------GATGAGTAGTAGTAGT---
ATT2/0
TGC10/13
ATTATTATGATTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGC---------TGATTGGAGGAGGTACGGCCG
------ATGATTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGATTGGAGGAGGTACGGCCG

Figure 5

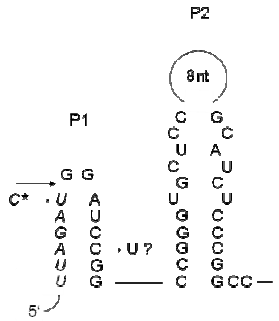
```
                                                                    50
ACGGAACTCGCACCAATAGGACCCCATGCCGTGATCAGTTCGCCGCCCTC
T>M E L>L A  P  I  G P>FH>YA>V V  I  S S>LP>LP>L S

                                                                   100
CCCGATCCCTATCGGTGTCTCTCCCCTTGGGGCGGGGAGTGGTCCAGAGA
 P>L I P>L I  G  V  S P>F L  G  A  G  S  G  P  E

                                                                   150
AAGCGCCGGCTCACGAGCGTGGTCCCGACCCACCCGATGATGCCAGAAAT
K A>VP>S A H>Y E R>C G P>F D  P P>F D  D  A  R  N

                                                                   200
CGTCCCGAGACAAGACCTCATCCTGTTCCCACATCACCCACCACCCCCGA
 R P>F E T>I R P>FH>YP>L V P>ST>IS>LP>FT>IT>IP>F D

                                                                   250
TTCGGAAGTCACCTCCTCCCCTCCCTGGGCCATCTCCCCCAAGTGGATCG
 S>L E  V  T S>FS>FP>F P  W  A  I  S P>L K  W  I

                                                                   300
ATCTGCTCGGATCCCGGTCCATGATCGTGCCCTCCCCGATACCGACCATA
D  L L>F G S>FR>W S  M  I  V P>FS>FP>L I P>L T  I

                                                                   350
GGGTCCTTCCATGAACGGAGACAGGGGGCGCCAGATCGGCACTCATCCCC
 G S>F F H>Y E R>W R  Q  G  A P>L D R>W H  S  S  P

               369
GATCGACCGGGGGGCTTGA
  I  D  R  G  A  *
```

Figure 6

A

```
                                                                   330
Selaginella AGCCCGCGTC GTACCAGGCG GCCGGTGAGT TCGGTAAAGG CCCACCGAGC
Pinus       AGCCCACGTA GTATCAGGTA GTTGGTTAG- ---GTAAAGG CTGACCAAGC

                                                                   380
Selaginella CGATGATGCG TGGCTGGTCT CTTCGGGGAT GATCAGCCAC ACCGGGACCG
Pinus       CAATGATGCT TAGCTGGTCT TTTCGG--AT GATCAGCCAC ACTGGGACTG

                                                                   429
Selaginella AGAGACGGCC CGGGCTGATT C-CGAGGCAG CAGCGGGGGA TATCGGGCAC
Pinus       AGACACGGCC CAGACTCCCT CGGGGGGCAG CAGTGGGGAA TATTGGACAA

                                                                   476
Selaginella CGGGCCCGCG CCCGACCCAG CAATGTCGCG TGAGTGATGA CGGGCA---T
Pinus       TGGGCTCACG CCCGATCCAG CAATATCGCG TGGGTGAAGA AGGGCACTGT

                                                                   522
Selaginella GGCTTGTAAA GCCCCATACT CG---GCCGA GTAT-CCCGC ACCGGCCGAG
Pinus       CGCTTGTCAA GCTCTTTCGT CGAGTTTCGA TCATGACAGG ACTCGAGGAA
```

B



Literature Cited

Abdelnoor RV, Yule R, Elo A, Christensen AC, Meyer-Gauen G, Mackenzie SA**.** 2003. Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. Proc Natl Acad Sci U S A. 100(10):5968-5973.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD**.** 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(15):9905-9912.

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ**.** 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics. 177(2):1173-1192.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ**.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res. 25(17):3389-3402.

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD**.** 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 27(6):1436-1448.

Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S**.** 2001. Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. Genetics. 158(2):851-864.

Arrieta-Montiel MP, Mackenzie S. A. 2011. Plant mitochondrial genomes and recombination. In: Kempken F, editor. Plant Mitochondria. New York: Springer. p. 65-82.

Banks JA**.** 2009. Selaginella and 400 million years of separation. Annu Rev Plant Biol. 60:223-238.

Bendich AJ**.** 1993. Reaching for the ring: the study of mitochondrial genome structure. Curr Genet. 24:279-290.

Bergthorsson U, Adams KL, Thomason B, Palmer JD**.** 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature. 424(6945):197-201.

Bullerwell CE, Schnare MN, Gray MW**.** 2003. Discovery and characterization of Acanthamoeba castellanii mitochondrial 5S rRNA. RNA. 9(3):287-292.

Chaw SM, Chun-Chieh SA, Wang D, Wu YW, Liu SM, Chou TY**.** 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short

interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol. 25(3):603-615.

Dombrovska E, Qiu YL. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol. 32(1):246-263.

Duchene AM, El Farouk-Ameqrane S., Sieber F., Maréchal-Drouard L. 2011. Import of RNAs into plant mitochondria. In: Kempken F, editor. Plant Mitochondria. New York: Springer. p. 241-260.

Duchene AM, Pujol C, Marechal-Drouard L. 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. Current Genetics. 55(1):1-18.

Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell. 53(3):413-422.

Giegé P, Brennicke A. 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. Proc Natl Acad Sci U S A. 96(26):15324-15329.

Glover KE, Spencer DF, Gray MW. 2001. Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria. J Biol Chem. 276(1):639-648.

Grewe F, Herres S, Viehoever P, Polsakiewicz M, Weisshaar B, Knoop V. 2010. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. Nucl Acids Res. in press (DOI:10.1093/nar/gkq1227).

Grewe F, Viehoever P, Weisshaar B, Knoop V. 2009. A *trans*-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. Nucl Acids Res. 37(15):5093-5104.

Groth-Malonek M, Wahrmund U, Polsakiewicz M, Knoop V. 2007. Evolution of a pseudogene: Exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. Mol Biol Evol. 24(4):1068-1074.

Haen KM, Pett W, Lavrov DV. 2010. Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA synthetases and mtDNA-encoded tRNAs in Cnidaria. Mol Biol Evol. 27(10):2216-2219.

Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. Nucl Acids Res. 31(20):5907-5916.

Kim E, Lane CE, Curtis BA, Kozera C, Bowman S, Archibald JM. 2008. Complete sequence and analysis of the mitochondrial genome of Hemiselmis andersenii CCMP644 (Cryptophyceae). BMC Genomics. 9:215.

Kim S, Yoon MK. 2010. Comparison of mitochondrial and chloroplast genome segments from three onion (*Allium cepa* L.) cytoplasm types and identification of a *trans*-splicing intron of *cox2*. Curr Genet. 56(2):177-188.

Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 46(3):123-139.

Knoop V**.** 2010. When you can't trust the DNA: RNA editing changes transcript sequences. Cell Mol Life Sci. 68:567-586.

Knoop V, Rüdinger M**.** 2010. DYW-type PPR proteins in a heterolobosean protist: Plant RNA editing factors involved in an ancient horizontal gene transfer? FEBS Lett. 584(20):4287-4291.

Knoop V, Volkmar U., Hecht J., Grewe F. 2010. Mitochondrial genome evolution in the plant lineage. In: Kempken F, editor. Plant Mitochondria. New York: Springer. p. 3-29.

Kumar S, Nei M, Dudley J, Tamura K**.** 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. 9(4):299-306.

Lang BF, Goff LJ, Gray MW**.** 1996. A 5 S rRNA gene is present in the mitochondrial genome of the protist Reclinomonas americana but is absent from red algal mitochondrial DNA. J Mol Biol. 261(5):407-413.

Lenz H, Rüdinger M, Volkmar U, Fischer S, Herres S, Grewe F, Knoop V**.** 2009. Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. Curr Genet. 56(2):189-201.

Li L, Wang B, Liu Y, Qiu YL**.** 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. J Mol Evol. 68(6):665-678.

Lohse M, Drechsel O, Bock R**.** 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 52(5-6):267-274.

Lowe TM, Eddy SR**.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucl Acids Res. 25(5):955-964.

Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffmann B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I**.** 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell. 16(8):2089-2103.

Malek O, Knoop V**.** 1998. Trans-splicing group II introns in plant mitochondria: the complete set of *cis*-arranged homologs in ferns, fern allies, and a hornwort. RNA. 4(12):1599-1609.

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(6):1403-1411.

Manchekar M, Scissum-Gunn K, Song D, Khazi F, McLean SL, Nielsen BL**.** 2006. DNA recombination activity in soybean mitochondria. J Mol Biol. 356(2):288-299.

Mower JP, Palmer JD**.** 2006. Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. Mol Genet Genom. 276(3):285-293.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K**.** 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Mol Genet Genom. 268(4):434-445.

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol. 223(1):1-7.

Rüdinger M, Funk HT, Rensing SA, Maier UG, Knoop V. 2009. RNA editing: 11 sites only in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. Mol Genet Genom. 281:473-481.

Salone V, Rüdinger M, Polsakiewicz M, Hoffmann B, Groth-Malonek M, Szurek B, Small I, Knoop V, Lurin C. 2007. A hypothesis on the identification of the editing enzyme in plant organelles. FEBS Lett. 581(22):4132-4138.

Shedge V, Arrieta-Montiel M, Christensen AC, Mackenzie SA. 2007. Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell. 19(4):1251-1264.

Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. 2010a. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. BMC Evol Biol. 10:274.

Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. 2010b. Extensive loss of RNA editing sites in rapidly evolving silene mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics. 185(4):1369-1380.

Smith DR. 2009. Unparalleled GC content in the plastid DNA of Selaginella. Plant Mol Biol. 71(6):627-639.

Smith DR, Lee RW. 2008. Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. Mol Biol Evol. 25(3):487-496.

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? J Mol Evol. 48(3):303-312.

Tasaki E, Sugita M. 2010. The moss *Physcomitrella patens*, a model plant for the study of RNA editing in plant organelles. Plant Signal Behav. 5(6):1-3.

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N. 2006. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol Biol Evol. 24(3):699-709.

Ullrich H, Lättig K, Brennicke A, Knoop V. 1997. Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. Plant Mol Biol. 33(1):37-45.

Wang B, Xue J, Li L, Liu Y, Qiu YL. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. Curr Genet. 55(6):601-609.

Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell. 25(3):793-803.

Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C. 2010. Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. Plant Cell. 22(4):1033-1045.

Xue JY, Liu Y, Li L, Wang B, Qiu YL**.** 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. Curr Genet. 56(1):53-61.

Zehrmann A, Verbitskiy D, van der Merwe JA, Brennicke A, Takenaka M**.** 2009. A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of *Arabidopsis thaliana*. Plant Cell. 21(2):558-567.

	Total		Size	Editing frequency	C to U codon changes														Total		
	C>U	silent	[kb]	[no. sites / kb]	**A>V**	**H>Y**	**L>F**	**P>F**	**P>L**	**P>S**	**Q>***	**R>***	**R>C**	**R>W**	**S>F**	**S>L**	**T>I**	**T>M**	codon changes	Start	Stop
***nad* 1**	137	20	0.984	139	7	9	3	4	31	14	1	0	4	8	7	14	4	3	109	○	●
***nad* 2**	249	73	1,506	165	11	10	1	15	46	16	0	0	11	3	16	7	8	2	146	●	○
***nad* 3**	77	20	0.369	208	2	4	1	10	9	2	0	0	1	3	5	3	4	1	45	●	○
***nad* 4**	237	44	1,518	156	11	12	1	18	50	13	0	1	8	4	10	16	14	6	164	●	●
***nad* 4L**	47	7	0.303	155	2	2	0	1	14	3	0	0	0	0	4	3	3	5	37	●	○
***nad* 5**	143	40	2,022	70	8	7	2	11	20	8	0	0	1	6	10	8	3	5	89	●	○
***nad* 6**	89	14	0.534	166	9	2	1	8	21	3	0	0	3	2	4	5	1	1	60	●	○
***nad* 7**	103	18	1,194	86	6	10	0	3	26	5	0	0	2	3	5	6	1	7	74	○	○
***nad* 9**	72	10	0.624	115	4	5	0	4	13	7	0	0	0	5	6	4	2	1	51	●	○
cob	122	25	1,191	102	8	6	6	5	19	8	0	0	2	6	10	7	10	1	88	○	●
***cox* 1**	181	24	1,608	112	11	8	7	17	37	10	0	0	2	7	12	6	12	9	138	●	●
***cox* 2**	97	23	0.774	125	10	7	1	3	10	5	0	1	6	5	6	6	5	2	67	○	●
***cox* 3**	133	22	0.810	164	10	7	4	12	18	6	1	0	3	11	7	7	6	4	96	●	●
***atp* 1**	160	20	1,584	101	22	7	2	7	27	18	0	1	3	1	3	16	14	8	129	●	●
***atp* 6**	80	18	0.756	105	6	3	1	9	11	4	0	0	1	0	5	4	6	2	52	●	○
***atp* 8**	35	6	0.504	69	0	2	1	0	10	3	0	1	2	2	0	4	1	1	27	●	○
atp9	44	8	0.255	172	4	1	0	5	3	3	1	0	0	0	2	5	3	2	29	●	●
tatC	133	32	0.789	171	6	3	6	8	27	3	0	0	5	6	7	3	9	4	87	●	○
Total	2139	424			137	105	37	140	392	131	2	4	54	72	119	124	106	64	1488	14	8

Table 4.7: (Supplementary Table 2)

4.4 Die mitochondriale DNA von *Gleichenia dicarpa*

Wie bereits eingangs erwähnt (siehe Abschnitt 2.2), liegen aus den Landpflanzengruppen der Lycophyten und Monilophyten noch keine vollständige mtDNA Sequenzen vor. Nachdem, wie in Abschnitt 4.1 und 4.3 beschrieben, nun zwei Lycophyten, *Isoetes engelmannii* und *Selaginella moellendorffii*, sequenziert wurden, fehlt noch eine Monilophytensequenz, um auch diese phylogenetische Lücke zu füllen.

Diese Lücke soll durch die mtDNA-Sequenz des Monilophyten *Gleichenia dicarpa* geschlossen werden (siehe auch Abschnitt 2.8). Das Pflanzenmaterial für die Herstellung einer Fosmidbank dieses leptosporangiaten Farns wurde dazu freundlicherweise vom Botanischen Garten Bonn bereitgestellt.

Die Klonierung, das Durchsuchen der Bank per Hybridisierungen und die Sequenzierung der identifizierten Fosmide wurde, wie in Abschnitt 4.1 und 4.3 (*Material and Methods*) beschrieben, durchgeführt. Die *G. dicarpa* Gesamt-DNA wird in der hergestellten Bank durch 18.532 sortiert abgelegte Fosmide repräsentiert. Ergebnisse erster Hybridisierungen bestätigten, dass in dieser Bank – anders als bei *S. moellendorffii* (siehe Abschnitt 4.3) – gegenüber den großen Vorkommen an nukleärer DNA bei Farnen [Leitch *et al.* 2005, Nakazato *et al.* 2006] der Anteil an mtDNA unterrepräsentiert vorliegt. Aus diesem Grund wurde beispielsweise unter Verwendung einer *cox*2-Sonde nur ein einziger mitochondrialer Fosmid innerhalb der gesamten Bank identifiziert. Auch nach der Sequenzierung von acht Fosmidklonen mit einer mitochondrialen Nettosequenz über 177 kb ist eine vollständige mtDNA-Sequenz noch nicht erreicht. Vergleichsweise war eine Sequenzierung von fünf Fosmiden für den Erhalt der vollständigen mtDNA von *I. engelmannii* mit einer Nettosequenz von 57 kb ausreichend (siehe Abschnitt 4.1).

Trotz der noch nicht vollständigen mtDNA lassen sich interessante und bisher unerkannte Eigenheiten in den partiellen Chondromsequenzen der mitochondrialen Fosmidinserts nachweisen. Das anschließende Kapitel *Numerous mobile DNA elements including an intact ty1-copia retrotransposon characterize the mitochondrial genome of the fern Gleichenia dicarpa* beinhaltet eine Zusammenfassung der bisherigen Ergebnisse der Sequenzierungsarbeiten und schildert die neuentdeckten Eigenheiten aus der ersten umfassenden Analyse einer Monilophyten-mtDNA.

4.4.1 Numerous mobile DNA elements including an intact ty1-*copia* retrotransposon characterize the mitochondrial genome of the fern *Gleichenia dicarpa*

Overview of the fosmid cloning approach of the *G. dicarpa* mtDNA

Information on the *Gleichenia dicarpa* mtDNA sequence was assembled from fosmid clones identified in an arrayed library by hybridization with mitochondrial gene probes and verified in their mitochondrial nature through complete sequencing of the inserts. Eight mitochondrial fosmid clones were detected (Gd09-2N22, Gd10-17B23, Gd22-38P8, Gd49-35M14, Gd54-13N7, Gd61-35K18, Gd64-22C22, Gd75-45E12) and completely sequenced inserts were annotated to characterize mitochondrial genomic features of *G. dicarpa* (Figure 4.24). By sequencing these clones we obtained altogether 290,179 bp of *G. dicarpa's* mtDNA, however high frequency of recombination events resulting in co-existing alternative gene arrangements reduce the amount of unique mitochondrial sequences. The net mtDNA sequence complexity of the eight analyzed *G. dicarpa* fosmid clones is 177,474 bp, with an overall GC content of 46.4% and 9.4% of coding sequences (Table 4.8).

