J Mol Evol (2012) 74:37–51
DOI 10.1007/s00239-012-9486-3

Nuclear DYW-Type PPR Gene Families Diversify with Increasing RNA Editing Frequencies in Liverwort and Moss Mitochondria

Mareike Rüdinger · Ute Volkmar · Henning Lenz · Milena Groth-Malonek · Volker Knoop

Received: 5 July 2011 / Accepted: 11 January 2012 / Published online: 3 February 2012

Abstract RNA editing in mitochondria and chloroplasts of land plants alters transcript sequences by site-specific conversions of cytidines into uridines. RNA editing frequencies vary extremely between land plant clades, ranging from zero in some liverworts to more than 2,000 sites in lycophytes. Unique pentatricopeptide repeat (PPR) proteins with variable domain extension (E/E+/DYW) have recently been identified as specific editing site recognition factors in model plants. The distinctive functions of these PPR protein domain additions have remained unclear, although deaminase function has been proposed for the DYW domain. To shed light on diversity of RNA editing and DYW proteins at the origin of land plant evolution, we investigated editing patterns of the mitochondrial *nad*5, *nad*4, and *nad*2 genes in a wide sampling of more than 100 liverworts and mosses using the recently developed PREPACT program (www.prepact.de) and exemplarily confirmed predicted RNA editing sites in selected taxa. Extreme variability in RNA editing frequency is seen both in liverworts and mosses. Only few editings exist in the liverwort *Lejeunea cavifolia* or the moss *Pogonatum urnigerum* whereas up to 20% of cytidines are edited in the liverwort *Haplomitrium mnioides* or the moss *Takakia lepidozioides*. Interestingly, the latter are taxa that branch very early within their respective clades. Amplicons targeting the E/E+/DYW domains and subsequent random clone sequencing show DYW domains among bryophytes to be highly conserved in comparison with their angiosperm counterparts and to correlate well with RNA editing frequencies regarding their diversities. We propose that DYW proteins are the key players of RNA editing at the origin of land plants.

Keywords RNA editing · DYW-type PPR proteins · Mitochondria · Mosses · Liverworts · PREPACT

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9486-3) contains supplementary material, which is available to authorized users.

M. Rüdinger (✉) · U. Volkmar · H. Lenz · M. Groth-Malonek · V. Knoop
IZMB – Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany
e-mail: mrueding@uni-bonn.de

Present Address:
U. Volkmar
Instituut voor Plantkunde en Microbiologie, Laboratorium voor Plantensystematiek, K.U. Leuven, Kasteelpark Arenberg 31, 3001 Leuven, Belgium

Present Address:
M. Groth-Malonek
KWS SAAT AG, Grimsehlstr. 31, 37555 Einbeck, Germany

Introduction

In land plants, RNA editing manifests itself as targeted conversions of cytidines (C) into uridines (U) in organellar transcripts by deamination (Covello and Gray 1989; Gualberto et al. 1989; Hiesel et al. 1989; Hoch et al. 1991; Lamattina et al. 1989; Maier et al. 1992). This phenomenon of correcting genetic information is absent in algae and seems to have emerged concomitant with the water-to-land transition of embryophytes (Covello and Gray 1993; Jobson and Qiu 2008; Maier et al. 2008; Steinhauser et al. 1999). RNA editing occurs in all land plant clades (Freyer et al. 1997; Groth-Malonek et al. 2007; Hiesel et al. 1994; Malek et al. 1996; Sper-Whitis et al. 1994; Sper-Whitis et al. 1996) with one unique exception. In the subclass of

complex thalloid, marchantiid liverworts no RNA editing has been found. This is evident from the completely sequenced mitochondrial genome of *Marchantia polymorpha* (Ohyama et al. 2009) and from investigations of the mitochondrial genes *cox*1, *cox*3, *nad*5, and *nad*7 in several other marchantiid liverwort species (Groth-Malonek et al. 2007; Malek et al. 1996; Sper-Whitis et al. 1996; Steinhauser et al. 1999). In flowering plants, mitochondrial transcriptomes contain some 300–500 RNA editing sites (Giegé and Brennicke 1999; Handa 2003; Mower and Palmer 2006; Notsu et al. 2002) and chloroplast transcriptomes contain approx. 20–40 editing sites (Inada et al. 2004; Sasaki et al. 2003; Tillich et al. 2005; Tsudzuki et al. 2001). In the extant lycophytes, which represent the most ancient surviving lineage of vascular plants, transcriptome analyses reveal enormous frequencies of RNA editing, with more than 1,700 editing sites in the quillwort *Isoetes engelmannii* (Grewe et al. 2011) and even more than 2,100 editing sites in the spike moss *Selaginella moellendorffii* (Hecht et al. 2011). Fern allies, ferns, and hornworts also display abundant "reverse" uridine-to-cytidine RNA editing (Kugita et al. 2003; Steinhauser et al. 1999; Vangerow et al. 1999; Wolf et al. 2004; Wolf et al. 2005; Yoshinaga et al. 1996).

In contrast to abundant RNA editing in vascular plants, the model moss *Physcomitrella patens* and its close relative *Funaria hygrometrica* show only 11 and 8 editing sites, respectively, in their entire mitochondrial transcriptomes (Rüdinger et al. 2009, 2011b). Studies in *Haplomitrium mnioides*, however, a member of the basal-most subclass of liverworts (Haplomitriidae), showed RNA editing at more than 20 positions in its *nad*7 gene alone (Groth-Malonek et al. 2007).

In the meantime, several specific recognition factors for RNA editing sites have been identified in chloroplasts and mitochondria of plant model species like *Arabidopis thaliana* (reviewed in Fujii and Small 2011), *Oryza sativa* (Kim et al. 2009), or *Physcomitrella patens* (Ohtani et al. 2010; Rüdinger et al. 2011b; Tasaki et al. 2010). All of them belong to the large pentatricopeptide repeat (PPR) protein family, which was first described for the *Arabidopsis thaliana* nuclear genome where some 450 members are encoded (Lurin et al. 2004). PPR proteins are characterized by tandem repeats of a loosely conserved 35 amino acid motif (Small and Peeters 2000). Canonical PPR proteins of this type (the P subfamily) exist in numerous eukaryotes but plant genomes also encode unique PPR variants, referred to as the "PLS" type. Members of the PLS subfamily are characterized by long (L) and short (S) PPR motif length variants and many have optional C-terminal protein domain additions, the E, the E+ and the DYW domains, as successive extensions that, when present, always appear in this order (Lurin et al. 2004).