Table 4.8: Overview of fosmid cloning approach

Fosmid	mt genes	size [bp]	GC [%]	AT [%]	coding [bp]	net [bp]
Gd61-35K18	5	39350	45.47	54.53	3411	30428
Gd22-38P8	10	33320	48.99	51.01	3955	32472
Gd54-13N7	8	32796	50.13	49.87	7912	11780
Gd64-22C22	5	37138	46.94	53.06	2976	37138
Gd10-17B23	5	36281	50.94	49.06	2830	32123
Gd49-35M14	2	33533	44.87	55.13	6163	33533
Σ		212418	47.83	52.17	27247	177474
Gd75-45E12	0	36077	42.82	57.18	0	36077
Gd09-2N22	0	41684	42.10	57.90	0	41877
Σ		290179	46.39	53.61	27247	255428

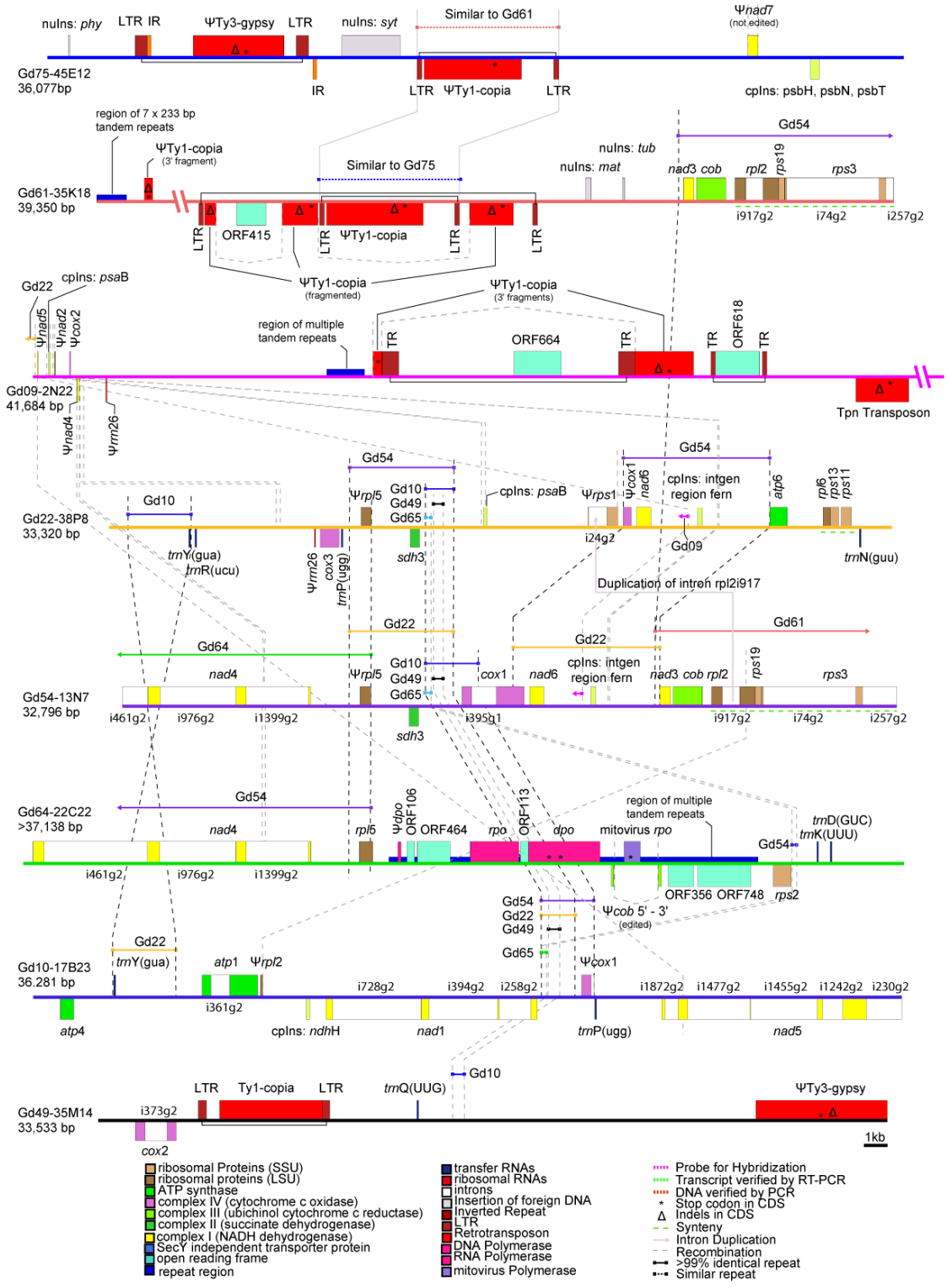


Figure 4.24: Gene maps of eight *Gleichenia dicarpa* mitochondrial fosmid clones. Protein-coding genes, ribosomal and transfer RNA genes and foreign DNA insertions, shown above or below the lines indicate directions of transcription to the right or to the left, respectively. Maps were generated with OGDRAW v1.2 software [Lohse *et al.* 2007] and manually modified.

The genes and introns identified in eight fosmid clones

We identified genes of a typical plant mitochondrial DNA complement encoding subunits of respiratory chain complex I (*nad*1, *nad*3, *nad*4, *nad*5, *nad*6), complex II (*sdh*3), complex III (*cob*), complex IV (*cox*1, *cox*2, *cox*3) and of complex V, the ATP synthase (*atp*1, *atp*4, *atp*6). Similarly, eight ribosomal proteins (*rpl*2, *rpl*5, *rpl*6, *rps*2, *rps*3, *rps*11, *rps*13, *rps*19) and seven intact tRNA genes: *trn*Y(gua), *trn*R(ucu), *trn*P(ugg), *trn*N(guu), *trn*K(uuu), *trn*D(guc) were found (Figure 4.24).

Genes of a typical plant mitochondrial gene set so far not identified are *nad*2, *nad*4L, *nad*7, *nad*9, *atp*8, *atp*9, *tat*C (encoding a subunit of the sec-independent transport pathway) and the genes encoding ribosomal RNAs. However, fragments of *nad*2, *nad*7 and *rrn*26 were identified in sequences of fosmids Gd09, Gd75 and Gd22, respectively. The *nad*7 fragment identified in Gd75 is in fact the only evidence of its mitochondrial origin, similar to fosmid Gd09, which was identified by its *nad*2 fragment. Other intact, full-length mitochondrial genes are missing from these two fosmid inserts for which reason their mitochondrial origins was not possible to be verified, e.g. by analyzing RNA editing in accompanying cDNA analyses. In contrast, the mitochondrial origins of all six other fosmid clones were clearly identified by verifying RNA editing of selected full-length genes they harbor as a characteristically mitochondrial feature (see 4.4.1).

A complement of 18 introns were identified in the 21 identified mitochondrial genes of *G. dicarpa* of which 17 were classified as group II introns and one as a group I intron, cox1i395g1. Three introns appear exclusively in monilophyte mtDNAs, atp1i362g2 [Wikström und Pryer 2005], nad1i258g2 [Dombrovska und Qiu 2004] and rps1i24g2. Three group II introns (nad5i1872g2, rpl2i917g2 and rps3i257g2) were apparently acquired by a common ancestor of monilophytes and seed plants and all remaining 12 group II introns have homologues in bryophytes, lycophytes and seed plants.

Repeated sequences were found in three different forms

Fragmented genes and repeated sequences indicate massive recombinational activity of the chondrome of *G. dicarpa*. Repeated sequences were found in three differing forms. Long repeats leading to redundant sequence stretches were identified by comparison of insert sequences. Five large regions of Gd54 were found

to be identically present in other fosmid inserts. These repeats are responsible for fragmented genes, e.g. *rpl*5 in Gd22 and Gd54 as well as *cox*1 in Gd10 and Gd22, of which complete intact gene copies were conversely found in different gene arrangements in Gd64 and Gd54, respectively (Figure 4.24). A full length version of the *rps*1 gene, fragmented in Gd22, is not yet available in another fosmid. Possibly, the start of *nad*3 and the end of the *cob* gene may as yet be misconcepted and alternative gene copies may actually reside at other locations.

Several sequence repeats of particularly short length (around 10-100 bp) are located at one end of the fosmid Gd09 insert. These contain small fragments of *nad*5, *nad*2, *cox*2, *nad*4 and *rrn*26 which could be assigned to the corresponding complete gene sequences when available (e.g. *nad*5 in Gd10, *cox*2 in Gd49, *nad*4 in Gd54 and Gd64). One would strongly expect complete coding sequences of mitochondrial *nad*2 and *rrn*26 in mtDNA of *G. dicarpa* as well although not present in the so far sequenced fosmid clones of this study. Numerous further short repeats not affecting coding sequences were also identified in this region by comparing with sequences of Gd22 and Gd54.

The third form of repeated sequence structures are directed tandem repeats as e.g. present at one end of fosmid Gd61, where a sequence stretch of 233 bp is repeated seven times until the fosmid insert breaks off (hence, this tandem repeat could even extend significantly longer in its true genomic location). Alternatively, direct sequence repeat motifs (around 2-66 bp) may be interrupted by unique sequence stretches like in another hotspot region present in Gd09 where from position 12268 different forward repeats shape the mtDNA for 1300 bp, then followed by a region of mobile elements. Likewise, in Gd64 the region of invasive elements from position 15067 to 30678 is interstratified by tandem repeats in contrast to other regions of this fosmid insert with no identified tandem repeat.

Conserved synteny for ribosomal gene clusters

Contrary to this massive recombinational activity resulting in repeats, conserved synteny was observed for ribosomal gene clusters *rpl*2-*rps*19-*rps*3 and *rpl*6-*rps*13-*rps*11 in Gd22, Gd54 and Gd61, respectively (Figure 4.24). Both gene continuities are also present in liverworts and moss mtDNAs [Oda *et al.* 1992, Terasawa *et al.* 2007, Wang *et al.* 2009] but got lost from mtDNAs of hornworts and lycophytes due to the frequent gene losses observed [reviewed in Knoop *et al.* 2011]. In hornworts only synteny of *rpl*6-*rps*13-*rps*11 was found with *rpl*6 and *rps*11 being

degraded to pseudogenes [Li *et al.* 2009, Xue *et al.* 2010]. In the lycophyte *Isoetes engelmannii* only *rps*3 is present [Grewe *et al.* 2009] whereas in another lycophyte *Selaginella moellendorffii* all genes of these cluster (and all other ribosomal protein genes) got lost (see 4.3). In contrast the highly conserved gene linkage *nad*4-*nad*2 obviously got lost in *G. dicarpa* with *rpl*5 following *nad*4 in Gd64 (if not present in an alternative arrangement which is not sequenced yet).

Peculiar RNA editing of mitochondrial genes and the verification of mitochondrial identity of fosmid clones

To gain insights into the peculiar mitochondrial feature of RNA editing, we analyzed cDNAs of ten mitochondrial genes (*atp*4, *nad*3, *cox*2, *cox*1, *cob*, *rps*1, *rps*11, *rps*13, *rpl*2, *rpl*6). All genes turned out to be RNA edited on transcript level, yet the frequencies of editing vary dramatically among these genes (Table

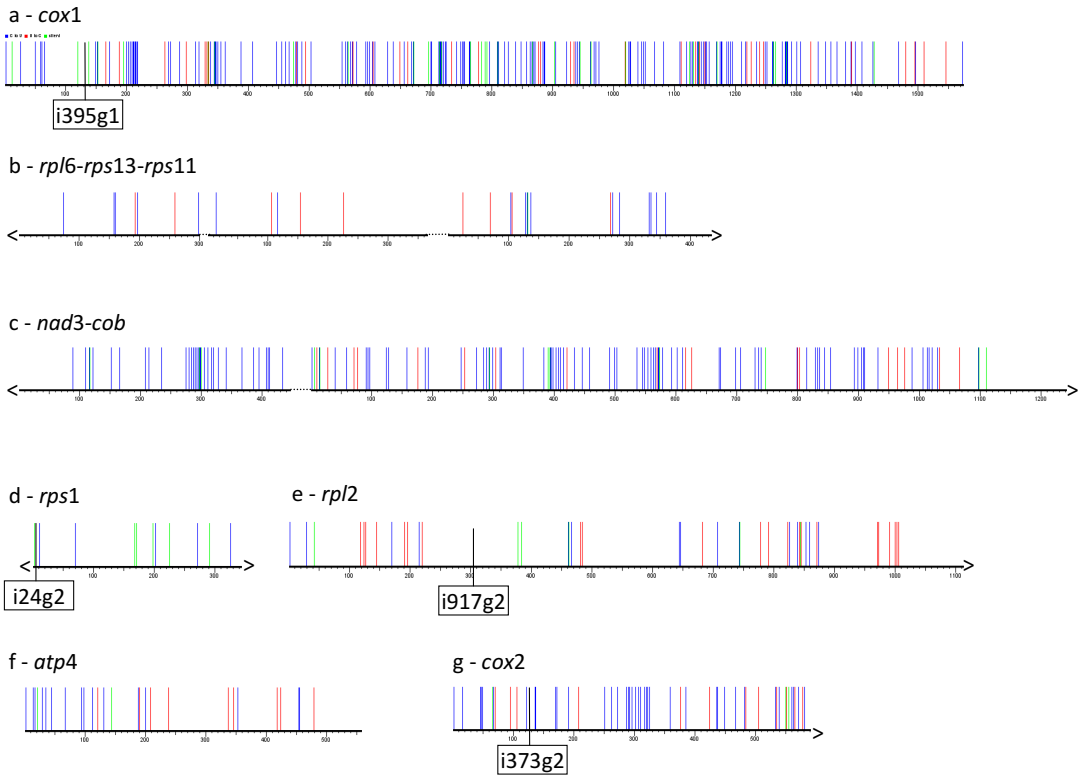


Figure 4.25: RNA editing sites in *G. dicarpa* mitochondrial genes. Non-silent C-to-U and U-to-C editings are indicated by blue or red lines, respectively, and green lines indicate silent editings. RNA-editing patterns are exemplarily shown for selected genes *cox*1 (**a**), the coherent co-transcripts of *rpl*6-*rps*13-*rps*11 (**b**) and *nad*3-*cob* (**c**), *rps*1 (**d**), *rpl*2 (**e**), *atp*4 (**f**) and *cox*2 (**g**). Group II intron insertion sites are indicated. Arrowheads signify partially analyzed reading frames.

4.9). On the one hand genes are heavy edited changing 217 pyrimidines (182 C-to-U and 35 U-to-C) in *cox*1 (Figure 4.25 a) even exceeding the hitherto highest *cox*1 editing number observed in *S. moellendorffii* affecting 181 sites (see 4.3).

On the other end of the spectrum the co-transcript of syntenic genes *rpl*6-*rps*13-*rps*11 is meagerly edited at 7, 5 and 15 sites, respectively (Figure 4.25 b). As outlined in figure 4.26, a marginal set of editing in co-transcript *rpl*6-*rps*13-*rps*11 represents a necessary complement of changes for proper protein function since editing re-establishes evolutionary conserved amino acid codons, corrects internal stop codons in reading frames and introduces an *rpl*6 stop codon. Most sites are partially edited which implies deficiently edited transcripts. Such low editing frequency is, however, not an generally attribute of unprocessed co-transcripts since a *nad*3-*cob* co-transcript is edited at 136 of 1689 positions (Figure 4.25 c).

Table 4.9: Overview of cDNA analysis of mitochondrial genes

gene	**freq.** [%]	**sum** [nt]	**C-to-U** [sites]	**U-to-C** [sites]	**analyzed** [nt]	**clone** [no.]	**remarks**
*cox*1	11.81	186	154	32	1575	1p55	
*rpl*6	2.56	7	5	2	273	cons.	co-trans
*rps*13	1.37	5	2	3	366	cons.	co-trans
*rps*11	3.71	15	11	4	404	cons.	co-trans
*nad*3	8.17	34	34	0	416	7p24	co-trans
cob	8.21	102	83	19	1242	7p24	co-trans
*rps*1	3.2	11	11	0	344	cons.	
*rpl*2	4.14	43	21	22	1038	7p28	
*atp*4	4.84	27	18	9	558	cons.	
*cox*2	9.05	54	42	12	597	1p68	

Further analyzes of individual *rps*1 and *rpl*2 transcripts reveal 11 and 43 editing sites within reading frames of 344 and 1038 nucleotides, respectively (Figure 4.25 d, e). In these two genes, cDNA analyses moreover verified correct intron splicing of rpl2i917g2 and its closely related relative rps1i24g2 (see below, 4.24 Gd22, Gd54 and 4.27). As a typical feature for monilophyte and hornwort mtDNAs and the one of the lycophyte [Grewe *et al.* 2010], RNA editing was observed in both directions of pyrimidine exchange in the mitochondrial transcripts of *G. dicarpa*. Notable exceptions were the *nad*3 and *rps*1 transcripts featuring exclusively cytidines edited into uridines at 34 and 11 positions, respectively (Figure 4.25 c, d), but not

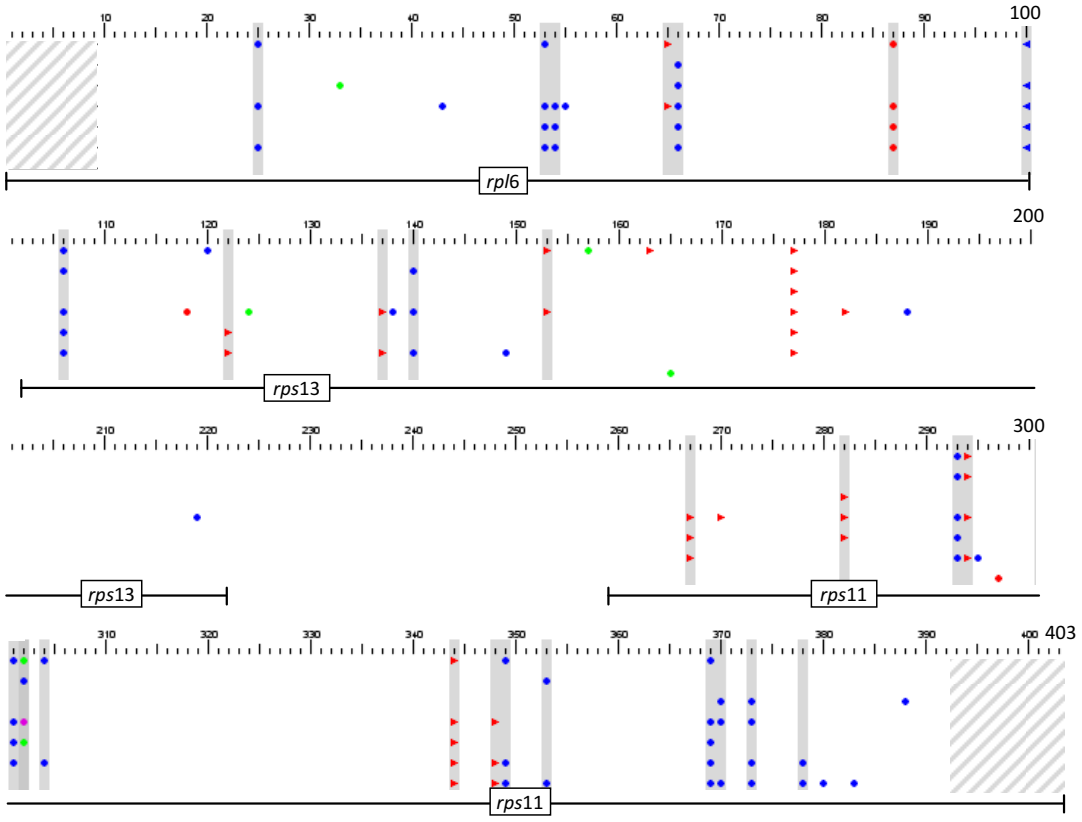


Figure 4.26: RNA-editing status in a population of seven cDNA clones of coherent co-transcripts of *rpl*6-*rps*13-*rps*11. The graphic is based on the cDNA variants overview of PREPACT [Lenz *et al.* 2010] with default settings: blue and red circles indicate codon sense changes derived from C-to-U or U-to-C editings, respectively, green circles indicate silent codon changes and purple circles reflect more than one editing in a codon. Forward arrows and reverse arrows indicate stop-codon removal and stop-codon creation, respectively. Positions of partial editing are highlighted with gray columns. Striped regions indicate primer binding sites.

in the reverse direction. C-to-U editing sites in *atp*4 are custering in upstream third and scattered reverse editing is distributed over the rest of the reading frame.

Identifying RNA editing as outlined above firmly corroborates the mitochondrial identities of fosmid inserts of Gd10, Gd22, Gd54, Gd61 and Gd64 carrying those coding sequences. Along the same lines, cDNA clones derived from RT-PCR with primer pairs anchoring in the *cox*2 5‘-UTR and intron cox2i373g2 or alternatively in the second exon of *cox*2 similarly seem to verify the mitochondrial nature of fosmid Gd49 insert: 54 editing sites (Figure 4.25 g) could be identified in the *cox*2 transcript in comparison to the genomic sequence of fosmid Gd49, which notably

harbors an apparently intact ty1-*copia* retrotransposon in close proximity to *cox*2 (see below and 4.24 Gd49). However, an indel of 6 nt in the Gd49 *cox*2 reading frame in comparison to the cDNA sequence (and the corresponding *cox*2 genes of other taxa such as *P. patens* and *C. taitungensis*) suggest the Gd49 *cox*2 copy to be a recently duplicated pseudogene copy.

Mitochondrial DNA of *Gleichenia* is massively interstratified by foreign mobile elements

As observed in former studies of mitochondrial DNA, integration of foreign DNA is suggested to be a consequence of recombination [Grewe *et al.* 2009]. The most surprising finding in partial mtDNA sequences of *G. dicarpa* is the massive presence of seven different kinds of foreign mobile elements (Introntransposition, promiscuous DNA originating from plastid and nuclear DNA, DNA and RNA polymerases derived from linear mitochondrial plasmids and mitoviruses, transposable elements, Ty3-*gypsy* and Ty1-*copia* retroelements), which was never observed before in such high frequency of occurrence.

Transposition of intron rpl2i917g2

Intron rps1i24g2 observed in mtDNA for the first time appears to be copied to its location in *rps*1 from the highly similar intron rpl2i917g2 in *G. dicarpa* (Figure 4.27). Considering the very highly conserved intron sequences (74% similarity in homologous positions), a very recent duplication scenario in that direction is most likely and makes the alternative scenario of reverse copying in a common ancestor of euphyllophytes and a subsequent loss of intron rps1i24g2 in the seed plant lineage less likely. Nevertheless, simultaneous existence of both introns seems to be a characteristic feature of monilophyte mtDNA (e.g. in the taxa *Todea barbata*, *Salvinia molesta* and *Dicksonia squarrosa*, unpublished observations). Interestingly, a similar, yet intragenic, recent intron duplication within *nad*5 was observed in former studies of mtDNA of *Huperzia* [Vangerow *et al.* 1998].

Promiscuous insertions from plastid and nuclear DNA

Promiscuous DNA representing integrated fragments of nuclear and chloroplast genes were identified at four and three positions, respectively. Fragments of nuclear phytochrome, synapotagmin, malonyltransferase and tubulin, ranging from 71 to 2458 bp in size, were identified in Gd61 and Gd75 inserts. Additionally in Gd75,

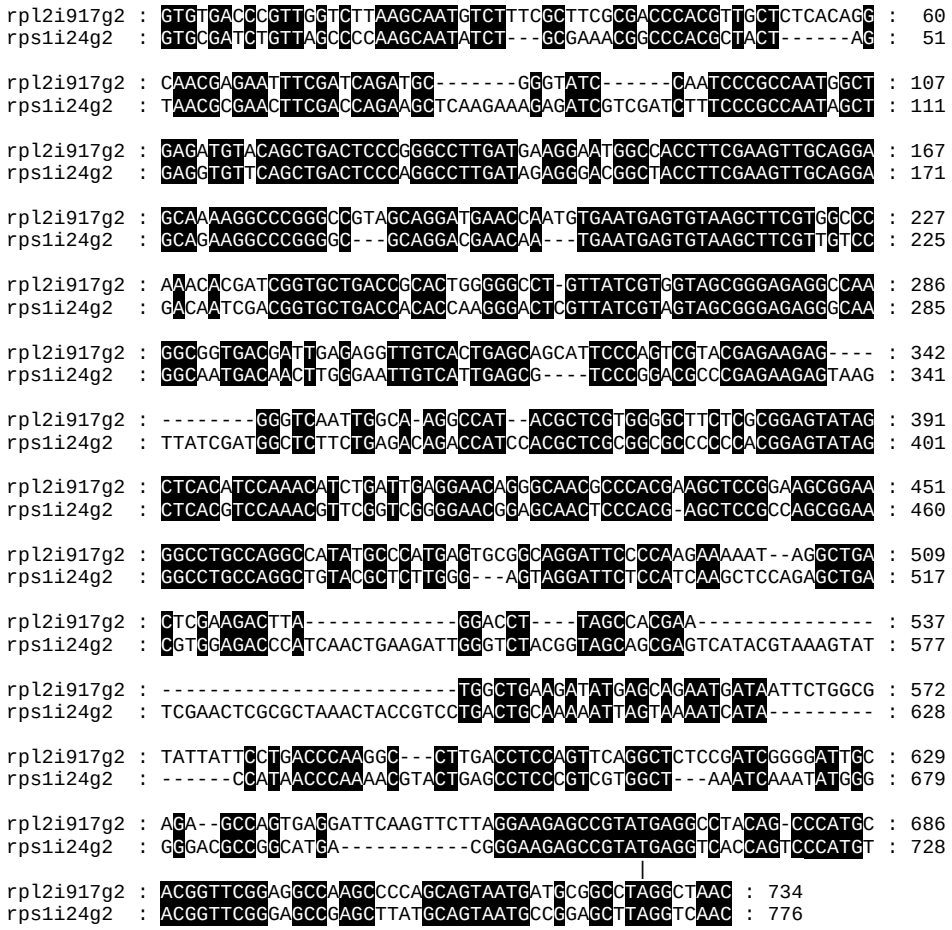


Figure 4.27: Alignment of introns rpl2i917g2 and rps1i24g2 of *G. dicarpa*. Nucleotide sequence alignment of intron rpl2i917g2 and its copy located in *rps*1 gene. Similar nucleotide positions are highlighted and contribute to 74% of homologous positions. Bar is indicating branching adenine located in domain six of both introns. Display and shading of sequence alignment was done with GeneDoc (http://www.nrbsc.org/gfx/genedoc/).

chloroplast photosystem II genes (*psb*H, *psb*N, *psb*T) were located covering 366 bp of its insert sequence. Further chloroplast fragments of photosystem II genes were detected in Gd22 and an intergenic region of fern chloroplast DNA were identified in Gd22 and Gd54 downstream of *nad*6 gene (Figure 4.24).

Polymerases from linear mitochondrial plasmids

The presence of linear self-replicating mitochondrial plasmids is already known in the plant lineage [Handa *et al.* 2002]. The DNA dependent polymerase (*dpo*) and RNA dependent polymerase (*rpo*) normally encoded on such linear plasmids were

also observed in an integrated form in the mtDNA [Gray *et al.* 2004]. Furthermore, transcription of integrated polymerases was verified in *Lolium perenne* and *Daucus carota* [McDermott *et al.* 2008, Robison und Wolyn 2004]. Analyzing the mtDNA of the monilophyte *G. dicarpa* reveals likewise open reading frames (ORF) for RNA dependent and DNA dependent polymerases in a head-to-tail arrangement separated by an ORF of unknown function typical for linear mitochondrial plasmid gene arrangements (Figure 4.24, Gd64).

Polymerases from mitovirus DNA

In close proximity of these, apparently linear plasmid-derived sequences in Gd64, resides an ORF coding for a mitovirus-type RNA dependent polymerase (Figure 4.24). Similar to the linear-plasmid like sequences, mitovirus *rpo* remnants have previously been detected in mtDNAs of *A. thaliana* and *Vicia faba* [Marienfeld *et al.* 1997]. Since the linear mitochondrial plasmid *dpo* and mitovirus *rpo* reading frames are interrupted by stop codons we aimed for cDNA analyses to verify if these genes are indeed not functional anymore or if stop codons might be corrected by RNA editing. Several attempts to amplify cDNAs from *dpo* failed and active transcription thus appears unlikely. With respect to mitovirus *rpo* integration, we were successful in amplifying transcripts with primer pairs binding in adjacent UTR and mitovirus *rpo* sequences, respectively. However, a stop codon was not corrected and other sequence differences to the Gd64 DNA sequence indicated that the amplified cDNA sequence was obviously transcribed from a different location.

Transposable elements

Transposable elements are divided into two major classes, according to their mode of propagation. In this study we identified elements of both classes to be integrated into mtDNA of *G. dicarpa*. A remnant of a classical transposable element, which moves by a DNA-DNA mechanism, was detected in Gd09 (Figure 4.24). This remnant of a transposon were found by BLASTX search identifying strong similarity (30% identical amino acids) to the TPN2 protein of a member of CACTA-type class II transposable element, called *Tgm9* present in soybean [Xu *et al.* 2010]. However, internal stop codons and frame shifts indicate that the transposon locus in Gd09 is defective.

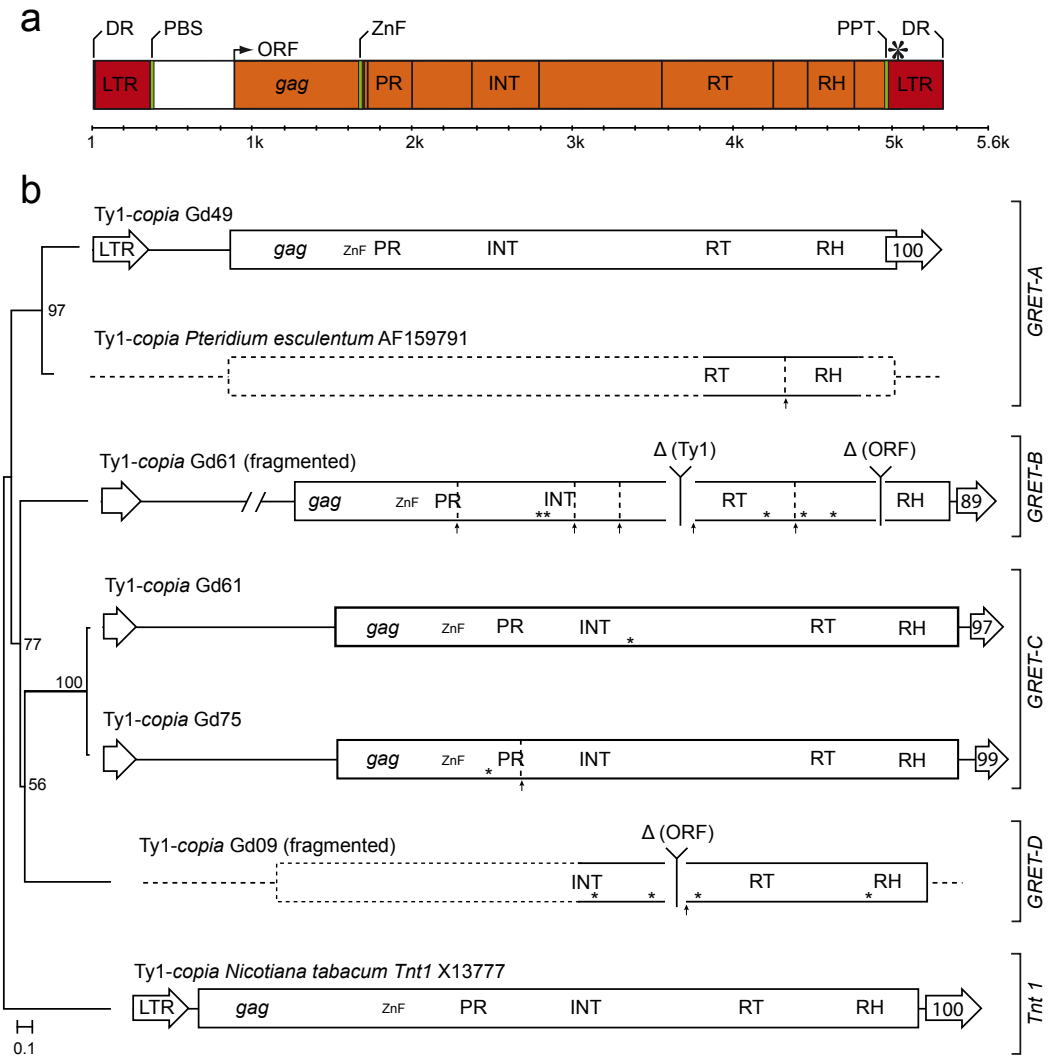


Figure 4.28: Correlation of Ty1-*copia* retrotransposons of *G. dicarpa*.
(**a**) Schematic demonstration of *Gleichenia* Ty1-*copia* retrotransposon A (*GRET-A*) of fosmid Gd49. Intact open reading frame is characterized by domain *gag* with a zinc finger motif (ZnF) and domains encoding for a protease (PR), integrase (INT), reverse transcriptase (RT) and RNase H (RH). Asterisk is indicating stop codon. Additional structural features for transposition are a primer binding site (PBS) and a polypurine tract (PPT). Long terminal repeats (LTR) bordering Ty1-*copia* retrotransposons are flanked by short direct repeats (DR) of the host target DNA, created upon insertion. (**b**) Schematic alignment of all Ty1-*copia* retrotransposons identified in fosmid inserts of *G. dicarpa*, the Ty1-*copia* sequence of *P. esculentum* and tobacco *Tnt1*. Phylogenetic tree is based on DNA alignment of Ty1-*copia* retrotransposon domains RT and RH (NJ tree [Saitou und Nei 1987] based on K2P distances [Kimura 1980], pairwise gap deletion, with bootstrap support from 1000 replicates [Felsenstein 1985] was conducted in MEGA5 [Tamura *et al.* 2007]). Reading frames of fragmented retrotransposons *GRET-B* and *GRET-D* are interrupted by another Ty1-*copia* retrotransposon *GRET-C* and/or regions containing an open reading frame (ORF). Upward directed arrows are indicating frame shifts. Numbers in LTRs are indicating percentage of similarity of Ty1-*copia* LTRs, respectively.

Ty3-*gypsy* retrotransposons

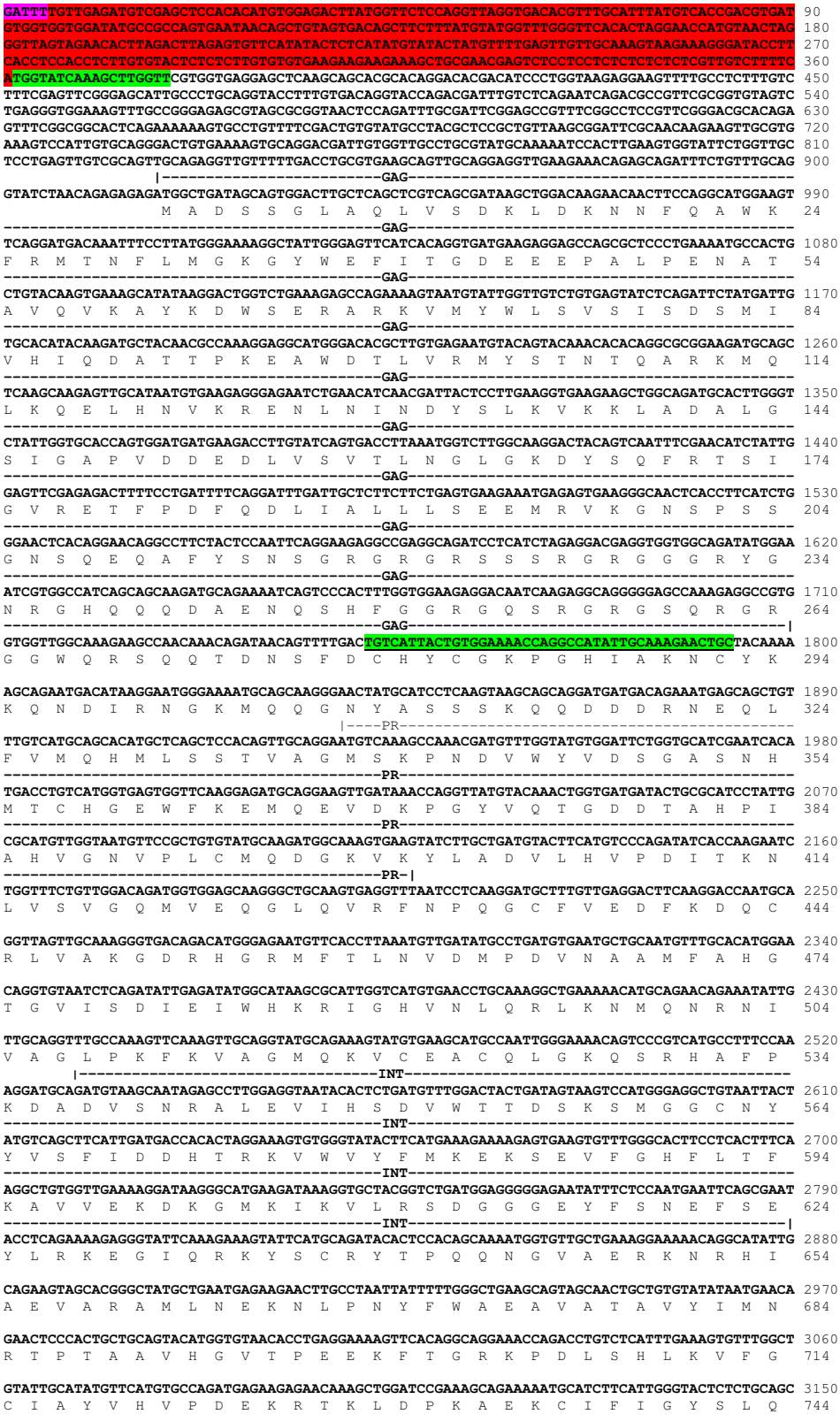
Similar to retroviruses, LTR-retrotransposons can move within a host genome from one position to another through an RNA intermediate by a "copy-and-paste" mechanism. Retrotransposons of this class are characterized by coding sequences homologous to the nucleic-acid binding (gag) and pol polyprotein of retroviruses and carry flanking Long Terminal Repeat (LTR) sequences.