All RNA editing factors that have been characterized so far are members of the PLS subfamily of PPR proteins and carry at least the E/E+ or the complete suite of extensions including the DYW domain. The DYW domain in particular received attention given its weak similarity with cytidine deaminases (Salone et al. 2007). Moreover, the appearance of the DYW domain and RNA editing perfectly correlate, with DYW genes neither found in green algae nor in marchantiid liverworts where no RNA editing has been detected until now (Rüdinger et al. 2008; Salone et al. 2007). Intriguingly, all hitherto identified editing factors in *Physcomitrella* are DYW-type PLS proteins and no E/E+ type proteins lacking the DYW domain are encoded in the *Physcomitrella* genome (O'Toole et al. 2008) suggesting that the DYW domain indeed is intimately correlated with editing, at least in early plant evolution.

A robust molecular phylogeny of bryophytes has emerged over the recent years and extensive data sets are now available for a wide bryophyte taxon sampling of the mitochondrial *nad*5 gene plus widely sampled data sets for the *nad*2 gene specifically for mosses and for the *nad*4 gene recently compiled specifically for liverworts (Beckert et al. 1998; Volkmar et al. 2011). We used the recently developed PREPACT tool for prediction of editing sites in these genes, which we exemplarily confirmed on cDNA level for selected taxa. A new amendment to PREPACT, which allows for the use of multiple reference sequences for more conservative RNA editing site prediction, is introduced. To further investigate the correlation of RNA editing frequency and DYW-type gene diversity, we compared the DYW domain diversity for selected pairs of liverworts (*Haplomitrium mnioides* and *Lejeunea cavifolia*) and mosses (*Takakia lepidozioides* and *Pogonatum urnigerum*), which differ extremely in their RNA editing rates. Complementing the previous work, our new data emphasize a correlation between RNA editing frequency and DYW protein diversity in the two most ancient land plant clades. The altogether 184 E–E+–DYW domain sequences now available from liverworts and mosses allow for comparison of amino acid conservation with their vascular plant homologues.

Materials and Methods

Identification of RNA Editing Sites on DNA and cDNA Level

Total plant nucleic acids were extracted using either the CTAB method (Doyle and Doyle 1990) employing cetyltrimethyl-ammonium bromide as a detergent for cell lysis or the NucleoSpin Plant kit (Macherey–Nagel, Düren, Germany). RNA was prepared from plant material using

the NucleoSpin® RNA Plant kit (Macherey–Nagel). RNA was additionally treated with DNase I (Fermentas Life Sciences, St. Leon-Rot, Germany) to remove potential vestiges of DNA. First strand cDNA was synthesized using the RevertAid™ M-MulV® Reverse Transcriptase kit (Fermentas Life Science) and the hexanucleotide random primer mix (10 µM per assay; Carl Roth, Karlsruhe, Germany).

Primers were designed to target conserved regions of the *nad*4 (nad4up: 5′-acagccaaatttcartttgtggaa-3′ and nad4do: 5′-tyaatsaaattttccatgttgcac-3′) and *nad*2 (nad2up: 5′-ggagttg tntttagtacctctaa-3′ and nad2do: 5′-agtagtaacgayttntcacgatc-cat-3′) genes. In some cases alternative primers (nad4dov2: 5′-tccatgttgcactaagttacttacggangtatgcat-3′; nad4up2: 5′-aaat ttcartttgtggaaannnttcgatggcttcc-3′ or nad4up3: 5′-aggaagcc ttattattttggtgatcc-3′) were used to amplify the *nad*4 gene regions. In liverworts primers n5up (5′-gcaggnttttyggncgt tttct-3′) and nad5do (5′-aacatnrcaaaggcataatgata-3′) were designed to amplify the coding region of *nad*5, whereas alternative primers (K: 5′-atatgtctgaggatccgcatag-3′, L: 5′-aactttggccaaggatcctacaaa-3′) were mostly used to amplify the gene region in mosses. PCR amplification assays in total contained 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each primer and 0.5 U of GoTaq polymerase (Promega, Mannheim, Germany) or alternatively the PCR extender system (Taq-Pfu mixture, 5prime, Hamburg, Germany) using the respective buffers supplied by the manufacturers and double distilled water in a volume of 25 µl. The touch-down temperature profile used in the PCR assays included an initial denaturation at 94°C for 3 min, followed by 10 cycles, each with a denaturation step at 94°C for 30 s, 30 s annealing initially at 50°C, then decreased by 0.8 K in each cycle and a synthesis step at 72°C for 3 min. This was followed by 30 further amplification cycles of 30 s at 94°C, 30 s at 42°C, and 3 min at 72°C and a final elongation step of 7 min at 72°C to complete strand syntheses. PCR products were cloned into the pGEM-T Easy vector (Promega) before sequencing (Macrogen Inc., Seoul, South Korea). Sequences obtained by clone sequencing and sequences already available in the NCBI database (Annotations see Tables 1, 2) were assembled and aligned manually using MEGA 4.0.2 (Tamura et al. 2007). RNA editing sites were predicted using the RNA editing prediction and analysis computer tool PREPACT (Lenz et al. 2010) and verified by comparison of DNA and cDNA sequences for several species (Tables 1, 2). Different PREPACT options including the new feature "Commons" were used for prediction of editing sites.

Phylogenetic Tree Construction

Phylogenetic tree construction was based on concatenated organelle genome DNA data sets: *nad*5 including group I intron nad5i753g1, *nad*2 including group II intron nad2i156g2, the *nad*5–*nad*4 intergenic spacer, the cobi420g1 group I intron locus, *rbc*L and *rps*4 for mosses and *nad*5 with nad5i753g1, *nad*4 including group II intron nad4i548g2, *rbc*L and *rps*4 for liverworts, aligned manually in MEGA 4.0.2 (Tamura et al. 2007) and divided into partitions (mitochondrial coding sequences, mitochondrial spacer sequences, group I intron, group II intron, chloroplast sequences). Maximum likelihood phylogenies were calculated with Treefinder (Jobb et al. 2004) under GTR+G+I substitution model selected with Modeltest (Posada and Crandall 1998) and node support was determined based on 1,000 bootstrap resampling replicates. Bayesian analyses were performed with MrBayes (Ronquist and Huelsenbeck 2003) for 1 million generations with every 100th tree stored. Trees sampled before log stationarity was reached were discarded as burn-in.

The New PREPACT Feature "Commons"

The prediction of RNA editing sites in multiple DNA sequences by "Plant RNA Editing Prediction and Analysis Computer Tool" (PREPACT) was improved to allow a batch prediction of multiple DNA sequences against multiple reference sequences included in the alignment. Results for each single prediction are displayed in tables using LivePipe JavaScript framework (livepipe.net). In an additional table called "commons" the prediction for each query is summarized to show the amount of overlapping predictions against all reference sequences. In these tables the number of predicted sites per reference and the overall number of sites supported by a user-defined percentage of reference sites is given, too. These data can be downloaded for spreadsheet analysis from the WWW interface (www.prepact.de).