We found two groups of LTR-retrotransposons to be present in the mtDNA of *G. dicarpa*, Ty1-*copia* and Ty3-*gypsy* retrotransposons. In fosmids Gd49 and Gd75 two different Ty3-*gypsy* retrotransposon homologies were identified, respectively. In this retrotransposon group the organization of coding domains of *pol* polyprotein is similar to that of retroviruses. In Gd75, LTRs with 97% identity are flanking the Ty3-*gypsy* retrotransposon ORF. The Ty3-*gypsy* retrotransposon homology in Gd49, however, is cut off with the end of the fosmid insert, thus no LTR can be identified. Reading frame parts of both Ty3-*gypsy* retroelements are frazzled by frameshifts and stop codons, thus proliferation activity of both Ty3-*gypsy* retroelements can be excluded.

Ty1-*copia* retrotransposons

With regard to the Ty1-*copia* retrotransposons, the mtDNA of *G. dicarpa* reflects a more vivid picture of active retrotransposition in recent evolutionary times.

Ty1-*copia* retroelements are characterized by an endonuclease domain (INT) upstream of the reverse transcriptase coding domain different to that of retroviruses (Figure 4.28 a). Within four of the analyzed fosmids five extended homologies with this group of retroelements were found (Gd09, Gd49, Gd61 and Gd75). Four of them are framed by LTRs with identities from 89% to 100% (Figure 4.28 b). For Ty1-*copia* homology in Gd09 no flanking LTR was identified since only a 3' fragment of this Ty1-*copia* ORF was found. This fragment itself is interrupted by the integration of ORF664 flanked by terminal repeats. Another fragmented Ty1-*copia* retrotransposon was found in Gd61. The three parts are interrupted by an additional full length Ty1-*copia* retrotransposon flanked by typical LTRs and ORF415.



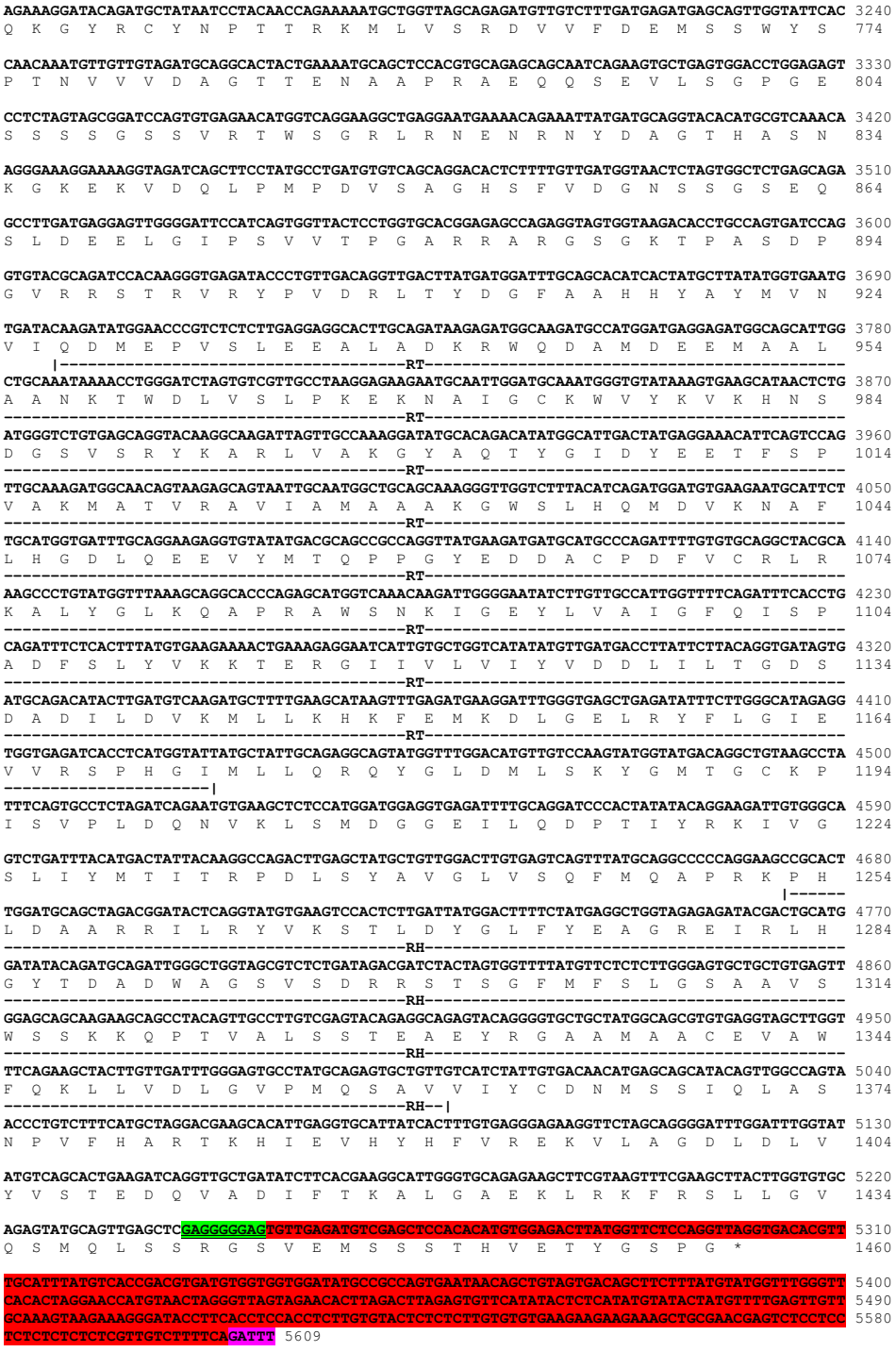


Figure 4.29: Nucleotide and amino acid sequence of *G. dicarpa* retrotransposon *GRET-A*. The identity of each domain is shown by horizontal lines above the sequences. Both 5‘ and 3‘ target site duplications (TSD) and long terminal repeats (LTR) are highlighted in magenta and red, respectively. Additional retrotransposon features, the primer binding site (PBS), the zinc finger domain (ZnF, underlined once) and the polypurine tract (PPT, underlined twice) are indicated in green color.

Ty1-*copia* reading frames found in Gd09 and Gd61 are most interrupted by stop codons and frameshifts and correspondingly the LTRs of the most fragmented Gd61 retroelement are only 89% identical (Figure 4.28 b). In contrast, only one stop codon each is interrupting the ORFs of the full length Ty1-*copia* elements in Gd61 and Gd75. Whereas the first one disrupts the continuity of the pol reading frame, which suggest degradation and inactivity of this retroelement, the latter establishes two separate reading frames of which one could encode the gag domains and second the pol domains. However, a stop codon read-through for the segregation of Gag and Pol protein expressions could be excluded since this stop codon is followed by a frameshift subsequently (Figure 4.28 b).

Most notable is the finding of a mitochondrial Ty1-*copia* retrotransposon in Gd49 next to the *cox*2 locus, which bears an uninterrupted ORF and 100% identical LTRs, indicating a functional, recently transposed retrotransposon. An alignment of all Ty1-*copia* retrotransposons identified in the *G. dicarpa* fosmids sequences discloses immediately that the non-interrupted Ty1-*copia* of Gd61 and Ty1-*copia* of Gd75 share a recent common ancestry. LTRs of both elements are very similar and phylogeny (calculated for the most similar regions, the reverse transcriptase and RNase H domains of *pol*) enhanced this hypothesis supporting the common node of these two elements with a 100% bootstrap value. These elements count for the first newly discovered Ty1-*copia* LTR-retrotransposon *GRET-C* (*Gleichenia* retrotransposon) (Figure 4.28 b).

```
CTCTCTCGTTGTCTTTTCATGGTATCAAAGCTTGGTTCGTGGTGAGGAGCTCAAGCAGCA
                  |||||||| |||  ||||
                3`-ACCATAGTCTCGGTCCAAAGCTAGGACCCTGGACA...-5`
```

Figure 4.30: Primer binding site (PBS) of *GRET-A*. Primer binding site (PBS) with complementarity to nuclear wheat tRNA-Met [Ghosh *et al.* 1982, Accession number V01383]. Long terminal repeat (LTR), primer binding site (PBS) and tRNA-Met are highlighted in red, green and yellow, respectively.

Another strong bootstrap support was given for the relationship of Ty1-*copia* of Gd49 and an undetermined sequence in the NCBI database obtained for *Pteridium esculentum* by an arbitrarily primed polymerase chain reaction (AP-PCR) [Thomson 2000] representing a Ty1-*copia* LTR-retrotransposons *GRET-A*. This finding indicates that *GRET-A* is also present in the fern *P. esculentum*. Moreover, the AP-PCR results coincidentally showed that this retrotransposon is not ubiquitously distributed among ferns, lacking in some species of the genus

Pteridium and also not to present in genera *Paesia*, *Dennstaedtia* and *Hypolepis* [Thomson 2000, referring to marker A03-1080]. Notably, the amplified region of *GRET-A* of *P. esculentum* bears a frameshift and therefore appears not to be an active element (Figure 4.28 b). In contrast, its pendant in the *G. dicarpa* mtDNA appears to be intact and possibly functionally active.

In a distance of 4887 bp containing the 4383 bp retrotransposon ORF of *GRET-A*, two identical LTR sequence stretches of 356 bp were located, both of which are terminated by the consensus sequences 5‘-TG..CA-3‘ and flanked by two 5 bp short directed repeats with the sequence: 5‘-GATTT-3‘ (Figure 4.28 a, 4.29). Such directed repeats are typical target site duplications (TSD) resulting from the retrotransposon insertion event. In the internal region between the two LTRs, two conserved sites were identified immediately bordering 3‘ downstream of the left LTR and 5‘ upstream of the right LTR, on the one hand the primer binding site (PBS) and on the other the polypurine tract site (PPT).

The PBS site is a conserved sequence, which is necessary to complement the 3‘ end sequence of tRNA-Met and to activate the transcription of mRNA. As no *trn*M sequence of *G. dicarpa* is available, we identified a 5‘-TGGTATCAAAGCTTGGTT-3‘ sequence complementary with nuclear wheat *trn*M [Ghosh *et al.* 1982, Accession number V01383] (Figure 4.30). Notably, despite its mitochondrial location, PBS still gives access to nuclear tRNA. The polypurine tract with the motive 5‘-GAGGGGGAG-3‘ is needed to switch to second strand DNA synthesis. Domains of retrotransposon ORF were identified with RPS-Blast (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and by aligning retrotransposon sequences. The order of domains (GAG-ZnF-PRO-INT-RT-RH) fully supports that the *GRET-A*-element falls into the Ty1-*copia* class of retrotransposons, contrasting with the Ty3-*gypsy* type in which the order of domains is GAG-ZnF-PRO-RT-RH-INT (Figure 4.28, 4.29).

To elucidate how often the newly identified Ty1-*copia* retrotransposon *GRET-A* is present in *G. dicarpa*, we used radioactively labeled fragments of its 3‘ region for hybridization of the sorted library filters and in a Southern blot experiment. The library filters represent the whole DNA of *G. dicarpa* randomly separated in 18.532 fragments of 33-43 kb size spotted in a 4x4 double offset pattern. With the hybridization approach, thousands of fosmids were identified carrying *GRET-A* retrotransposon homology, confirming its massive distribution (Figure 4.31 a). This high copy number was supported by the results of Southern blot analysis.

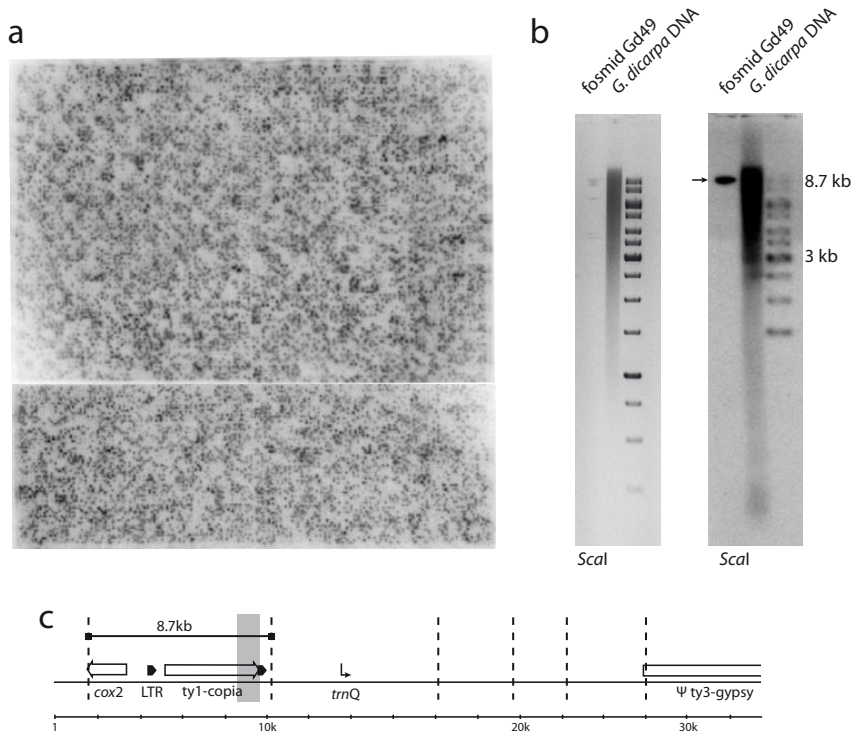


Figure 4.31: Hybridization of sorted fosmid library filters and Southern blot analysis. (**a**) Hybridization result of the sorted library. Filters are representing the whole *G. dicarpa* DNA randomly separated in 18.532 fosmids spotted in a 4x4 double offset pattern. (**b**) Southern blot analysis of fosmid Gd49 and whole *G. dicarpa* DNA cut with *Sca*I (left) and probed with an internal PCR-derived fragment of *GRET-A* (right). Arrow is indicating location of a 8.7 kb fragment containing Ty1-*copia* retrotransposon *GRET-A*. (**c**) Schematic overview of Southern blot experiment by means of Gd49 gene map. Horizontal line and arrows are representing fosmid Gd49 insert and the location of genes, respectively. The location of *Sca*I restriction sites is shown in dashed lines. PCR-derived probe of Ty1-*copia* retrotransposon *GRET-A* (highlighted with gray column) is covering internal and LTR region and was used for hybridization as presented in (a) and (b).

After hybridization, signals were observed in *G. dicarpa* DNA even under high stringency washing conditions (2x SSC with 0.1% SDS at 65°C) (Figure 4.31 c, d). These results suggest that the *GRET-A* retrotransposon is massively distributed within *G. dicarpa* species whereas apparently no detection was possible in several other fern species [Thomson 2000].

To verify arrangement of *cox*2 and retrotransposon *GRET-A* on DNA level, we further amplified *G. dicarpa* DNA with one primer binding within the *GRET-A*-LTR and a second primer annealing upstream in *cox*2 sequence. PCR resulted in a smear (not shown) possibly due to thousands of LTR primer binding options

Figure 4.32: Sequence alignment of Ty1-*copia* retrotransposon *GRET-A* copies in different genomic locations. Sequences verify transposition of Ty1-*copia* retrotransposon *GRET-A* in different locations. All fragments were 3‘ *GRET-A* orientated but arose from different locations as their 5‘ upstream regions show. Background shading colors indicate sequence similarity of *GRET-A* LTR, whereas no simililarity is given for upstream regions. Display and shading of sequence alignment was done with GeneDoc (http://www.nrbsc.org/gfx/genedoc/)

indicated by hybridization results (Figure 4.31). We eluted several DNA fragments of the smear and cloned them into pGEM-T Easy vectors (Promega). Sequencing nine products revealed that all fragments were 3‘ *GRET-A* orientated but arose from different locations as their 5‘ upstream regions verified (Figure 4.32). The sequences clarified that a Ty1-*copia* retrotransposon *GRET-A* transposes in different locations and verified results represented in figure 4.31.

5 Abschließende Diskussion

Da die Veröffentlichungen der vorangegangenen Kapitel jeweils eine Diskussion enthalten, stehen in der abschließenden Diskussion die neusten Ergebnisse der mtDNA-Sequenzierung von *G. dicarpa* im Mittelpunkt. Es werden ein Vergleich der rekombinanten mitochondrialen Strukturen basaler Tracheophyten und die hohe Anzahl mobiler Elemente in der mtDNA des Monilophyten im Kontext aktueller Forschung dargestellt.

5.1 Neuorganisation der mtDNA durch Rekombinationen

Über die in dieser Arbeit erstmals sequenzierten mtDNAs der Lycophyten und Monilophyten und die damit verbundenen Strukturanalysen kann die Entstehung der rekombinanten Chondromstruktur auf einen gemeinsamen Vorfahren der Tracheophyten datiert werden. Die vielen identifizierten Sequenzwiederholungen (*repeats*) in *S. moellendorffii* sind – ähnlich wie in Samenpflanzen – offenbar der Schlüssel für die Entstehung rekombinanter Moleküle und unterscheiden die untersuchten mtDNAs von denen der basalsten Landpflanzengruppen, der Bryophyten, in denen homogene Moleküle mit weitgehend syntenisch angeordneten Genarrangements nachgewiesen wurden.

Die Chondromstruktur rekombinanter mtDNA hat sich jedoch in den verschiedenen Entwicklungslinien der Tracheophyten jeweils individuell ausgeprägt. Innerhalb der Lycophyten entstand in *I. engelmannii* eine extrem kompakte Chondromstruktur, die mit sehr kleinen intergenischen Bereichen und Introns einen einzigartigen evolutiven Trend in der Landpflanzenevolution darstellt (siehe Abschnitt 4.1.1). Vollkommen verschieden gestaltet sich dagegen die lose Struktur der mtDNA von *S. moellendorffii* mit zehn langen *repeats* als Basis für homologe Rekombinationen (siehe Abschnitt 4.3.1). Beide mtDNAs sind sehr komplex aufgebaut,

wobei *I. engelmannii* mit einer kompakten Struktur eher denen basaler Landpflanzen entspricht, während die Chondromstruktur von *S. moellendorffii* an die rekombinanten mtDNAs der Angiospermen erinnert.

Die gesamte mtDNA von *I. engelmannii* ist über 24 asymmetrische Rearrangements, sogenannte Rekombinationspunkte, strukturiert. Diese könnten einen ähnlichen Ursprung haben wie manche der Rekombinationen einer *Beta vulgaris*-Linie, die ebenfalls nicht an *repeats* assoziiert sind. Man vermutet in diesem Fall, dass die Rekombinationen nicht mehr aktiv sind und die *repeats* nachträglich eliminiert wurden [Satoh *et al.* 2006]. Dies wäre im kompakten Chondrom von *I. engelmannii* jedoch keine Erklärung für jene Rekombinationspunkte, durch die chimäre Gene (*atp*8/*cox*1) resultieren. Wahrscheinlicher ist es, dass die Rekombinationspunkte jeweils Anfang und Ende von langen *repeats* darstellen, über die auch in *I. engelmannii* homologe Rekombinationen stattfinden könnten. In diesem Fall könnte ein Gemisch ineinander liegender *repeats* für viele weitere Arrangements verantwortlich sein, wie es auch für manche *repeats* im *S. moellendorffii*-Chondrom vermutet wird (siehe Abschnitt 4.3.1).

In *G. dicarpa* wurden drei Arten von Sequenzwiederholungen ermittelt. Geprägt wird die Chondromstruktur des Monilophyten durch lange und kurze *repeats*, die sich wiederum auf weit entfernten Positionen der mtDNA befinden können oder sich als *direct repeats* in naher Umgebung finden lassen. Oft liegen sie auch unmittelbar aneinander als sogenannte *tandem repeats* verbunden vor. Diese vielen *repeats* führen zu einer bisher ungesehenen komplexen Struktur mitochondrialer DNA. Dabei gleichen die unterschiedlichen Arten von Sequenzwiederholungen den Strukturen, wie sie auch in den Angiospermen nachgewiesen wurden. An langen *repeats* finden dort häufig homologe Rekombinationen statt, die zu neuen intra- und intermolekularen Rearrangements führen und oft in einem *mastercircle* zusammengefasst dargestellt werden [Kazuyoshi und Tomohiko 2010]. Die über kurze *repeats* (>50 bp, <1 kb) vermittelten Rekombinationen sind dagegen selten und führen zu wenigen subgenomischen Strukturen, sogenannte *sublimons* [Small *et al.* 1987], deren Konzentration aber – reguliert von Kern-kodierten Faktoren – stark ansteigen kann [Arrieta-Montiel und Mackenzie 2011, Maréchal und Brisson 2010]. Noch kürzere *repeats* (<25 bp) wurden bei eingehenderen Untersuchungen von intergenischen Regionen in Angiospermen identifiziert [Satoh *et al.* 2006] und sind entweder an nicht-homologen Rekombinationen (*nonhomologous end joining*, NEHJ) beteiligt [Arrieta-Montiel und Mackenzie 2011] oder stellen Rudimente

vergangener Rearrangements durch homologe Rekombinationen dar [Kazuyoshi und Tomohiko 2010].

Syntenische Gen-*Cluster* liegen konserviert in den homogenen Chondromen von Grünalgen und Bryophyten vor [Terasawa *et al.* 2007] und sind in den rearrangierten Angiospermen-mtDNAs selten. Interessanterweise blieben jedoch trotz der erheblichen Neuorganisation der Farn-mtDNA in zwei Fällen syntenische Gen-*Cluster* (*rpl*2-*rps*19-*rps*3 und *rpl*6-*rps*13-*rps*11) erhalten. Diese *Cluster* findet man auch in den homogenen mtDNAs der Leber- und Laubmoose. Aufgrund des erhöhten Genverlusts in Hornmoosen und Lycophyten liegen dort viele Gene der *Cluster*, wenn sie nicht vollständig verloren gingen, nur noch als degradierte Pseudogene vor.

Neben den Genen der beiden *Cluster* wurden mit *rps*2, *rpl*5 und *sdh*3 noch weitere drei Gene in der mtDNA von *G. dicarpa* identifiziert, die oft von den mtDNAs anderer Taxa verloren gegangen sind (siehe Abschnitt 2.4). Es kann aufgrund der bisher unvollständigen mtDNA-Sequenzen noch keine eindeutige Aussage über den Genbestand der mtDNA von *G. dicarpa* getroffen werden. Die bisherigen Funde deuten jedoch darauf hin, dass es in *G. dicarpa* keinen erhöhten Genverlust wie von den mtDNAs der Lycophyten (siehe Abschnitt 4.1.1 und 4.3.1), Hornmoose [Li *et al.* 2009, Xue *et al.* 2010] und Caryophyllales [Kubo *et al.* 2000, Sloan *et al.* 2010a] gegeben hat.

5.2 Rekombinationen reguliert von nukleären Faktoren

Als besonders interessant gestaltet sich die Identifikation und Betrachtung von Faktoren, die mitochondriale Rekombinationen regulieren. Im Kerngenom von *A. thaliana* wies man bereits den Genen *msh*1, *rec*A und *OSB1* eine Beteiligung an mitochondrialen Rekombinationen nach [Abdelnoor *et al.* 2003, Arrieta-Montiel *et al.* 2009, Shedge *et al.* 2007, Zaegel *et al.* 2006]. Im Laubmoosgenom von *P. patens* wurde ein *rec*A-Homolog (*PpRec*A1) mit mitochondrialem Funktionsort identifiziert [Odahara *et al.* 2007].

Homologe der Gene *msh*1 und *rec*A gibt es auch in Eubakterien. Für *rec*A wurde nachgewiesen, dass es vermutlich aus der frühen mtDNA in das Kerngenom transferiert wurde [Lin *et al.* 2006]. Im Prokaryoten vermitteln sie DNA-Austausch

nach Konjugationen und sind an Reparatur- und DNA-Replikationsprozessen über homologe Sequenzbereiche beteiligt [Kowalczykowski *et al.* 1994].

Wird die Funktion der homologen Gene in Pflanzen unterbunden, führt dies in der mtDNA zu einem Verlust der Kontrolle über Rekombinationen und einer Erhöhung von irregulären Rearrangements [Maréchal und Brisson 2010]. *PpRec*A1 in *P. patens* wurde eine Beteiligung an Reparaturmechanismen von beschädigter DNA [Odahara *et al.* 2007] sowie der Organisation der mtDNA-Struktur über kurze *repeats* (<100 bp) nachgewiesen [Odahara *et al.* 2009]. *P. patens* gewinnt zunehmend als Modellpflanze an Bedeutung, da in der dominanten Haplophase des Laubmooses über homologe Rekombination die Integration von DNA in das Kerngenom mit großer Effizienz abläuft [Schaefer und Zryd 1997].

Weiterführende Studien darüber, wie diese rekombinationsregulierenden Faktoren im Kerngenom der Lycophyten ausgeprägt sind, werden die gewonnen Erkenntnisse über die individuelle Beschaffenheit der mtDNA-Strukturen erweitern. Zukünftige Genomanalysen (wie die angekündigte vollständige Genomanalyse von *S. moellendorffii* [Banks 2009]) durch *next-generation*-Sequenzierungen werden dabei eine Grundlage für diese weiterführenden Analysen bilden.

5.3 Integration "promisker" DNA in Tracheophyten

Trotz der Tendenz zu einer komprimierten mtDNA wurden in der Sequenz von *I. engelmannii* Regionen "promisker" DNA des Plastiden- oder Kerngenoms identifiziert (siehe Abschnitt 4.1.1).

Die bisher identifizierten "promisken" Kerngenom-Fragmente stammen hauptsächlich von mobilen Retroelementen ab. Solch eine Integration von Retrotransposon-Fragmenten wurde erstmals in über 5% des mitochondrialen Genoms von *A. thaliana* nachgewiesen [Knoop *et al.* 1996]. Mit der wachsenden Anzahl an verfügbaren Kerngenomsequenzen, die zu Vergleichsanalysen genutzt werden können, wird eine Identifikation "promisker" Kernregionen erleichtert. Mittlerweile wurden deshalb in fast allen untersuchten mtDNAs von Angiospermen Fragmente transponierender Elemente entdeckt, die mit bis zu 13% Bestandteil der mtDNA von *O. sativa* sind [Notsu *et al.* 2002].

In *I. engelmannii* wurden interessanterweise keine Fragmente von Retroelementen, sondern ausschließlich von nicht *per se* mobilen Kerngenen gefunden. Damit wurden neben den Fragmenten einer Mandelonitril-Lyase und des Gens *nonexineformation1* (*nef*1), beide im Chondrom von *B. vulgaris* identifiziert [Kubo *et al.* 2000, Satoh *et al.* 2006], in *I. engelmannii* zwei weitere "promiske" DNA-Fragmente eines *Auxin Response* Transkriptionsfaktors und eines Phytochrom-Gens festgestellt (siehe Abschnitt 4.1.1). Später wurden zudem Fragmente in beiden sequenzierten mtDNAs der Cucurbitaceae *C. lanatus* und *C. citrullus* entdeckt, eine Lektin-Protein-Kinase und ebenfalls Teile einer Mandelonitril-Lyase [Alverson *et al.* 2010].

Da in allen sequenzierten mtDNAs der Tracheophyten "promiske" DNA nachgewiesen wurde, jedoch keine in den Bryophyten-Gruppen, kann die Fähigkeit zu einem solchen Einbau zugleich mit dem Aufkommen massiv rearrangierter mtDNA in einem gemeinsamen Vorfahren der Tracheophyten entstanden sein. Die häufige, teilweise unkontrollierte Neuverknüpfung von DNA-Strängen fördert möglicherweise eine Integration fremder DNA-Fragmente. In diesem Zusammenhang wurde bei einer Vergleichsanalyse der mtDNA von *B. vulgaris* mit der einer CMS-Linie ein massiv rearrangiertes Chondrom und zugleich ein erhöhtes Aufkommen von integrierten "promisken" Sequenzen beobachtet [Satoh *et al.* 2006].

Innerhalb der Lycophyten zeigt *S. moellendorffii* als Schwestertaxon von *I. engelmannii* interessanterweise einen anderen Trend: Dort wurden trotz großer nichtkodierender intergenischer Regionen keine erkennbaren fremden DNA-Fragmente integriert. In Anbetracht einer Verbindung von rearrangierter mtDNA und der Integration fremder DNA, kann für *S. moellendorffii* als Schwestergruppe von *I. engelmannii* angenommen werden, dass andere nötige Voraussetzungen für eine rekombinationsbegleitete Integration fremder DNA möglicherweise fehlen oder dass die Mitochondrien den Zugang fremder DNA-Fragmente verwehren, sodass es gar nicht erst zu einem physischen Kontakt mit der mtDNA kommt. Eine Möglichkeit des DNA Imports in die Mitochondrien wurde anhand des Modellsystems der Kartoffel nachgewiesen [Koulintchenko *et al.* 2003] und das seltene Vorkommen von "promisker" DNA in Plastiden [Keeling 2009] wurde mit dem Fehlen eines solchen Systems in der Chloroplastenmembran begründet [Shapiro 2010].

Auch in der mtDNA des Monilophyten *G. dicarpa* wurden "promiske" Sequenzen identifiziert. Dort wurden in einer bisher ungesehenen Anzahl neben den Fragmenten transponierender Elemente und nicht *per se* mobiler Kerngene noch viele

weitere fremde, mobile und invasive Elemente identifiziert: Ein transponiertes Intron, integrierte DNA- und RNA-Polymerasen von linearen mitochondrialen Plasmiden und von Mitoviren, darüber hinaus ein DNA-Transposon und vollständige Ty3-*gypsy*- und Ty1-*copia*-Retroelemente. Diese Elemente charakterisieren die Monilophyten-mtDNA als besonders empfänglich für die Integration verschiedenster Typen mobiler und promisker Sequenzen, wie es bislang in keinem anderen Chondrom gesehen wurde.

5.4 Introntransposition innerhalb der Monilophyten

Das erstmalig in *G. dicarpa* identifizierte Intron rps1i24g2 ist vermutlich eine Kopie des Introns rpl2i917g2 und damit das Ergebnis einer Introntransposition. Eine intragenische Intronduplikation wurde zuvor innerhalb des *nad*5-Gens in der Lycophyten mtDNA von *Huperzia selago* beschrieben [Vangerow *et al.* 1998]. Aufgrund einer späteren individuellen Größenzunahme der variablen Domäne IV des ursprünglichen Introns unterscheiden sich beide Introns jedoch um beinahe 2 kb. Ähnlichkeitsstudien von Gruppe II Introns von *M. polymorpha* verdeutlichten zudem vier Intronduplikation in dem Lebermoos nach einer Trennung dieser Pflanzenlinie von den anderen Landpflanzen [Ohyama und Takemura 2008].

Die hohe Sequenzübereinstimmung der Introns rpl2i917g2 und rps1i24g2 von *G. dicarpa* deutet auf ein sehr rezentes Duplikationsereignis innerhalb des Monilophyten. Die beschriebenen mobilen Funktionen der Introns (siehe Abschnitt 2.6) könnten zusammen mit einem reversen Spleißmechanismus für die Duplikation verantwortlich sein [Augustin *et al.* 1990]. Da das duplizierte *G. dicarpa*-Intron allerdings keine ORFs trägt, sind diese entweder später verloren gegangen oder es müssen *trans*-Faktoren für die Transposition verantwortlich gewesen sein.

Erste phylogenetische Analysen anderer Monilophytensequenzen bestätigen, dass diese Intronduplikation eine charakteristische Eigenschaft der Monilophyten darstellt. An welcher Position der Monilophytenphylogenie dieses Ereignis stattgefunden hat, gilt es, über weitere Taxa noch zu bestimmen.

5.5 Mitoviren-RNA-Polymerase

Neben den Regionen "promisker" DNA und der Intronduplikation wurden über die Identifikation auffallend großer ORFs in der mtDNA von *G. dicarpa* weitere integrierte mobile Elemente erkannt. Manche dieser ORFs kodieren für RNA-Polymerasen, die möglicherweise ihren Ursprung in RNA-Mitoviren von pflanzenparasitären Pilzen haben. Bekannt sind diese integrierten Fragmente bereits aus *A. thaliana* und *Vicia faba* [Marienfeld *et al.* 1997], wobei der Leserahmen in *V. faba* gleich dem von *G. dicarpa* von Stop-Codons unterbrochen ist. Zwei vollständig intakte Leserahmen fand man hingegen in *A. thaliana*.

Wegen der großen Ähnlichkeit der mitochondrialen und der pilzlichen Mitovirensequenzen wird ein horizontaler Sequenztransfer der viralen RNA zwischen Pilz und Pflanze und eine anschließende Integration über einen reversen Transkriptionsmechanismus angenommen [Marienfeld *et al.* 1999]. Dies würde neben dem oben angesprochenen reversen Spleißen und dem u.a. in *I. engelmannii* festgestellten *retro-processing* (siehe Abschnitt 4.2.1) eine weitere Sequenzintegration über RNA-Intermediate in mtDNA darstellen.

5.6 RNA- und DNA-Polymerasen linearer mitochondrialer Plasmide

Weitere große ORFs der mtDNA von *G. dicarpa* kodieren für DNA- und RNA-Polymerasen ähnlich der Gene, die lineare mitochondriale Plasmide tragen. Lineare Plasmide wurden erstmals in den Mitochondrien von Pilzen und Pflanzen [Meinhardt *et al.* 1990], später auch in dem Protisten *Physarum polycephalum* nachgewiesen [Sakurai *et al.* 2004]. Charakterisiert sind die linearen mitochondrialen Plasmide durch endständige *inverted repeats* und ORFs, wovon einer für eine Familie-B-DNA-Polymerase und ein anderer für eine Phagentyp-RNA-Polymerase kodiert [Handa *et al.* 2002].

Die Plasmidmoleküle können über Rekombinationen in das Chondrom integriert werden [Sakurai *et al.* 2004]. Oft beweisen rudimentäre Fragmente der Polymerasen einen vorangegangenen Einbau von Plasmiden [Übersicht in Gray *et al.* 2004]. Solche Überreste wurden selbst in der mtDNA von *M. polymorpha* nachgewiesen [Weber *et al.* 1995], dessen Mitochondrien keine linearen Plasmide enthalten [McDermott *et al.* 2008] und diese möglicherweise nach einem Einbau verloren

haben. In anderen mtDNAs wie z.B. *Lolium perenne* und *Daucus carota* sind die Leserahmen noch vollständig enthalten und werden sogar transkribiert [McDermott *et al.* 2008, Robison und Wolyn 2004].

Es ist zu vermuten, dass lineare mitochondriale Plasmide von viralen Vorläufern ähnlich den T3- oder T7-Phagen abstammen, von denen eventuell ebenso die Replikations- und Transkriptionsmaschinerie der Mitochondrien erworben wurde [Shutt und Gray 2006].

Über eine Sequenzierung mitochondrialer Fosmide von *G. dicarpa* wurde erstmals die Integration von Plastiden-DNA in ein Monilophyten-Chondrom nachgewiesen. Die Leserahmen werden durch Stop-Codons unterbrochen, die einen Verlust der Genfunktion signalisieren – auch eine Transkription der Gene konnte nicht nachgewiesen werden.

Eine Vermutung, die RNA-Polymerasen der linearen Plasmide (oder der mitochondrial integrierten Gene) könnten für mitochondriale Transkription herangezogen werden, wird durch eine klare phylogenetische Trennung der RNA-Polymerasesequenzen der Plasmide von den kernkodierten Gensequenzen widerlegt [Handa *et al.* 2002]. In *Vitis vinifera* wurde darüber hinaus trotz eines intakten Leserahmens einer mitochondrial kodierten RNA-Polymerase ein koexistierendes kernkodiertes Gen nachgewiesen, von dem eine mitochondriale Aktivität anzunehmen ist [Goremykin *et al.* 2009, Velasco *et al.* 2007]. Daraus resultierend kann angenommen werden, dass auch die Polymerasen der mitochondrialen Replikations- und Transkriptionsmaschinerie von *G. dicarpa* von Kerngenen kodiert werden.

5.7 Integrierte transponierende Elemente

Neben den "promisken" Fragmenten von nicht *per se* mobilen Genen des Chloroplasten und des Kerns wurden in der mtDNA von *G. dicarpa* zudem große Regionen von DNA-Transposons, Ty3-*gypsy*- und Ty1-*copia*-Retrotransposons entdeckt.

Fünf der acht in die mtDNA von *G. dicarpa* integrierten mobilen Elemente sind in voller Länge mit einer Größe über 4000 bp enthalten und werden typischerweise von langen terminalen Sequenzwiederholungen (*long terminal repeats* LTRs) flankiert. Die meisten Leserahmen der mobilen Elemente sind jedoch durch wenige Stop-Codons und Indels (Insertionen und Deletionen) unterbrochen. Die Retroelemente haben damit ihre Funktion und Mobilität verloren. Eine bisher unentdeckte

Besonderheit ist der Einbau eines vollständigen und damit möglicherweise funktionsfähigen Ty1-*copia*-Retrotransposons.

Keine der bisher identifizierten "promisken" Retroelemente in anderen mtDNAs sind ähnlich vollständig. Die bisher größte zusammenhängende "promiske" Region, eine 1555 bp große 5'-Hälfte eines Ty1-*copia*-Retrotransposons, wurde im Reis Chondrom entdeckt [Notsu *et al.* 2002]. Dieses Fragment ist ebenfalls weder durch Stop-Codons noch Indels degradiert, was auf ein rezentes Transpositionsereignis und eine nachfolgende Fragmentierung durch Rekombination deutet.

Transponierende Elemente gestalten einen erheblichen Sequenzanteil des Kerngenoms vieler Eukaryoten [Kidwell 2002]. Verschiedene transponierende Elemente wurden in allen untersuchten Landpflanzengenomen und dem der Grünalge *Volvox carteri* nachgewiesen und gelten als ubiquitär im Pflanzenreich verteilt [Bennetzen 1996, Lindauer *et al.* 1993].