Identification of E–DYW Domain Extensions of PPR Protein Genes on DNA Level and Consensus Creation

Degenerated primers F (5′-gshtaygtdytbhtrtcmaacatwta-3′) and R (5′-tyaccartartcnctacaagaaca-3′) were designed based on available partial carboxyterminal E/E+/DYW domain sequences from bryophytes (Rüdinger et al. 2008) and used to amplify the C-terminal part of DYW-type PPR genes. PCR amplification assays with ingredients as described above included 3 min initial denaturation at 94°C followed by 10 cycles each with 30 s denaturation at 94°C, 30 s annealing at 45°C to 35°C (decreasing 1 K/cycle) and 3 min synthesis at 72°C, additionally 30 cycles with annealing at 35°C and a final step of synthesis for 7 min at 72°C were performed. PCR products were cloned into the pGEM-T Easy vector and commercially sequenced (Macrogen Inc., Seoul, South Korea). The cloning approach was tested with random sequencing of 30 clones of *Funaria*

Table 1 Liverwort taxon sampling and RNA editing in *nad*5 and *nad*4 gene regions

Liverworts	*nad*5 C-U								*nad*4 C-U								DYW-type PPR genes		
Species	*P. patens* CDS	*C. vulgaris*	*A. thaliana* CDS	*M. polymorpha*	Commons	cDNA confirmation	coverage (bp)	Accession number	*P. patens* CDS	*C. vulgaris*	*A. thaliana* CDS	*M. polymorpha*	Commons	cDNA confirmation	coverage (bp)	Accession number	PCR amplification	Number of diff. (c)DNA seqs	Accession numbers
Haplomitriids																			
Apotreubia nana	16	13	15	17	11		1104	EU519192	16	14	17	15	11		1311	JF451366			
Treubia lacunosa	2	1	2	1	0		1104	GU188738	3	2	5	1	0		1311	JF451367			
Treubia pygmaea	3	2	3	2	1		1104	JF451415	3	2	5	1	0		1311	JF451368			
Haplomitrium mnioides	**52**	**50**	**53**	**54**	**48**	**56**	**1104**	**AJ409111**	51	50	49	55	41		1311	DQ267608	yes	40	EU495454-70/EU495560-87
Marchantiids																			
Asterella grollei/ blumeana	1	0	2	0	0		1104	DQ268909	2	2	4	0	0		1311	JF451370	no		
Corsinia coriandra	**1**	**0**	**2**	**0**	**0**	**0**	**1104**	**AJ001034**	2	2	4	0	0		1311	JF451373	no		
Conocephalum conicum	1	0	2	0	0		1104	EU519188	2	2	4	0	0		1311	JF451372	no		
Oxymitra incrassata	1	0	2	0	0		1104	EU515190	2	2	4	0	0		1311	JF451376	no		
Riccia huebeneriana/ breidleri	1	0	1	0	0		1104	DQ268957	2	2	4	0	0		1311	JF451378	no		
Riccia fluitans	1	0	2	0	0		1104	DQ268956	*2*	*2*	*4*	*0*	*0*		*1230*	*AJ310802*	no		
Ricciocarpos natans	**1**	**0**	**2**	**0**	**0**	**0**	**1104**	**AJ001032**	2	2	4	0	0		1311	JF451379	no		
Monosolenium tenerum	1	0	2	0	0		1104	DQ268944	2	2	4	0	0		1311	JF451375	no		
Targionia hypophylla	*1*	*0*	*2*	*0*	*0*		*1092*	*DQ268964*	*2*	*2*	*4*	*0*	*0*		*1230*	*JF451382*			
Bucegia romanica	1	0	2	0	0		1104	AJ001031	2	2	4	0	0		1311	JF451371	no		
Marchantia polymorpha	1	0	2	0	0		1104	NC001660	2	2	4	0	0		1311	NC001660	no		
Lunularia cruciata	**1**	**0**	**2**	**0**	**0**	**0**	**1104**	**AJ001002**	***2***	***2***	***4***	***0***	***0***	***0***	***1230***	***AJ310803***	no		
Sphaerocarpos donnellii	1	0	2	0	0		1104	AJ001033	2	2	4	0	0		1311	JF451381			
Blasia pusilla	1	0	2	0	0		1104	EU519187	2	1	3	0	0		1311	JF451369	no		
Jungermanniids																			
Symphyogyna hymenophyllum/ brongniartii	*16*	*15*	*16*	*15*	*13*		*981*	*AY688772*	17	17	19	21	12		1311	JF451411			
Symphyogyna brasiliensis	*12*	*10*	*11*	*12*	*11*		*749*	*JF451432*	16	15	17	20	11		1311	JF451410			
Fossombronia pusilla	**14**	**12**	**14**	**14**	**12**	**14**	**1104**	**JN204577**	**30 (14)**	**27 (13)**	**26 (11)**	**31 (15)**	**20 (9)**	**15**	**1311**	**JF451407**	yes	6	EU495593-98
Fossombronia alaskana/angulosa	*8*	*7*	*8*	*8*	*7*		*707*	*AY688750*	34	31	30	35	24		1311	JF451406	yes	1	EU495494
Noteroclada confluens	13	12	11	13	10		1104	AJ622816	*24*	*22*	*24*	*26*	*18*		*1291*	*JF451408*			
Pellia cf endiviifolia	**22**	**18**	**21**	**21**	**17**	**21**	**1104**	**JF451431**	***22***	***19***	***20***	***23***	***16***	***22***	***1070***	***JF451409***	yes	4	EU495511-14
Calycularia crispula	**17**	**14**	**17**	**16**	**14**	**16**	**1104**	**EU519191**	**18**	**17**	**16**	**20**	**13**	**19**	**1311**	**JF451405**			
Pleurozia purpurea	4	3	5	3	3		1104	NC013444	5	4	5	4	0		1311	NC_013444			
Metzgeria conjugata/ furcata	12	11	12	11	10	11	1104	AJ000703	14	14	14	16	8		1311	JF451414	yes	6	EU495503-04/EU495610-13
Apometzgeria pubescens/ frontipilis	11	10	11	10	9		1104	DQ268906	14	13	14	16	7		1311	JF451413			
Aneura pinguis	*12*	*11*	*9*	*11*	*8*		*880*	*AY688744*	36	32	30	39	22		1311	JF451412	yes		
Ptilidium pulcherrimum	6	5	7	5	5		1104	JF451420	10	7	9	9	3		1311	JF451387	yes	4	EU495518-19/21-22
Porella navicularis/ platyphylla	*4*	*3*	*4*	*3*	*3*		*681*	*AY688767*	7	6	8	6	3		1311	JF451386	yes	4	JN204520-23
Lepidogyna hodgsoniae	**7**	**6**	**7**	**6**	**5**	**5**	**1104**	**JF451419**	**15**	**14**	**15**	**14**	**9**	**13**	**1311**	**JF451385**	yes	9	EU495497/EU495599-5606
Frullania tamarisci	2	1	3	1	1		1104	JF451417	5	4	6	3	1		1311	JF451383	yes	1	JN204519
Radula complanata	*0*	*0*	*1*	*0*	*0*		*420*	*AY608311*	4	3	6	2	1		1311	JF451388			
Lejeunea cavifolia	2	1	3	1	1		1104	AJ000701	4	3	6	2	1		1311	JF451384	yes	9	JN204537-45
Mylia cons/taylorii	3	2	3	2	2		1104	JF451426	*7*	*7*	*8*	*5*	*4*		*1255*	*JF451399*			
Herbertus alpinus/sendtneri	*4*	*3*	*5*	*3*	*3*		*860*	*DQ268927*	6	5	7	4	2		1311	JF451396			
Trichocolea tomentella	5	4	6	4	4		1104	AJ000707	11	10	12	12	7		1311	JF451403	yes	1	EU495531
Plagiochila asplenioides	**13**	**12**	**14**	**13**	**12**	**13**	**1104**	**AJ000704**	***18***	***16***	***18***	***17***	***12***	***15***	***1230***	***JF451401***			
Lophocolea heterophylla	7	6	8	6	6		1104	DQ268932	17	16	17	15	12	14	1311	JF451398	yes	2	EU495607-08
Blepharostoma cons /trichophyllum	5	4	6	4	4		1104	JF451422	7	5	8	6	2		1311	JF451391			
Bazzania trilobata	11	9	12	10	9		1104	AJ622815	*17*	*16*	*16*	*18*	*11*		*1230*	*AJ310800*	yes	3	EU495474-76
Harpanthus flotovianus	3	2	4	2	2		1104	JF451425	9	7	9	9	4		1311	JF451395			
Nardia scalaris	*5*	*4*	*5*	*4*	*4*		*552*	*JF451427*	*12*	*11*	*14*	*11*	*8*		*1290*	*JF451400*	yes	6	EU495505-10
Gymnomitrion concinnatum	5	4	6	4	4		1104	JF451424	8	7	9	6	4		1311	JF451394			
Calypogeia muelleriana	7	6	8	6	6		1104	EU519191	8	7	9	7	4		1311	JF451392	yes	3	EU495477-79
Anthelia julacea	9	8	10	8	8		1104	JF451421	11	10	12	10	7		1311	JF451390	yes	3	EU495471/EU495588-89
Jamesoniella autumnalis	7	6	8	6	6		1104	AJ000700	13	10	11	11	6		1311	JF451397			
Tritomaria quinquedentata	7	5	8	6	5		1104	JF451429	6	3	4	4	0		1311	JF451404			
Scapania nemorea	**7**	**5**	**8**	**7**	**5**	**7**	**1104**	**AJ000706**	**6**	**2**	**4**	**4**	**0**	**1**	**1311**	**JF451402**	yes	4	EU495527-30
Diplophyllum albicans	9	7	9	8	6		1104	JF451423	9	5	6	7	2		1311	JF451393	yes	11	EU495480-84/87-88/90/93/591-92
Anastrophyllum michauxii	5	4	6	4	4		1104	AY688743	*7*	*4*	*5*	*5*	*1*		*1230*	*JF451389*	yes	2	EU495472-73

Table 1 continued

Listed are liverwort taxa, the corresponding database accessions of the *nad*5 and *nad*4 gene regions investigated and the numbers of putative RNA editing sites depending on reference taxon sequences (*Marchantia polymorpha*, *Chara vulgaris* and cDNA sequences of *Physcomitrella patens* and *Arabidopsis thaliana*). The respective fifth column under "commons" lists the RNA editing sites congruently predicted with all four reference taxa. *Numbers* are given in *italics* when regions shorter than the regular amplicon have been investigated. Exemplary verification of RNA editing sites on cDNA level is highlighted by *gray shading*. Two epithets are given where sequences from different species of a genus were used for *nad*5 and *nad*4, respectively. *Fossombronia* cDNA clones cover an only 591 bp region of *nad*4, potential editing site numbers in that region are shown in *brackets*. The right part of the table documents results for PCR amplification of the E–DYW amplicon with 'Number of diff. (c)DNA seqs' indicating the number of different DYW-type protein coding regions obtained in this or recent studies

hygrometrica PCR products which revealed three of its nine known DYW genes (Rüdinger et al. 2011b).

Bioinformatic Work and Statistical Analyses

Deduced protein sequences were aligned with MEGA 4.0.2 (Tamura et al. 2007) using the ClustalW algorithm and manually adjusted. Consensus sequences of the E, E+, and DYW domains were created using sequences obtained by clone sequencing and sequences already available in the NCBI database (Tables 1, 2) and displayed using the weblogo server at http://weblogo.berkeley.edu (Crooks et al. 2004). Phylogenetic analyses were conducted in MEGA 4.0.2 (Tamura et al. 2007). Mathematical simulations for identification of differing numbers of different genes within a limited random clone sampling of a gene family were conducted using R (R Development Core Team 2011) and displayed with the graphic package ggplot2 as shown in Supplementary Fig. 1 (Wickham 2009). The Fisher's exact test, which assesses the likelihood that two different subsets are equal, was used to test for statistical significance of different DYW population diversities.

Results

Amending PREPACT for Stringent Editing Site Predictions

RNA editing in plant mitochondrial genes can be predicted quite reliably by comparison with homologous genes in non-editing species like the marchantiid liverwort *Marchantia polymorpha* or green algae or with known cDNAs in editing taxa. Among several other features, the recently developed PREPACT allows automatic prediction of RNA editing sites in multiple sequence alignments or in full organelle genomes (Lenz et al. 2010). The latter option has recently been used successfully to identify candidate RNA editing sites even in the mtDNA of a phylogenetic distant protist, which could subsequently be confirmed (Knoop and Rüdinger 2010; Rüdinger et al. 2011a). However, RNA editing prediction with this strategy frequently proves to be too sensitive and identifies false positive candidate sites when based on individual reference sequences alone. Consequently, we have now added new options to PREPACT, which allow simultaneous inclusion of multiple reference sequences to identify intersections of independently predicted editing sites ("commons") for more stringent prognoses.