Die Genomgrößen der Eukaryoten korrelieren mit der enthaltenen Kopienanzahl der transponierenden Elemente [Kidwell 2002]. Im vergleichsweise kleinen Hefegenom von *Saccharomyces cerevisiae* mit einer Größe von 12 Mb wurden insgesamt nur 331 Retrotransposons gezählt [Kim *et al.* 1998]. Das Mais-Kerngenom mit einer Größe von ungefähr 2500 Mb enthält dagegen bis zu jeweils 30.000 Kopien verschiedener Retrotransposon-Familien [SanMiguel *et al.* 1996]. Hauptsächlich befinden sich diese mobilen Elemente in den repetitiven Genomregionen, die mit bis 80% den größten Anteil des Kerngenoms ausmachen können [SanMiguel und Bennetzen 1998].

Retrotransposons vervielfältigen sich über einen "*copy and paste*"-Mechanismus und liegen zu einer viel höheren Kopienanzahl als DNA-Transposons vor, die über einen "*cut and paste*"-Mechanismus oft ihre ursprüngliche Position im Genom verlassen und an einer anderer Stelle reintegrieren. Dies ist möglicherweise auch ein Grund dafür, dass die "promisken" Sequenzanteile in der mtDNA dominierend von Retrotransposons abstammen [Knoop *et al.* 1996, Notsu *et al.* 2002] und auch in den Fosmidsequenzen der mtDNA von *G. dicarpa* nur ein DNA-Transposon verglichen zu acht Retrotransposons gefunden wurde.

Durch ihre vielfältigen Eigenschaften (Transposition, Restriktion, Insertion und Einfluss auf Translokationen sowie Rekombinationen) können transponierende Elemente große Auswirkungen auf die Genfunktionen und Genomstruktur ihres Wirtsgenoms haben [Bennetzen 2000]. Die Aktivität und eine Verbreitung der

funktionalen Retrotransposons wird deshalb im Kerngenom durch Methylierungen ständig unterdrückt [Yoder *et al.* 1997]. Nur drei pflanzlichen Retrotransposons (*Wis-2*, *Bs1* und *Tnt1*) wurde bisher unter speziellen Bedingungen eine Aktivität nachgewiesen [Grandbastien 1992].

Es wird angenommen, dass Retrotransposons erst in bestimmten Situationen, durch einen "genomischen Schock", aktiviert werden, der dann zu einer massiven Neuordnung des Genoms führt [McClintock 1984]. Polyploidie, die zu hunderten Chromosomen in manchen Farnspezies geführt hat [Khandelwan 1990], könnte nach Wendel [2000] einen solchen Schock verursacht haben, der die Stilllegung der Retrotransposons wieder aufhebt. Infolge der erhöhten Retrotransposon-Kopien könnte es zusammen mit der Polyploidie zu der massiven Vergrößerung der Farngenome gekommen sein [Leitch *et al.* 2005]. Die nachgewiesene große Verbreitung eines Retrotransposons (*GRET-A*) im Genom von *G. dicarpa* unterstützt eine solche Annahme. Der Einbau mobiler Elemente, hauptsächlich Ty1-*copia*- und Ty3-*gypsy*-Retrotransposons, in das mitochondriale Genom ist damit ein kollaterales Ereignis der bisher drastischsten Verbreitungen von Retroelementen in Landpflanzen.

5.8 Schlussbetrachtung

"Mitochondrial Genomes: anything goes"

Mit diesen Worten überschrieben Burger *et al.* [2003] ihre Publikation, um damit auf die vielen verschiedenen Besonderheiten mitochondrialer DNA aufmerksam zu machen. Alle darauf folgenden Sequenzierungen pflanzlicher mitochondrialer DNA bestätigten diese Aussage, denn jede neue mitochondriale Sequenz eröffnete jeweils bis dahin ungesehene Besonderheiten dieser Moleküle.

Die in der vorliegenden Arbeit vorgestellten mtDNA-Analysen von basalen Tracheophyten verdeutlichen erneut, dass bezüglich mitochondrialer DNA alles möglich sein kann: Durch die bisher umfassendste Analyse einer Monilophyten mtDNA von *Gleichenia dicarpa* und die zwei vollständigen Sequenzierungen der Lycophyten *Isoetes engelmannii* und *Selaginella moellendorffii* wurde eine Lücke der Landpflanzenphylogenie geschlossen. Es wurden dadurch neue Erkenntnisse über die mitochondrialen Genome gewonnen und neue individuelle Evolutionswege zu extrem komplexen Strukturen erkannt. Diese Strukturen führten zu weiteren mitochondrialen Besonderheiten, wie dem Einbau "promisker" Fragmente aus

Zellkern und Chloroplasten bis zu vollständigen Retrotransposons und zu einem in diesen Studien erstmalig gesehenen *trans*-spleißenden Gruppe I Intron.

Daneben zeichnen sich die untersuchten Taxa durch weitere post-transkriptionelle Eigenheiten aus, z.B. den Höchstwerten an *RNA Editing* von proteinkodierenden Transkripten und tRNAs. Ebenso wie in ihrer Struktur unterscheiden sich beide Lycophyten auch in diesem Zusammenhang. Während in *S. moellendorffii* ausschließlich Cytidin-zu-Uridin-Konversionen zu finden sind, handelt es sich bei 20% der *Editing*-Ereignisse in *I. engelmannii* um Austausche in Gegenrichtung von Uridin zu Cytidin. Die geplante mtDNA-Sequenzierung von *Huperzia* [Prof. Y.-L. Qiu, Ann Arbor, pers. Mitt.] aus der Ordnung der Lycopodiales, der Schwestergruppe der Isoetales und Selaginellales, wird interessante zusätzliche Information über die mitochondriale Evolution der Lycophyten liefern.

Durch immer kostengünstigere Verfahren werden zukünftig *next-generation* Sequenzierungen zu einer hohen Zunahme an Sequenzdaten führen. Komplettsequenzierungen pflanzlicher Genome liefern dabei weitere Einblicke in kernkodierte mitochondriale Faktoren. Zudem wird die Anzahl vollständig sequenzierter mtDNAs zunehmen, denn oft erhält man bereits eine vollständige mtDNA (und die cpDNA des Plastiden) als eine Art Nebenprodukt bei Kerngenom-Sequenzierungen. Doch bei der Sequenzierung komplexer mitochondrialer Strukturen, wie sie in dieser Arbeit vor allem in der Farn-mtDNA gesehen wurde, muss der Vorteil von *next-generation* Sequenzierungen erwogen werden. Trotz einer extrem hohen möglichen Sequenzabdeckung (*coverage*) ist beispielsweise ohne eine Fosmidbank als Orientierung ein korrektes Sequenzassembly unter Umständen gar nicht möglich.

Diese komplexen mitochondrialen Strukturen sind ein Grund dafür, warum bisher im Vergleich zu tierischen mtDNAs deutlich weniger pflanzliche mtDNAs vollständig sequenziert wurden. Sie begründen aber auch die Fülle an mitochondrialen Besonderheiten und die immer wieder neu erlangten Erkenntnisse zu bisher unbekannten mitochondrialen Evolutionswegen. Zukünftige Sequenzierungen werden weitere Geheimnisse mitochondrialer Genome entschlüsseln und ganz sicher werden die Taxa der basalen Tracheophytengruppen dafür eine äußerst interessante Grundlage bieten.

6 Literaturverzeichnis

Abdelnoor, R., Yule, R., Elo, A., Christensen, A., Meyer-Gauen, G. und Mackenzie, S. (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to MutS. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10):5968–5973.

Adams, K., Qiu, Y., Stoutemyer, M. und Palmer, J. (2002) Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99(15):9905–9912.

Allen, J., Jackson, A., Rigden, D., Willis, A., Ferguson, S. und Ginger, M. (2008) Order within a mosaic distribution of mitochondrial c-type cytochrome biogenesis systems? *FEBS Journal*, 275(10):2385–2402.

Allen, J. O., Fauron, C. M., Minx, P., Roark, L., Oddiraju, S., Lin, G. N., Meyer, L., Sun, H., Kim, K., Wang, C., Du, F., Xu, D., Gibson, M., Cifrese, J., Clifton, S. W. und Newton, K. J. (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics*, 177(2):1173–1192.

Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. und Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–3402.

Alverson, A. J., Wei, X., Rice, D. W., Stern, D. B., Barry, K. und Palmer, J. D. (2010) Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). *Molecular Biology and Evolution*, 27(6):1436 –1448.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. und Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290(5806):457–465.

Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C. M., Podowski, R. M., Naslund, A. K., Eriksson, A., Winkler, H. H. und Kurland, C. G. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature*, 396(6707):133–140.

Antes, T., Costandy, H., Mahendran, R., Spottswood, M. und Miller, D. (1998) Insertional editing of mitochondrial tRNAs of Physarum polycephalum and Didymium nigripes. *Molecular and Cellular Biology*, 18(12):7521–7527.

Arrieta-Montiel, M., Lyznik, A., Woloszynska, M., Janska, H., Tohme, J. und Mackenzie, S. (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. *Genetics*, 158(2):851–864.

Arrieta-Montiel, M. P. und Mackenzie, S. A. (2011) Plant Mitochondrial Genomes and Recombination. In Kempken, F., editor, *Plant Mitochondria*, volume 1 of *Advances in Plant Biology*, Seiten 65–82. Springer New York.

Arrieta-Montiel, M. P., Shedge, V., Davila, J., Christensen, A. C. und Mackenzie, S. A. (2009) Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics*, 183(4):1261–1268.

Augustin, S., Muller, M. und Schweyen, R. (1990) Reverse self-splicing of group II intron RNAs in vitro. *Nature*, 343(6256):383–386.

Banks, J. A. (2009) Selaginella and 400 million years of separation. *Annual Review of Plant Biology*, 60(1):223–238.

Beagley, C., Okada, N. und Wolstenholme, D. (1996) Two mitochondrial group I introns in a metazoan, the sea anemone Metridium senile: one intron contains genes for subunits 1 and 3 of NADH dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 93(11):5619–5623.

Bendich, A. (1993) Reaching for the ring: the study of mitochondrial genome structure. *Current Genetics*, 24(4):279–290.

Bendich, A. J. (2007) The size and form of chromosomes are constant in the nucleus, but highly variable in bacteria, mitochondria and chloroplasts. *BioEssays*, 29(5):474–483.

Bennetzen, J. (1996) The contributions of retroelements to plant genome organization, function and evolution. *Trends in Microbiology*, 4(9):347–353.

Bennetzen, J. (2000) Transposable element contributions to plant gene and genome evolution. *Plant Molecular Biology*, 42(1):251–269.

Bentolila, S., Chateigner-Boutin, A. und Hanson, M. (2005) Ecotype allelic variation in C-to-U editing extent of a mitochondrial transcript identifies RNA-editing quantitative trait loci in Arabidopsis. *Plant Physiology*, 139(4):2006–2016.

Bentolila, S., Elliott, L. E. und Hanson, M. R. (2008) Genetic architecture of mitochondrial editing in Arabidopsis thaliana. *Genetics*, 178(3):1693–1708.

Bergthorsson, U., Adams, K., Thomason, B. und Palmer, J. (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, 424(6945):197–201.

Binder, S., Marchfelder, A. und Brennicke, A. (1994) RNA editing of tRNAPhe and tRNACys in mitochondria of Oenothera berteriana is initiated in precursor molecules. *Molecular and General Genetics*, 244(1):67–74.

Binder, S., Marchfelder, A., Brennicke, A. und Wissinger, B. (1992) RNA editing in trans-splicing intron sequences of nad2 mRNAs in Oenothera mitochondria. *Journal of Biological Chemistry*, 267(11):7615 –7623.

Blanc, V., Litvak, S. und Araya, A. (1995) RNA editing in wheat mitochondria proceeds by a deamination mechanism. *FEBS Letters*, 373(1):56–60.

Bock, R., Hagemann, R., Kössel, H. und Kudla, J. (1993) Tissue-and stage-specific modulation of RNA editing of the psbF and psbL transcript from spinach plastids-a new regulatory mechanism? *Molecular and General Genetics*, 240(2):238–244.

Bonen, L. (2008) Cis- and trans-splicing of group II introns in plant mitochondria. *Mitochondrion*, 8(1):26–34.

Bonen, L. (2011) RNA Splicing in Plant Mitochondria. In Kempken, F., editor, *Plant Mitochondria*, volume 1 of *Advances in Plant Biology*, Seiten 131–155. Springer New York.

Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. und Schuster, W. (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Letters*, 325(1-2):140–145.

Bresinsky, A., Jarosch, B. und Körner, C. (2008) *Strasburger-Lehrbuch der Botanik*. Spektrum, Akademischer Verlag.

Bullerwell, C., Schnare, M. und Gray, M. (2003) Discovery and characterization of Acanthamoeba castellanii mitochondrial 5S rRNA. *RNA*, 9(3):287–292.

Burger, G., Gray, M. W. und Lang, B. F. (2003) Mitochondrial genomes: anything goes. *Trends in Genetics*, 19(12):709–716.

Carrillo, C., Chapdelaine, Y. und Bonen, L. (2001) Variation in sequence and RNA editing within core domains of mitochondrial group II introns among plants. *Molecular and General Genetics*, 264(5):595–603.

Castandet, B., Choury, D., Bégu, D., Jordana, X. und Araya, A. (2010) Intron RNA editing is essential for splicing in plant mitochondria. *Nucleic Acids Research*. DOI: 10.1093/nar/gkq591.

Cech, T., Damberger, S. und Gutell, R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nature Structural and Molecular Biology*, 1(5):273–280.

Chapdelaine, Y. und Bonen, L. (1991) The wheat mitochondrial gene for subunit I of the NADH dehydrogenase complex: A trans-splicing model for this gene-in-pieces. *Cell*, 65(3):465–472.

Chateigner-Boutin, A. und Small, I. (2010) Plant RNA editing. *RNA Biology*, 7(2):213–219.

Chaw, S., Shih, A., Wang, D., Wu, Y., Liu, S. und Chou, T. (2008) The mitochondrial genome of the gymnosperm Cycas taitungensis contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Molecular Biology and Evolution*, 25(3):603–615.

Cho, Y. und Palmer, J. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Molecular Biology and Evolution*, 16(9):1155–1165.

Choquet, Y., Goldschmidt-Clermont, M., Girard-Bascou, J., Kück, U., Bennoun, P. und Rochaix, J. (1988) Mutant phenotypes support a trans-splicing mechanism for the expression of the tripartite psaA gene in the C. reinhardtii chloroplast. *Cell*, 52(6):903–913.

Clifton, S., Minx, P., Fauron, C., Gibson, M., Allen, J., Sun, H., Thompson, M., Barbazuk, W., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R. und Newton,

K. (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology*, 136(3):3486–3503.

Covello, P. und Gray, M. (1993) On the evolution of RNA editing. *Trends in Genetics*, 9(8):265–268.

Covello, P. S. und Gray, M. W. (1989) RNA editing in plant mitochondria. *Nature*, 341(6243):662–666.

Dombrovska, O. und Qiu, Y. (2004) Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and molecular evolutionary implications. *Molecular Phylogenetics and Evolution*, 32(1):246–263.

Duchêne, A., Farouk-Ameqrane, S., Sieber, F. und Maréchal-Drouard, L. (2011) Import of RNAs into Plant Mitochondria. In Kempken, F., editor, *Plant Mitochondria*, volume 1 of *Advances in Plant Biology*, Seiten 241–260. Springer New York.

Duchêne, A. M., Pujol, C. und Maréchal-Drouard, L. (2009) Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Current Genetics*, 55(1):1–18.

Ellis, J. (1982) Promiscuous DNA-chloroplast genes inside plant mitochondria. *Nature*, 299:678–679.

Fauron, C., Casper, M., Gao, Y. und Moore, B. (1995) The maize mitochondrial genome: dynamic, yet functional. *Trends in Genetics*, 11(6):228–235.

Feagin, J., Abraham, J. und Stuart, K. (1988) Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. *Cell*, 53(3):413–422.

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39(4):783–791.

Fey, J., Tomita, K., Bergdoll, M. und Maréchal-Drouard, L. (2000) Evolutionary and functional aspects of C-to-U editing at position 28 of tRNA(Cys)(GCA) in plant mitochondria. *RNA*, 6(4):470–474.

Fox, T. D. und Leaver, C. J. (1981) The zea mays mitochondrial gene coding cytochrome oxidase subunit II has an intervening sequence and does not contain TGA codons. *Cell*, 26(3):315–323.

Fujii, S., Kazama, T., Yamada, M. und Toriyama, K. (2010) Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. *BMC Genomics*, 11(1):209.

Gensel, P. G. und Berry, C. M. (2001) Early lycophyte evolution. *American Fern Journal*, 91(3):74–98.

Ghosh, H. P., Ghosh, K., Simsek, M. und RajBhandary, U. L. (1982) Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. *Nucleic Acids Research*, 10(10):3241 –3247.

Giegé, P. und Brennicke, A. (1999) RNA editing in Arabidopsis mitochondria effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26):15324–15329.

Giegé, P., Grienenberger, J. und Bonnard, G. (2008) Cytochrome c biogenesis in mitochondria. *Mitochondrion*, 8(1):61–73.

Gissi, C., Iannelli, F. und Pesole, G. (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*, 101(4):301–320.

Glover, K. E., Spencer, D. F. und Gray, M. W. (2001) Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria. *Journal of Biological Chemistry*, 276(1):639 –648.

Goremykin, V. V., Salamini, F., Velasco, R. und Viola, R. (2009) Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution*, 26(1):99–110.

Grandbastien, M. (1992) Retroelements in higher plants. *Trends in Genetics*, 8(3):103–108.

Gray, M., Burger, G. und Lang, B. (1999) Mitochondrial evolution. *Science*, 283(5407):1476–1481.

Gray, M. W., Lang, B. F. und Burger, G. (2004) Mitochondria of protists. *Genetics*, 38(1):477.

Grewe, F., Herres, S., Viehöver, P., Polsakiewicz, M., Weisshaar, B. und Knoop, V. (2010) A unique transcriptome: 1782 positions of RNA editing alter 1406

codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. *Nucleic Acids Research*. DOI:10.1093/nar/gkq1227.

Grewe, F., Viehöver, P., Weisshaar, B. und Knoop, V. (2009) A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. *Nucleic Acids Research*, 37(15):5093–5104.

Grosskopf, D. und Mulligan, R. (1996) Developmental-and tissue-specificity of RNA editing in mitochondria of suspension-cultured maize cells and seedlings. *Current Genetics*, 29(6):556–563.

Groth-Malonek, M., Pruchner, D., Grewe, F. und Knoop, V. (2005) Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Molecular Biology and Evolution*, 22(1):117–125.