We here use the recently proposed nomenclature to label RNA editing sites (Rüdinger et al. 2009). Briefly, editing site labels are composed of the name of the respective gene followed by an 'e' (for editing), the respective nucleotide introduced by the editing event (U), the nucleotide position in the transcript (with position 1 corresponding to the A of the AUG start codon) followed by the resulting amino acid change, e.g., nad5eU598RC.

RNA Editing Variability in Mosses and Liverworts

Mitochondrial gene sequences of *nad*4 and *nad*5 from 52 liverworts and gene sequences of *nad*2 and *nad*5 from 54 mosses were included in our analyses (Tables 1, 2; *nad* genes encode subunits of the NADH ubiquinone oxidoreductase, complex 1). RNA editing site numbers identified in the available mitochondrial transcriptomes of widely divergent plant species (*Physcomitrella patens*, *Funaria hygrometrica*, *Selaginella moellendorffii, Oryza sativa*, *Silene noctiflora*, *Arabidopsis thaliana*, *Brassica napus*) were compared with RNA editing frequencies in *nad*2, *nad*4, and *nad*5 transcripts alone to test for their use in extrapolation from a limited transcript sample (Supplementary Table S1). This revealed that the three genes allow to extrapolate very reasonably from the limited gene sampling to total mitochondrial RNA editing numbers over a wide range of RNA editing frequencies in the different taxa (i.e., edited/coding nucleotides), ranging from 0.02% in *Funaria* (0.04% estimated) to 10.2% in *Selaginella* (12.7% estimated).

RNA editing sites were predicted with PREPACT using alternative reference sequences (the homologous gene sequences of *Marchantia polymorpha* and the alga *Chara vulgaris* and the corresponding cDNA sequences of *Arabidopsis thaliana* and *Physcomitrella patens*). Both in liverworts (Table 1) as well as in mosses (Table 2) the

Table 2 Moss taxon sampling and RNA editing in *nad*5 and *nad*2 gene regions

Mosses	*nad*5 C-U								*nad*2 C-U								DYW-type PPR genes		
Species	*P. patens* CDS	*C. vulgaris*	*A. thaliana* CDS	*M. polymorpha*	Commons	cDNA confirmation	coverage (bp)	Accession number	*P. patens* CDS	*C. vulgaris*	*A. thaliana* CDS	*M. polymorpha*	Commons	cDNA confirmation	coverage (bp)	Accession number	PCR amplification	Number of diff. DNA seqs	Accession numbers
Takakia lepidozioides	35	31	33	32	27		1104	AJ291553	26	22	24	25	22		1252	AJ299525	yes	26	JN204546-71
Sphagnum fallax	6	4	8	6	4		1104	AJ001225	2	1	3	2	1		1252	AJ299524	yes		
Andreaea nivalis	7	6	7	6	5		1104	AJ001226	4	4	6	5	4		1252	AJ299526			
Tetraphis pellucida	7	7	8	7	6		1104	AJ224855	1	0	3	1	0		1252	AJ299529			
Dawsonia superba/ spec	2	2	4	2	2		1104	AY908804	0	0	2	0	0		1252	EU095309			
Atrichum undulatum	**5**	**4**	**6**	**4**	**4**	**4**	**1104**	**AJ001229**	**2**	**2**	**4**	**2**	**2**	**2**	**1252**	**AJ299527**	yes	3	JN204526-28
Oligotrichum hercynicum/ parallelum	*0*	*0*	*1*	*0*	*0*		*274*	*EU095271*	*3*	*2*	*4*	*3*	*2*		*1216*	*EU095310*			
Pogonatum urnigerum	2	2	4	2	2	2	1104	AJ291554	0	0	2	0	0		1252	AJ299528	yes	4	JN204572-75
Buxbaumia aphylla	5	4	5	5	3		1104	AJ291555	2	0	3	2	0		1252	AJ299531			
Diphyscium sessile	1	1	3	1	1		1104	Z98972	2	2	5	2	2		1252	AJ299530			
Physcomitrium pyriforme/ lorentzii	*2*	*2*	*3*	*2*	*2*		*272*	*EU095280*	0	0	2	0	0		1252	EU095312	yes	5	JN204529-33
Physcomitrella patens	**2**	**2**	**4**	**2**	**2**	**2**	**1104**	**NC_007945**	**0**	**0**	**2**	**0**	**0**	**0**	**1252**	**NC_007945**	yes	10	O'Toole et al. 2008
Funaria hygrometrica	**2**	**2**	**4**	**2**	**2**	**2**	**1104**	**Z98959**	**0**	**0**	**2**	**0**	**0**	**0**	**1252**	**AJ299534**	yes	9	JF501595-603
Discelium nudum	*2*	*2*	*3*	*2*	*2*		*272*	*EU095281*	0	0	2	0	0		1252	AY908956			
Encalypta streptocarpa	3	3	5	3	3		1104	AJ291556	1	1	3	1	1		1252	AJ299533			
Bryobrittonia longipes	*2*	*2*	*3*	*2*	*2*		*275*	*EU095277*	0	0	2	0	0		1252	EU095311			
Chamaebryum pottioides	*2*	*2*	*3*	*2*	*2*		*252*	*FJ870750*	*1*	*1*	*2*	*1*	*0*		*1226*	*FJ870757*			
Oedipodiella australis	*3*	*3*	*4*	*3*	*3*		*748*	*FJ870752*	0	0	2	0	0		1252	FJ870759			
Gigaspermum repens	*1*	*1*	*2*	*1*	*1*		*270*	*FJ870751*	0	0	2	0	0		1252	FJ870758			
Cinclidotus riparius	2	2	4	2	2		1104	AJ291563	0	0	2	0	0		1252	AJ299545			
Pottia truncata	5	5	7	5	5		1104	Z98957	1	1	3	1	1		1252	AJ299543			
Tortula latifolia	3	3	5	3	3		1104	AJ291562	0	0	2	0	0		1252	AJ299544			
Ceratodon purpureus	**3**	**3**	**5**	**3**	**3**	**3**	**1104**	**Z98955**	**0**	**0**	**2**	**0**	**0**	**0**	**1252**	**AJ299538**			
Ditrichum cylindricum	2	2	4	2	2		1104	AJ291559	2	2	4	2	2		1252	AJ299539			
Schistostega pennata	**2**	**2**	**4**	**2**	**2**	**2**	**1104**	**AJ224856**	1	1	3	1	1		1252	AJ299546			
Fissidens cristatus	1	1	3	1	1		1104	Z98954	0	0	2	0	0		1252	AJ299541	yes	1	JN204524
Orthodicranum montanum	7	7	9	7	7		1104	AJ291558	2	2	4	2	2		1252	AJ299537			
Blindia acuta	*2*	*2*	*3*	*2*	*2*		*304*	*EU095286*	0	1	2	0	0		1252	AY908928			
Racomitrium lanuginosum	2	2	4	2	2		1104	AJ291558	0	1	2	0	0		1252	AJ299542			
Grimmia plagiopodia	5	4	7	5	4		1104	AY908919	1	2	3	1	1		1252	AY908919			
Coscinodon cribrosus/ calyptratus	*2*	*2*	*3*	*2*	*2*		*265*	*EU095283*	0	1	2	0	0		1252	AY908918	yes	1	EU495532
Leucobryum glaucum	7	7	8	6	6		1104	AJ291560	1	0	3	1	0		1252	AJ299540			
Timmiella anomala/ spec	*2*	*2*	*3*	*2*	*2*		*270*	*EU095293*	0	0	2	0	0		1252	EU095317			
Catoscopium nigritum	*2*	*2*	*3*	*2*	*2*		*274*	*FJ870753*	0	0	2	0	0		1252	FJ870760			
Timmia bavarica	6	6	8	5	5		1104	AJ409093	0	0	2	0	0		1252	AJ299532	yes	5	EU495533-34, JN204534-36
Timmia megapolitana/ austriaca	*3*	*3*	*4*	*2*	*2*		*368*	*EU095276*	0	0	2	0	0		1252	FJ870755			
Bartramia halleriana	2	2	4	2	2		1104	Z98961	0	0	2	0	0		1252	AJ299547			
Plagiopus oederi	2	2	4	2	2		1104	Z98962	0	0	2	0	0		1252	AJ299548			
Hedwigia ciliata	3	3	5	3	3		1104	Z98966	1	1	3	1	1		1252	AJ299554			
Rhacocarpus purpurascens	2	2	4	2	2		1104	Z98967	0	0	2	0	0		1252	AJ299555			
Mnium hornum	2	2	4	2	2		1104	AJ291567	0	0	2	0	0		1252	AJ299552			
Pohlia nutans	2	2	4	2	2		1104	AJ291565	0	0	2	0	0		1252	AJ299550			
Splachnum ampullaceum	1	1	3	1	1		1104	EU095308	1	1	3	1	1		1252	EU095318			
Ulota crispa	**3**	**3**	**4**	**3**	**3**	**3**	**1104**	**AJ291568**	**0**	**0**	**2**	**0**	**0**	**0**	**1252**	**AJ299553**			
Orthodontium lineare	7	6	8	6	6		1104	AJ291566	1	1	2	1	0		1252	AJ299551	yes	1	JN204525
Aulacomnium androgynum	4	4	6	4	4		1104	AJ291564	4	4	5	4	3		1252	AJ299549			
Fontinalis antipyretica	3	3	4	3	2		1104	AJ291570	4	4	6	4	4		1252	AJ299558			
Herzogiella seligeri	3	3	5	3	3		1104	AJ291573	2	2	4	2	2		1252	AJ299561			
Homalia trichomanoides	**5**	**4**	**6**	**4**	**4**	**5**	**1104**	**JN204576**	**1**	**1**	**3**	**1**	**1**	**1**	**1252**	**AJ299557**			
Thamnobryum alopecurum	9	8	10	8	8		1104	AJ291571	0	0	2	0	0		1252	AJ299559			
Pterogonium gracile	6	5	6	5	4		1104	Z98968	0	0	2	0	0		1252	AJ299556			
Tomentypnum nitens	4	3	5	3	3		1104	AJ291572	2	1	4	2	1		1252	AJ299560			
Scorpidium scorpioides	7	6	7	6	5		1104	AJ291575	2	2	4	2	2		1252	AJ299563			
Hygrohypnum ochraceum	3	2	3	2	2		1104	AJ291574	2	1	4	2	1		1252	AJ299562			

Listed are moss taxa, the corresponding database accessions of the *nad*5 and *nad*2 gene regions investigated and the numbers of putative RNA editing sites depending on reference taxon sequences (*Marchantia polymorpha*, *Chara vulgaris* and cDNA sequences of *Physcomitrella patens* and *Arabidopsis thaliana*). The respective fifth column under "commons" lists the RNA editing sites congruently predicted with all four reference taxa. Numbers are given in *italics* when regions shorter than the regular amplicon have been investigated. Exemplary verification of RNA editing sites on cDNA level is highlighted by *gray shading*. Two epithets are given where sequences from different species of a genus were used for *nad*5 and *nad*2, respectively. The right part of the table documents results for PCR amplification of the E–DYW amplicon with 'Number of diff. DNA seqs' indicating the number of different DYW-type protein coding regions obtained in this or recent studies

numbers of predicted RNA editing sites differ widely in different taxa. Notably, the restricted taxon sampling of Steinhauser et al. (1999) for *nad*5 of only seven marchantiid (complex thalloid) liverworts is now extended to 14 taxa and the data sampling now includes *nad*4 as an independent second locus. For none of the marchantiid liverworts only a single site of RNA editing was predicted using coding regions of *nad*5 and *nad*4 of *Marchantia polymorpha* as a reference whereas up to four sites would be predicted using *Chara vulgaris* or the *Physcomitrella* or *Arabidopsis* cDNAs as references (Table 1). None of the ambiguously predicted sites using the phylogenetically more distant taxa was corroborated in exemplary cDNA analyses of *nad*5 in *Corsinia, Lunularia*, or *Ricciocarpos* (Steinhauser et al. 1999) or of *nad*4 in *Lunularia* (this study) in support for the new "commons" concept for more restrictive editing site prediction (Table 1).

In contrast to the marchantiid liverworts, editing sites were consistently predicted for the jungermanniid (i.e., leafy and simple thalloid) liverworts, even when using the new restrictive "commons" mode of prediction. Again, we wished to test predictions with exemplary cDNA sequencing, for which we selected six taxa. In particular, this included *Pellia* cf. *endivifolia* and *Calycularia crispula* with some 20 or more editing sites predicted for each gene (Fig. 1; Table 1). Sequencing on cDNA level for the investigated gene regions of *nad*5 and *nad*4 confirmed a total of 43 and 35 C-to-U editing events, respectively, for those two taxa. All sites predicted using the stringent "commons" prediction were confirmed and all additional sites were correctly predicted using the homologous liverwort *Marchantia* sequence as a reference. An example of editing prediction in *nad*5 based on graphic output from PREPACT is exemplarily shown in Fig. 1 for a sample of selected liverworts and mosses.