Groth-Malonek, M., Wahrmund, U., Polsakiewicz, M. und Knoop, V. (2007) Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial nad7 Gene Supports Haplomitrium as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae. *Molecular Biology and Evolution*, 24(4):1068 –1074.

Gualberto, J., Lamattina, L., Bonnard, G., Weil, J. und Grienenberger, J. (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. *Nature*, 341(6243):660–662.

Gualberto, J. M., Weil, J. und Grienenberger, J. (1990) Editing of the wheat coxIII transcript: evidence for twelve C to U and one U to C conversions and for sequence similarities around editing sites. *Nucleic Acids Research*, 18(13):3771–3776.

Haen, K., Pett, W. und Lavrov, D. (2010) Parallel loss of nuclear-encoded mitochondrial aminoacyl-tRNA synthetases and mtDNA-encoded tRNAs in Cnidaria. *Molecular Biology and Evolution*, 27(10):2216–2219.

Handa, H. (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. *Nucleic Acids Research*, 31(20):5907 –5916.

Handa, H., Itani, K. und Sato, H. (2002) Structural features and expression analysis of a linear mitochondrial plasmid in rapeseed (Brassica napus L.). *Molecular Genetics and Genomics*, 267(6):797–805.

Hasebe, M., Wolf, P. G., Pryer, K. M., Ueda, K., Ito, M., Sano, R., Gastony, G. J., Yokoyama, J., Manhart, J. R., Murakami, N., Crane, E. H., Haufler, C. H. und Hauk, W. D. (1995) Fern phylogeny based on rbcL nucleotide sequences. *American Fern Journal*, 85(4):134–181.

Haugen, P., Simon, D. und Bhattacharya, D. (2005) The natural history of group I introns. *Trends in Genetics*, 21(2):111–119.

Hiesel, R., Combettes, B. und Brennicke, A. (1994) Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proceedings of the National Academy of Sciences of the United States of America*, 91(2):629–633.

Hiesel, R., Wissinger, B., Schuster, W. und Brennicke, A. (1989) RNA editing in plant mitochondria. *Science*, 246(4937):1632–1634.

Hoch, B., Maier, R. M., Appel, K., Igloi, G. L. und Kossel, H. (1991) Editing of a chloroplast mRNA by creation of an initiation codon. *Nature*, 353(6340):178–180.

Howad, W. und Kempken, F. (1997) Cell type-specific loss of atp6 RNA editing in cytoplasmic male sterile Sorghum bicolor. *Proceedings of the National Academy of Sciences of the United States of America*, 94(20):11090–11095.

Janke, A. und Pääbo, S. (1993) Editing of a tRNA anticodon in marsupial mitochondria changes its codon recognition. *Nucleic Acids Research*, 21(7):1523–1525.

Jobson, R. und Qiu, Y. (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? *Biology Direct*, 3(1):43.

Joyce, P. B. und Gray, M. W. (1989) Chloroplast-like transfer RNA genes expressed in wheat mitochondria. *Nucleic Acids Research*, 17(14):5461–5476.

Karcher, D. und Bock, R. (2002a) Temperature sensitivity of RNA editing and intron splicing reactions in the plastid ndhB transcript. *Current Genetics*, 41(1):48–52.

Karcher, D. und Bock, R. (2002b) The amino acid sequence of a plastid protein is developmentally regulated by RNA editing. *Journal of Biological Chemistry*, 277(7):5570–5574.

Kazuyoshi, K. und Tomohiko, K. (2010) Cost of having the largest mitochondrial genome: Evolutionary mechanism of plant mitochondrial genome. *Journal of Botany*, 2010. DOI:10.1155/2010/620137.

Keeling, P. (2009) Role of horizontal gene transfer in the evolution of photosynthetic eukaryotes and their plastids. *Methods in Molecular Biology*, 532:501–515.

Kempken, F., Bolle, N., Forner, J. und Binder, S. (2007) Transcript end mapping and analysis of RNA editing in plant mitochondria. In Leister, D. und Herrmann, J., editors, *Mitochondria: Practical Protokols*, volume 372, Seiten 177–192. Humana Press, Totowa, NJ.

Keren, I., Bezawork-Geleta, A., Kolton, M., Maayan, I., Belausov, E., Levy, M., Mett, A., Gidoni, D., Shaya, F. und Ostersetzer-Biran, O. (2009) AtnMat2, a nuclear-encoded maturase required for splicing of group-II introns in Arabidopsis mitochondria. *RNA*, 15(12):2299–2311.

Khandelwan, S. (1990) Chromosome evolution in the genus Ophioglossum L. *Botanical Journal of the Linnean Society*, 102(3):205–217.

Kidwell, M. (2002) Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, 115(1):49–63.

Kim, E., Lane, C., Curtis, B., Kozera, C., Bowman, S. und Archibald, J. (2008) Complete sequence and analysis of the mitochondrial genome of Hemiselmis andersenii CCMP 644(Cryptophyceae). *BMC Genomics*, 9(1):215.

Kim, J., Vanguri, S., Boeke, J., Gabriel, A. und Voytas, D. (1998) Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence. *Genome Research*, 8(5):464–478.

Kim, S. und Yoon, M. (2010) Comparison of mitochondrial and chloroplast genome segments from three onion (Allium cepa L.) cytoplasm types and identification of a trans-splicing intron of cox2. *Current Genetics*, 56(2):177–188.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2):111–120.

Knoop, V. (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics*, 46(3):123–139.

Knoop, V. (2010) When you can't trust the DNA: RNA editing changes transcript sequences. *Cellular and Molecular Life Sciences*, 68(4):567–586.

Knoop, V. und Rüdinger, M. (2010) DYW-type PPR proteins in a heterolobosean protist: Plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Letters*, 584(20):4287–4291.

Knoop, V., Schuster, W., Wissinger, B. und Brennicke, A. (1991) Trans splicing integrates an exon of 22 nucleotides into the nad5 mRNA in higher plant mitochondria. *The EMBO Journal*, 10(11):3483–3493.

Knoop, V., Unseld, M., Marienfeld, J., Brandt, P., Sunkel, S., Ullrich, H. und Brennicke, A. (1996) copia-, gypsy- and LINE-Like Retrotransposon Fragments in the Mitochondrial Genome of Arabidopsis thaliana. *Genetics*, 142(2):579–585.

Knoop, V., Volkmar, U., Hecht, J. und Grewe, F. (2011) Mitochondrial genome evolution in the plant lineage. In Kempken, F., editor, *Plant Mitochondria*, volume 1 of *Advances in Plant Biology*, Seiten 3–29. Springer New York.

Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A. und Ohyama, K. (1997) Interorganellar gene transfer in bryophytes: the functional nad7 gene is nuclear encoded in Marchantia polymorpha. *Molecular and General Genetics*, 256(5):589–592.

Koller, B., Fromm, H., Galun, E. und Edelman, M. (1987) Evidence for in vivo trans splicing of pre-mRNAs in tobacco chloroplasts. *Cell*, 48(1):111–119.

Kotera, E., Tasaka, M. und Shikanai, T. (2005) A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, 433(7023):326–330.

Koulintchenko, M., Konstantinov, Y. und Dietrich, A. (2003) Plant mitochondria actively import DNA via the permeability transition pore complex. *The EMBO Journal*, 22(6):1245–1254.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. und Rehrauer, W. M. (1994) Biochemistry of homologous recombination in Escherichia coli. *Microbiology and Molecular Biology Reviews*, 58(3):401–465.

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. und Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, 31(1):147–157.

Kubo, T. und Mikami, T. (2007) Organization and variation of angiosperm mitochondrial genome. *Physiologia Plantarum*, 129(1):6–13.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A. und Mikami, T. (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNACys(GCA). *Nucleic Acids Research*, 28(13):2571–2576.

Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T. und Yoshinaga, K. (2003) RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Research*, 31(9):2417–2423.

Kumar, S., Nei, M., Dudley, J. und Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics*, 9(4):299–306.

Kumar, S., Tamura, K. und Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics*, 5(2):150–163.

Laforest, M., Roewer, I. und Lang, B. (1997) Mitochondrial tRNAs in the lower fungus Spizellomyces punctatus tRNA editing and UAG stop codons recognized as leucine. *Nucleic Acids Research*, 25(3):626–632.

Lang, B., Goff, L. und Gray, M. (1996) A 5 S rRNA gene is present in the mitochondrial genome of the protist Reclinomonas americana but is absent from red algal mitochondrial DNA. *Journal of Molecular Biology*, 261(5):607–613.

Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M. und Gray, M. W. (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature*, 387(6632):493–497.

Lang, B. F., Laforest, M. und Burger, G. (2007) Mitochondrial introns: a critical view. *Trends in Genetics*, 23(3):119–125.

Lavrov, D., Brown, W. und Boore, J. (2000) A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede Lithobius forficatus. *Proceedings of the National Academy of Sciences of the United States of America*, 97(25):13738–13742.

Lavrov, D. V. (2007) Key transitions in animal evolution: a mitochondrial DNA perspective. *Integrative and Comparative Biology*, 47(5):734–743.

Leitch, I., Soltis, D., Soltis, P. und Bennett, M. (2005) Evolution of DNA amounts across land plants (Embryophyta). *Annals of Botany*, 95(1):207–217.

Lenz, H., Rüdinger, M., Volkmar, U., Fischer, S., Herres, S., Grewe, F. und Knoop, V. (2010) Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Current Genetics*, 56(2):189–201.

Li, L., Wang, B., Liu, Y. und Qiu, Y. (2009) The complete mitochondrial genome sequence of the hornwort Megaceros aenigmaticus shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *Journal of Molecular Evolution*, 68(6):665–678.

Lin, Z., Kong, H., Nei, M. und Ma, H. (2006) Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences of the United States of America*, 103(27):10328–10333.

Lindauer, A., Fraser, D., Brüderlein, M. und Schmitt, R. (1993) Reverse transcriptase families and a copia-like retrotransposon, Osser, in the green alga Volvox carteri. *FEBS Letters*, 319(3):261–266.

Lohse, M., Drechsel, O. und Bock, R. (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics*, 52(5):267–274.

Lonergan, K. und Gray, M. (1993) Editing of transfer RNAs in Acanthamoeba castellanii mitochondria. *Science*, 259(5096):812–816.

Lonsdale, D. M., Brears, T., Hodge, T. P., Melville, S. E. und Rottmann, W. H. (1988) The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 319(1193):149–163.

Lowe, T. und Eddy, S. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, 25(5):955–964.

Lukeš, J., Leander, B. und Keeling, P. (2009) Cascades of convergent evolution: The corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proceedings of the National Academy of Sciences of the United States of America*, 106(Supplement 1):9963–9970.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M.-L., Mireau, H., Peeters, N., Renou, J.-P., Szurek, B., Taconnat, L. und Small, I. (2004) Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant Cell*, 16(8):2089–2103.

Maier, R. M., Zeltz, P., Kössel, H., Bonnard, G., Gualberto, J. M. und Grienenberger, J. M. (1996) RNA editing in plant mitochondria and chloroplasts. *Plant Molecular Biology*, 32(1):343–365.

Maier, U., Bozarth, A., Funk, H., Zauner, S., Rensing, S., Schmitz-Linneweber, C., Börner, T. und Tillich, M. (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biology*, 6(1):36.

Malek, O. und Knoop, V. (1998) Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA*, 4(12):1599–1609.

Malek, O., Lättig, K., Hiesel, R., Brennicke, A. und Knoop, V. (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. *The EMBO Journal*, 15(6):1403–1411.

Manchekar, M., Scissum-Gunn, K., Song, D., Khazi, F., Mclean, S. und Nielsen, B. (2006) DNA recombination activity in soybean mitochondria. *Journal of Molecular Biology*, 356(2):288–299.

Marande, W. und Burger, G. (2007) Mitochondrial DNA as a genomic jigsaw puzzle. *Science*, 318(5849):415.

Marchfelder, A., Brennicke, A. und Binder, S. (1996) RNA editing is required for efficient excision of tRNA (Phe) from precursors in plant mitochondria. *Journal of Biological Chemistry*, 271(4):1898–1903.

Maréchal, A. und Brisson, N. (2010) Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, 186(2):299–317.

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

Marienfeld, J., Unseld, M. und Brennicke, A. (1999) The mitochondrial genome of Arabidopsis is composed of both native and immigrant information. *Trends in Plant Science*, 4(12):495–502.

Marienfeld, J. R., Unseld, M., Brandt, P. und Brennicke, A. (1997) Viral nucleic acid sequence transfer between fungi and plants. *Trends in Genetics*, 13(7):260–261.

Martick, M., Horan, L. H., Noller, H. F. und Scott, W. G. (2008) A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, 454(7206):899–902.

McClintock, B. (1984) The significance of responses of the genome to challenge. *Science*, 226(4676):792–801.

McDermott, P., Connolly, V. und Kavanagh, T. A. (2008) The mitochondrial genome of a cytoplasmic male sterile line of perennial ryegrass (Lolium perenne L.) contains an integrated linear plasmid-like element. *Theoretical and Applied Genetics*, 117(3):459–470.

Meinhardt, F., Kempken, F., Kämper, J. und Esser, K. (1990) Linear plasmids among eukaryotes: fundamentals and application. *Current Genetics*, 17(2):89–95.

Michel, F. und Ferat, J. (1995) Structure and activities of group II introns. *Annual Review of Biochemistry*, 64(1):435–461.

Michel, F., Umesono, K. und Ozeki, H. (1989) Comparative and functional anatomy of group II catalytic introns–a review. *Gene*, 82(1):5–30.

Michel, F. und Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *Journal of Molecular Biology*, 216(3):585–610.

Miyata, Y., Sugita, C., Maruyama, K. und Sugita, M. (2008) RNA editing in the anticodon of tRNA Leu (CAA) occurs before group I intron splicing in plastids of a moss Takakia lepidozioides S. Hatt. & Inoue. *Plant Biology*, 10(2):250–255.

Miyata, Y. und Sugita, M. (2004) Tissue-and stage-specific RNA editing of rps14 transcripts in moss (Physcomitrella patens) chloroplasts. *Journal of Plant Physiology*, 161(1):113–115.

Mower, J. und Palmer, J. (2006) Patterns of partial RNA editing in mitochondrial genes of Beta vulgaris. *Molecular Genetics and Genomics*, 276(3):285–293.

Nakazato, T., Jung, M.-K., Housworth, E. A., Rieseberg, L. H. und Gastony, G. J. (2006) Genetic map-based analysis of genome structure in the homosporous fern Ceratopteris richardii. *Genetics*, 173(3):1585–1597.

Nakazono, M., Nishiwaki, S., Tsutsumi, N. und Hirai, A. (1996) A chloroplast-derived sequence is utilized as a source of promoter sequences for the gene for subunit 9 of NADH dehydrogenase (nad9) in rice mitochondria. *Molecular and General Genetics*, 252(4):371–378.

Notsu, Notsu, Y., Masood, Masood, S., Nishikawa, Nishikawa, T., Kubo, Kubo, N., Akiduki, Akiduki, G., Nakazono, Nakazono, M., Hirai, Hirai, A., Kadowaki und Kadowaki, K. (2002) The complete sequence of the rice (Oryza sativa L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, 268(4):434–445.

Nugent, J. und Palmer, J. (1991) RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. *Cell*, 66(3):473–481.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T. und . (1992) Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. *Journal of Molecular Biology*, 223(1):1–7.

Odahara, M., Inouye, T., Fujita, T., Hasebe, M. und Sekine, Y. (2007) Involvement of mitochondrial-targeted RecA in the repair of mitochondrial DNA in the moss, Physcomitrella patens. *Genes and Genetic Systems*, 82(1):43–51.

Odahara, M., Kuroiwa, H., Kuroiwa, T. und Sekine, Y. (2009) Suppression of Repeat-Mediated Gross Mitochondrial Genome Rearrangements by RecA in the Moss Physcomitrella patens. *The Plant Cell*, 21(4):1182–1194.

Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S. und Tsunewaki, K. (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Research*, 33(19):6235–6250.

Ohta, E., Oda, K., Yamato, K., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Ohyama, K. und Michel, F. (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. *Nucleic Acids Research*, 21(5):1297–1305.

Ohyama, K., Oda, K., Ohta, E. und Takemura, M. (1993) Gene organization and evolution of introns of a liverwort, Marchantia polymorpha, mitochondrial genome. In *Plant Mitochondria*, Seiten 115–129. Weinheim: VCH Verlagsgesellschaft.

Ohyama, K. und Takemura, M. (2008) Molecular evolution of mitochondrial introns in the liverwort Marchantia polymorpha. *Proceedings of the Japan Academy, Series B*, 84(1):17–23.

Oldenburg, D. und Bendich, A. (1998) The structure of mitochondrial DNA from the liverwort, Marchantia polymorpha. *Journal of Molecular Biology*, 276(4):745–758.

Oldenburg, D. und Bendich, A. (2001) Mitochondrial DNA from the liverwort Marchantia polymorpha: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. *Journal of Molecular Biology*, 310(3):549–562.

Palmer, J. D. und Shields, C. R. (1984) Tripartite structure of the Brassica campestris mitochondrial genome. *Nature*, 307(5950):437–440.

Palmer, J. D., Soltis, D. und Soltis, P. (1992) Large size and complex structure of mitochondrial DNA in two nonflowering land plants. *Current Genetics*, 21(2):125–129.

Panarese, S., Rainaldi, G., Benedetto, C. D. und Gallerani, R. (2008) Sequencing of a segment of a monilophyte species mitochondrial genome reveals features highly similar to those of seed plant mtDNAs. *The Open Plant Science Journal*, 2(1):15–20.

Phillips, T. L. und DiMichele, W. A. (1992) Comparative ecology and life-history biology of arborescent lycopsids in Late Carboniferous swamps of Euramerica. *Annals of the Missouri Botanical Garden*, 79(3):560–588.

Picardi, E., Horner, D., Chiara, M., Schiavon, R., Valle, G. und Pesole, G. (2010) Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Research*, 38(14):4755–4767.

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S. und Sipes, S. D. (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature*, 409(6820):618–622.

Pryer, K. M., Schuettpelz, E., Wolf, P. G., Schneider, H., Smith, A. R. und Cranfill, R. (2004) Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *American Journal of Botany*, 91(10):1582–1598.

Pryer, K. M., Smith, A. R. und Skog, J. E. (1995) Phylogenetic relationships of extant ferns based on evidence from morphology and rbcL sequences. *American Fern Journal*, 85(4):205–282.

Qiu, Y. (2008) Phylogeny and evolution of charophytic algae and land plants. *Journal of Systematics and Evolution*, 46(3):287–306.

Qiu, Y., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G., Hendry, T., Taylor, D., Testa, C., Ambros, M., Crandall-Stotler, B., Duff, R., Stech, M., Frey, W., Quandt, D. und Davis, C. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42):15511–15516.

Qiu, Y. und Palmer, J. D. (2004) Many independent origins of trans splicing of a plant mitochondrial group II intron. *Journal of Molecular Evolution*, 59(1):80–89.