In general, false positive predictions using individual references mainly affect conservative amino acid exchanges (such as GCN alanine to GUN valine) potentially subject to editing (Supplementary Tables S2–S5). A potential editing event nad4eU1394AV erroneously predicted for *nad*4 in many jungermanniid taxa is a typical example that remained consistently unconfirmed in cDNAs.

Interestingly, *Pellia* and *Calycularia* are closely related genera (Fig. 2a) and share 13 editing sites in *nad*5 and 14 sites in *nad*4 (Fig. 1; Supplementary Tables S2, S3). Other closely related species like *Metzgeria furcata*, *Apometzgeria frontipilis* and the *Symphyogyna* species (Fig. 2a; Supplementary Tables S2, S3) also share the majority of their editing sites in both investigated gene regions. RNA editing frequencies differ widely among the jungermanniid liverworts, whereas the frequencies of RNA editing in the two analyzed genes of a given jungermannid species are rather similar in most cases, confirming that RNA editing is much more taxon-dependent than locus-dependent. As a single exception, in *Aneura pinguis* the editing site prediction in *nad5* is significantly lower than in *nad4*. Overall, a decrease of RNA editing frequency is apparent in diversification of the jungermannid liverworts with higher editing rates in the early-branching taxa (Fig. 2a).

In the haplomitriid liverworts, the earliest diverging clade of liverworts (Fig. 2a), the genera *Haplomitrium*, *Apotreubia*, and *Treubia* show the most extreme discrepancies in RNA editing frequencies. *Haplomitrium mnioides* has the highest degree of RNA editing in all liverworts with 56 (confirmed) editing sites in *nad*5 (1104 bp), whereas in *Treubia* only single editing events are predicted in *nad*5 and *nad*4, respectively.

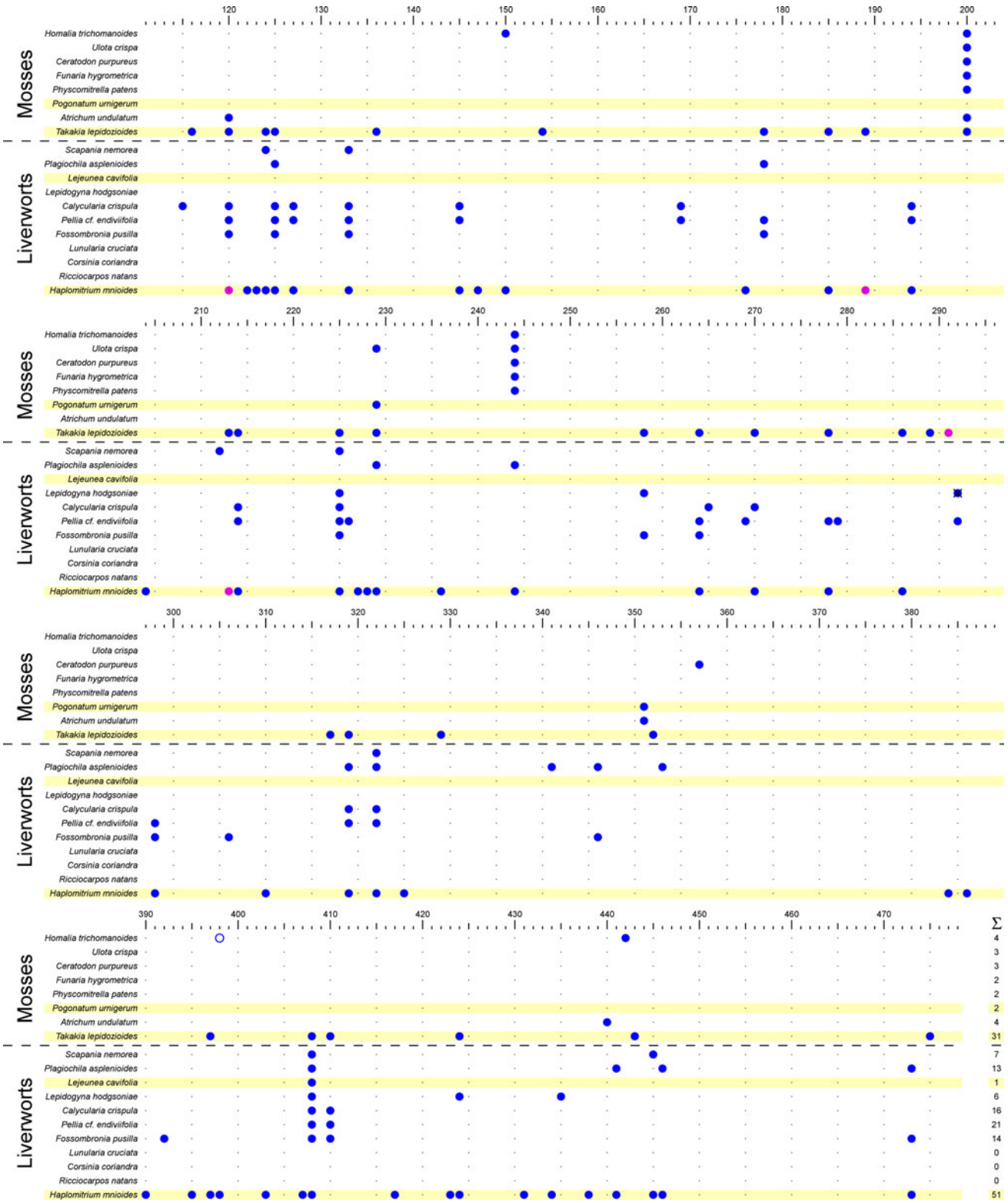
In mosses, RNA editing levels also vary between species, but two editing sites nad5eU598RC and nad5eU730RW seem to be highly conserved in nearly all arthrodontous mosses (Fig. 1; Supplementary Table S4). Overall editing frequencies are lower than in the jungermanniid liverworts (Table 2; Fig. 3a) with even less editing sites in *nad*2 than in *nad*5. A unique exception is *Takakia lepidozioides*, for which 27 editing sites in the *nad5* gene and 22 editing sites in the *nad2* gene are predicted using the commons feature of PREPACT. Very similar to *Haplomitrium* among the liverworts, *Takakia* represents an early branch in the phylogeny of its clade.