Qiu, Y. L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R. W. und Hendry, T. A. (2007) A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences*, 168(5):691–708.

Rajasekhar, V. und Mulligan, R. (1993) RNA editing in plant mitochondria:[alpha]-phosphate is retained during C-to-U conversion in mRNAs. *The Plant Cell*, 5(12):1843–1852.

Randau, L., Munch, R., Hohn, M. J., Jahn, D. und Soll, D. (2005) Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature*, 433(7025):537–541.

Randau, L. und Söll, D. (2008) Transfer RNA genes in pieces. *EMBO Reports*, 9(7):623–628.

Rasmusson, A. G., Handa, H. und Møller, I. M. (2008) Plant mitochondria, more unique than ever. *Mitochondrion*, 8(1):1–4.

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P., Lindquist, E. A., Kamisugi, Y., Tanahashi, T., Sakakibara, K., Fujita, T., Oishi, K., Shin-I, T., Kuroki, Y., Toyoda, A., Suzuki, Y., ichi Hashimoto, S., Yamaguchi, K., Sugano, S., Kohara, Y., Fujiyama, A., Anterola, A., Aoki, S., Ashton, N., Barbazuk, W. B., Barker, E., Bennetzen, J. L., Blankenship, R., Cho, S. H., Dutcher, S. K., Estelle, M., Fawcett, J. A., Gundlach, H., Hanada, K., Heyl, A., Hicks, K. A., Hughes, J., Lohr, M., Mayer, K., Melkozernov, A., Murata, T., Nelson, D. R., Pils, B., Prigge, M., Reiss, B., Renner, T., Rombauts, S., Rushton, P. J., Sanderfoot, A., Schween, G., Shiu, S., Stueber, K., Theodoulou, F. L., Tu, H., de Peer, Y. V., Verrier, P. J., Waters, E., Wood, A., Yang, L., Cove, D., Cuming, A. C., Hasebe, M., Lucas, S., Mishler, B. D., Reski, R., Grigoriev, I. V., Quatrano, R. S. und Boore, J. L. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science*, 319(5859):64–69.

Richardson, A. und Palmer, J. (2007) Horizontal gene transfer in plants. *Journal of Experimental Botany*, 58(1):1–9.

Robart, A. und Zimmerly, S. (2005) Group II intron retroelements: function and diversity. *Cytogenetic and Genome Research*, 110(1-4):589–597.

Robison, M. M. und Wolyn, D. J. (2004) A mitochondrial plasmid and plasmid-like RNA and DNA polymerases encoded within the mitochondrial genome of carrot (Daucus carota L.). *Current Genetics*, 47(1):57–66.

Rüdinger, M., Funk, H., Rensing, S., Maier, U. und Knoop, V. (2009) RNA editing: only eleven sites are present in the Physcomitrella patens mitochondrial transcriptome and a universal nomenclature proposal. *Molecular Genetics and Genomics*, 281(5):473–481.

Rüdinger, M., Polsakiewicz, M. und Knoop, V. (2008) Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Molecular Biology and Evolution*, 25(7):1405 –1414.

Ruf, S. und Kössel, H. (1997) Tissue-specific and differential editing of the two ycf3 editing sites in maize plastids. *Current Genetics*, 32(1):19–23.

Saitou, N. und Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406 –425.

Sakurai, R., Nomura, H., Moriyam, Y. und Kawano, S. (2004) The mitochondrial plasmid of the true slime mold Physarum polycephalum bypasses uniparental inheritance by promoting mitochondrial fusion. *Current Genetics*, 46(2):103– 114.

Saldanha, R., Mohr, G., Belfort, M. und Lambowitz, A. (1993) Group I and group II introns. *The FASEB Journal*, 7(1):15.

Salmans, M. L., Chaw, S., Lin, C., Shih, A. C., Wu, Y. und Mulligan, R. M. (2010) Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. *Current Genetics*, 56(5):439–446.

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Grothmalonek, M., Szurek, B., Small, I., Knoop, V. und Lurin, C. (2007) A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Letters*, 581(22):4132– 4138.

Sambrook, J., Fritsch, E. und Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, volume 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SanMiguel, P. und Bennetzen, J. (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Annals of Botany*, 82(Suppl A):37.

SanMiguel, P., Tikhonov, A., Jin, Y., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P., Edwards, K., Lee, M., Avramova, Z. und Bennetzen, J. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science*, 274(5288):765–768.

Satoh, M., Kubo, T. und Mikami, T. (2006) The Owen mitochondrial genome in sugar beet (Beta vulgaris L.): possible mechanisms of extensive rearrangements and the origin of the mitotype-unique regions. *Theoretical and Applied Genetics*, 113(3):477–484.

Schaefer, D. G. und Zryd, J. (1997) Efficient gene targeting in the moss Physcomitrella patens. *The Plant Journal*, 11(6):1195–1206.

Schuster, W., Hiesel, R., Wissinger, B. und Brennicke, A. (1990) RNA editing in the cytochrome b locus of the higher plant Oenothera berteriana includes a U-to-C transition. *Molecular and Cellular Biology*, 10(5):2428–2431.

Shapiro, J. (2010) Mobile DNA and evolution in the 21st century. *Mobile DNA*, 1:4.

Shedge, V., Arrieta-Montiel, M., Christensen, A. C. und Mackenzie, S. A. (2007) Plant Mitochondrial Recombination Surveillance Requires Unusual RecA and MutS Homologs. *The Plant Cell*, 19(4):1251–1264.

Shutt, T. E. und Gray, M. W. (2006) Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics*, 22(2):90–95.

Simon, D. M., Clarke, N. A., McNeil, B. A., Johnson, I., Pantuso, D., Dai, L., Chai, D. und Zimmerly, S. (2008) Group II introns in Eubacteria and Archaea: ORF-less introns and new varieties. *RNA*, 14(9):1704 –1713.

Sloan, D. B., Alverson, A. J., Storchová, H., Palmer, J. D. und Taylor, D. R. (2010a) Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. *BMC Evolutionary Biology*, 10(1):274.

Sloan, D. B., MacQueen, A. H., Alverson, A. J., Palmer, J. D. und Taylor, D. R. (2010b) Extensive loss of RNA editing sites in rapidly evolving Silene mitochondrial genomes: Selection vs. retroprocessing as the driving force. *Genetics*. DOI:10.1534/genetics.110.118000.

Small, I., Isaac, P. und Leaver, C. (1987) Stoichiometric differences in DNA molecules containing the atpA gene suggest mechanisms for the generation of mitochondrial genome diversity in maize. *The EMBO Journal*, 6(4):865–869.

Smith, A., Pryer, K., Schuettpelz, E., Korall, P., Schneider, H. und Wolf, P. (2006) A classification for extant ferns. *Taxon*, 55(3):705–731.

Smith, D. R. (2009) Unparalleled GC content in the plastid DNA of Selaginella. *Plant Molecular Biology*, 71(6):627–639.

Smith, D. R. und Lee, R. W. (2008) Mitochondrial genome of the colorless green alga Polytomella capuana: A linear molecule with an unprecedented GC content. *Molecular Biology and Evolution*, 25(3):487–496.

Soma, A., Onodera, A., Sugahara, J., Kanai, A., Yachie, N., Tomita, M., Kawamura, F. und Sekine, Y. (2007) Permuted tRNA genes expressed via a circular RNA intermediate in Cyanidioschyzon merolae. *Science*, 318(5849):450–453.

Steinhauser, S., Beckert, S., Capesius, I., Malek, O. und Knoop, V. (1999) Plant Mitochondrial RNA Editing: extreme in hornworts and dividing the liverworts? *Journal of Molecular Evolution*, 48(3):303–312.

Stern, D. und Lonsdale, D. (1982) Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature*, 299(5885):698–702.

Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A. und Sugiura, M. (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Molecular Genetics and Genomics*, 272(6):603–615.

Takenaka, M., Verbitskiy, D., van der Merwe, J. A., Zehrmann, A. und Brennicke, A. (2008) The process of RNA editing in plant mitochondria. *Mitochondrion*, 8(1):35–46.

Tamura, K., Dudley, J., Nei, M. und Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24(8):1596 –1599.

Tasaki, E., Hattori, M. und Sugita, M. (2010) The moss pentatricopeptide repeat protein with a DYW domain is responsible for RNA editing of mitochondrial ccmFc transcript. *The Plant Journal*, 62(4):560–570.

Taylor, W. C. und Hickey, R. J. (1992) Habitat, evolution, and speciation in Isoetes. *Annals of the Missouri Botanical Garden*, 79(3):613–622.

Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M. und Sato, N. (2007) The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. *Molecular Biology and Evolution*, 24(3):699–709.

Thomson, J. (2000) Morphological and genomic diversity in the genus Pteridium (Dennstaedtiaceae). *Annals of Botany*, 85(suppl 2):77–99.

Tian, X., Zheng, J., Hu, S. und Yu, J. (2006) The rice mitochondrial genomes and their variations. *Plant Physiology*, 140(2):401–410.

Tillich, M., Lehwark, P., Morton, B. R. und Maier, U. G. (2006) The Evolution of chloroplast RNA editing. *Molecular Biology and Evolution*, 23(10):1912–1921.

Turmel, M., Otis, C. und Lemieux, C. (2002a) The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeridium globosum: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 99(17):11275–11280.

Turmel, M., Otis, C. und Lemieux, C. (2002b) The complete mitochondrial DNA sequence of Mesostigma viride identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Molecular Biology and Evolution*, 19(1):24–38.

Turmel, M., Otis, C. und Lemieux, C. (2003) The mitochondrial genome of Chara vulgaris: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *The Plant Cell*, 15(8):1888–1903.

Turmel, M., Otis, C. und Lemieux, C. (2007) An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga Chlorokybus atmophyticus. *BMC Genomics*, 8(1):137.

Ullrich, H., Lättig, K., Brennicke, A. und Knoop, V. (1997) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 Arabidopsis thaliana ecotypes. *Plant Molecular Biology*, 33(1):37–45.

Unseld, M., Marienfeld, J., Brandt, P. und Brennicke, A. (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. *Nature Genetics*, 15(1):57–61.

Vaidya, A. B., Akella, R. und Suplick, K. (1989) Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-kilobase-pair DNA of a malarial parasite. *Molecular and Biochemical Parasitology*, 35(2):97–107.

Vangerow, S., Teerkorn, T. und Knoop, V. (1998) Phylogenetic information in the mitochondrial nad5 gene of pteridophytes: RNA editing and intron sequences. *Plant Biology*, 1:235–243.

Vaughn, J. C., Mason, M. T., Sper-Whitis, G. L., Kuhlman, P. und Palmer, J. D. (1995) Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric coxI gene of Peperomia. *Journal of Molecular Evolution*, 41(5):563–572.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D. A., Cestaro, A., Pruss, D., Pindo, M., FitzGerald, L. M., Vezzulli, S., Reid, J., Malacarne, G., Iliev, D., Coppola, G., Wardell, B., Micheletti, D., Macalma, T., Facci, M., Mitchell, J. T., Perazzolli, M., Eldredge, G., Gatto, P., Oyzerski, R., Moretto, M., Gutin, N., Stefanini, M., Chen, Y., Segala, C., Davenport, C., Demattè, L., Mraz, A., Battilana, J., Stormo, K., Costa, F., Tao, Q., Si-Ammour, A., Harkins, T., Lackey, A., Perbost, C., Taillon, B., Stella, A., Solovyev, V., Fawcett, J. A., Sterck, L., Vandepoele, K., Grando, S. M., Toppo, S., Moser, C., Lanchbury, J., Bogden, R., Skolnick, M., Sgaramella, V., Bhatnagar, S. K., Fontana, P., Gutin, A., Van de Peer, Y., Salamini, F. und Viola, R. (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE*, 2(12):1326.

Wang, B., Xue, J., Li, L., Liu, Y. und Qiu, Y. (2009) The complete mitochondrial genome sequence of the liverwort Pleurozia purpurea reveals extremely conservative mitochondrial genome evolution in liverworts. *Current Genetics*, 55(6):601–609.

Wang, D., Wu, Y., Shih, A. C., Wu, C., Wang, Y. und Chaw, S. (2007) Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 MYA. *Molecular Biology and Evolution*, 24(9):2040–2048.

Wang, W., Tanurdzic, M., Luo, M., Sisneros, N., Kim, H., Weng, J., Kudrna, D., Mueller, C., Arumuganathan, K., Carlson, J., Chapple, C., de Pamphilis, C., Mandoli, D., Tomkins, J., Wing, R. und Banks, J. (2005) Construction of a bacterial artificial chromosome library from the spikemoss Selaginella moellendorffii: a new resource for plant comparative genomics. *BMC Plant Biology*, 5(1):10.

Ward, B. L., Anderson, R. S. und Bendich, A. J. (1981) The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, 25(3):793–803.

Weber, B., Börner, T. und Weihe, A. (1995) Remnants of a DNA polymerase gene in the mitochondrial DNA of Marchantia polymorpha. *Current Genetics*, 27(5):488–490.

Wendel, J. (2000) Genome evolution in polyploids. *Plant Molecular Biology*, 42(1):225–249.

Weng, J., Akiyama, T., Bonawitz, N., Li, X., Ralph, J. und Chapple, C. (2010) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants. *The Plant Cell*, 22(4):1033–1045.

Wikström, N. und Pryer, K. M. (2005) Incongruence between primary sequence data and the distribution of a mitochondrial atp1 group II intron among ferns and horsetails. *Molecular Phylogenetics and Evolution*, 36(3):484–493.

Wissinger, B., Schuster, W. und Brennicke, A. (1991) Trans splicing in Oenothera mitochondria: nad1 mRNAs are edited in exon and trans-splicing group II intron sequences. *Cell*, 65(3):473–482.

Wolf, P. G., Rowe, C. A. und Hasebe, M. (2004) High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern Adiantum capillus-veneris. *Gene*, 339:89–97.

Xu, M., Brar, H. K., Grosic, S., Palmer, R. G. und Bhattacharyya, M. K. (2010) Excision of an active CACTA-like transposable element from DFR2 causes variegated flowers in soybean [Glycine max (L.) Merr.]. *Genetics*, 184(1):53–63.

Xue, J., Liu, Y., Li, L., Wang, B. und Qiu, Y. (2010) The complete mitochondrial genome sequence of the hornwort Phaeoceros laevis: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics*, 56(1):53–61.

Yoder, J., Walsh, C. und Bestor, T. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics*, 13(8):335–340.

Yoshinaga, K., Iinuma, H., Masuzawa, T. und Uedal, K. (1996) Extensive RNA editing of U to C in addition to C to U substitution in the rbcL transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Research*, 24(6):1008–1014.

Yu, W. und Schuster, W. (1995) Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria. *Journal of Biological Chemistry*, 270(31):18227 –18233.

Zaegel, V., Guermann, B., Ret, M. L., Andres, C., Meyer, D., Erhardt, M., Canaday, J., Gualberto, J. M. und Imbault, P. (2006) The plant-specific ssDNA binding protein OSB1 is involved in the stoichiometric transmission of mitochondrial DNA in Arabidopsis. *The Plant Cell*, 18(12):3548–3563.

Zehrmann, A., van der Merwe, J., Verbitskiy, D., Brennicke, A. und Takenaka, M. (2008) Seven large variations in the extent of RNA editing in plant mitochondria between three ecotypes of Arabidopsis thaliana. *Mitochondrion*, 8(4):319–327.

Zehrmann, A., Verbitskiy, D., van der Merwe, J. A., Brennicke, A. und Takenaka, M. (2009) A DYW domain-containing pentatricopeptide repeat protein is required for RNA editing at multiple sites in mitochondria of Arabidopsis thaliana. *The Plant Cell*, 21(2):558–567.

Zimmerly, S., Hausner, G. und chu Wu, X. (2001) Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Research*, 29(5):1238–1250.

Danksagung

Mein Dank gilt als erstes Herrn Professor Dr. Volker Knoop für die Möglichkeit, dass ich meine Doktorarbeit in seiner Arbeitsgruppe anfertigen konnte, und er meine Arbeit mit seiner ansteckenden Begeisterung sowie zahlreichen Ratschlägen und Ideen stets bereichert hat. Dieser Dank gilt auch für die Unterstützung beim Aufbau von Kooperationen und den großen Einsatz beim Erstellen der Publikationen.

Ich danke Herrn Professor Dr. Dietmar Quandt für die Übernahme des Co-Referats, seine Hilfsbereitschaft und sein Interesse an meiner Arbeit. Ebenso danke ich Herrn Professor Dr. Walter Witke und Herrn Professor Dr. Thomas Litt für die Bereitschaft zur Begutachtung dieser Arbeit.

Herrn Professor Bernd Weisshaar und Prisca Viehöver aus unserer Kooperationsgruppe an der Universität Bielefeld sowie Diana Lehmann und Yvonnne Steinbüchel vom MPI für Züchtungsforschung in Köln danke ich für die gute Zusammenarbeit und die große Unterstützung bei der Sortierung der Fosmidbanken und der Sequenzierung von Subbanken.

Außerdem möchte ich mich bei Herrn Professor Dr. Jeff Palmer und Herrn Dr. Erin Badenhop der Indiana University für die freundliche Bereitstellung der von Jerry Gastony gesammelten *Isoetes*-Pflanzen bedanken. Herrn Bernd Reinken und Herrn Dr. Wolfram Lobin vom Botanischen Garten Bonn danke ich für die mühsame Kultivierung von *Isoetes* und *Selaginella* sowie der Bereitstellung des seltenen *Gleichenia*-Pflanzenmaterials.

Ich danke meinem Freund und Kollegen Michael Gebert für eine interessante Zeit voller humorvoller Gespräche und hilfreicher Diskussionen. Bei Julia Hecht, Simon Fischer und Stefan Herres bedanke ich mich für eine gute und effiziente Zeit der Zusammenarbeit. Ein großes Dankeschön geht an Monika Polsakiewicz für ihre geduldige Unterstützung und die zahlreichen Ratschläge.

Schließlich danke ich allen anderen Mitarbeitern und ehemaligen Kollegen aus der Arbeitsgruppe Molekulare Evolution, insbesondere Julia Dreistein, Karolin Eifler, Milena Groth-Malonek, Katja Kirsch, Julia Lee, Henning Lenz, Karoline Meschenmoser, Theresia Rein, Mareike Rüdinger, Johanna Schmitz, Jan Seidenberg, Ute Volkmar und Katrin Weyand für die freundliche Atmosphäre, die gute Zusammenarbeit und die Anteilnahme an den Arbeiten anderer.

Ganz besonderer Dank geht an meine Freundin Christina, die mich stets in meiner Arbeit unterstützt hat und mir besonders in den letzten Monaten meiner Dissertation stets hilfreich zur Seite stand sowie meinen Eltern für ihre immerwährende Unterstützung und ihr Vertrauen.