The Detailed Inventory of RNA Editing Sites

A detailed listing of editing sites following the recently proposed nomenclature (Rüdinger et al. 2009) is given in Supplementary Tables S2–S5. The few editing sites which were not confirmed in cDNA analyses of related taxa were omitted from the listings. For the exemplary cDNA analyses (Tables 1, 2) at least three clones per gene region were sequenced to increase chances of identifying potentially partial and/or silent RNA editing events. In liverworts generally one to two editing sites per taxon were found to be partially edited. In mosses, except for one single editing site in *Homalia trichomanoides*, all editing sites were completely edited in all sequenced cDNA clones. Silent RNA editing sites which do not change the amino acid sequence were rarely observed. No single reverse U-to-C RNA editing site was detected in any of the moss or liverwort cDNAs.

RNA Editing Frequencies Correlate with DYW Domain Diversity in Selected Species

With the observation of highly variable editing frequencies both in liverworts and in mosses, the question arises whether the number of DYW-type PPR genes, recently



identified as RNA editing recognition factors, varies correspondingly. In a previous study we already showed a high diversity of E–DYW domains in *Haplomitrium mnioides*, the taxon with apparently the highest RNA editing rate among liverworts (Rüdinger et al. 2008). Targeting a PCR amplicon encompassing the E–DYW domain continuity with degenerated primers, we now wished to check the diversity of the corresponding gene family in the

◀ **Fig. 1** RNA editing site prediction is shown for the mitochondrial *nad5* gene amplicon encompassing 368 codons using the graphical output of PREPACT (with *Marchantia* as the reference sequence) for a selection of eight mosses (*upper part*) and 11 liverworts (*lower part*). The sampling includes all species for which RNA editing was checked on cDNA level and pairs of taxa with high versus low editing among the mosses (*Takakia* and *Pogonatum*) and the liverworts (*Haplomitrium* and *Lejeunea*) also investigated for DYW gene diversity (*yellow shading*). *Blue circles* indicate codon sense changes after single C-to-U editing and *purple circles* indicate codon changes double C-to-U editings (multistep). The *blue open circle* indicates one editing site (confirmed) in *Homalia trichomanoides* (codon 398) that was predicted with *Physcomitrella patens,* but not with *Marchantia polymorpha* as reference. Conversely, one predicted editing site in *Lepidogyna hodgsoniae* (codon 292, crossed out) was not confirmed on cDNA level (Color figure online)

liverwort *Lejeunea cavifolia,* in which we now identified a particularly low amount of predicted RNA editing. Indeed, in a total of 30 E–DYW amplicon clones only nine different sequences could be identified (Fig. 2b) contrasting the diversity of 40 different among 51 E–DYW clones previously identified in *Haplomitrium*.

A fully congruent picture emerged for the mosses. 30 E–DYW amplicon clones of *Takakia lepidozioides,* the moss with the highest degree of putative RNA editing sites in mitochondria and likely also in chloroplasts (Sugita et al. 2006; Yura et al. 2008), resulted in 26 different E–DYW domain sequences. Conversely, in a total of 30 E–DYW amplicon clones of *Pogonatum urnigerum,* with only two editing sites predicted in the investigated gene regions, only four different E–DYW domain sequences could be identified (Fig. 3b), again demonstrating clearly that numbers of RNA editing sites and DYW gene diversity correlate.

The restricted clone samplings naturally cover only a fraction of the true DYW gene diversity in a given genome. Firstly, degenerated primers will bind preferentially to a subset of DYW gene family members. This was tested with *Funaria hygrometrica* DNA as a control, where we identified three of the nine known DYW genes (Rüdinger et al. 2011b) among 30 clones and—with the benefit of knowing primer target sites in this species—these turned out to be the DYW genes, where primer sequences fitted best. Nevertheless, the *Funaria* result matches very well with the only four DYW sequences identified in *Pogonatum* given that both moss species show only two editing sites in the sampled genes. Secondly, even if primer binding would be fully unbiased for DYW genes, theoretical mathematical modelling shows that a total genomic diversity of n different DYW genes requires a sampling of at least $2n$ clones for reliable estimates, which limits the approach for very high DYW gene numbers (see Supplementary Fig. 1). Conversely, the high diversity values of 26 or 40 different DYW domains in samplings of 30 or 51 clones in *Takakia* and *Haplomitrium*, respectively, would translate into minimally 50 or 60 and up to a few hundreds of DYW domains in their genomes within 95% confidence limits. To test, whether our observations from the clone samplings for the two species pairs of mosses and liverworts reflect true differences we used the Fisher's exact test, which reveals the likelihood that the two different observations may actually reflect equal true diversities. The likelihood of equal probabilities for the liverwort species pair *Haplomitrium/Lejeunea* (40/51 vs. 9/30) is 4×10^{-5} and for the moss species pair *Takakia/Pogonatum* (26/30 vs. 4/30) is 1.3×10^{-8}, therefore strongly rejecting hypotheses of equal probabilities in both species pairs.

Conservation of C-terminal Domain Additions Among Land Plant Clades

The C-terminal domain additions E/E+/DYW in more than 100 PLS-type PPR proteins of this type each are highly conserved among the dicot *Arabidopsis thaliana* and the monocot *Oryza sativa.* The moss *Physcomitrella patens* encodes only ten DYW genes in its nuclear genome and it seemed interesting to check for conservation of the domains in the now available wide sequence samplings of liverworts and mosses. The domain sequences obtained in this study were combined with those from previous studies (O'Toole et al. 2008; Rüdinger et al. 2008) resulting in a total of 119 E–DYW domain sequences from 19 different liverworts and of 65 E–DYW domain sequences from 10 mosses, which were used to derive liverwort- and moss-specific consensus sequences of the C-terminal domain extensions (Tables 1, 2). The weblogo profiles (Crooks et al. 2004) and consensus sequences derived for liverworts and mosses did not show any characteristic differences (not shown), suggesting that no functional adaptation occurred after this earliest phylogenetic split of plant evolution. Therefore, liverwort and moss sequence alignments were combined to create a collective bryophyte data set, which was used to derive a joint weblogo conservation plot (Fig. 4). The comparison with the corresponding consensus sequence profiles of the E, E+, and DYW domains of *Arabidopsis thaliana* (Lurin et al. 2004) showed no single significant change in conservation patterns across all three C-terminal domains, suggesting no significant functional adaptation in land plant phylogeny. Only two conserved amino acid sequence stretches in the E domain and two in the DYW domain seem to be under relaxed conservation in *Arabidopsis thaliana* sequences (Fig. 4). Comparison with a consensus sequence (not shown) of 77 different E–DYW domains of the lycophyte *Selaginella moellendorffii* (http://wiki.genomics.purdue.edu-/index.php-/PPR_gene_family) likewise shows a comparably high conservation of these protein domains in the early-branching vascular plant lineage.

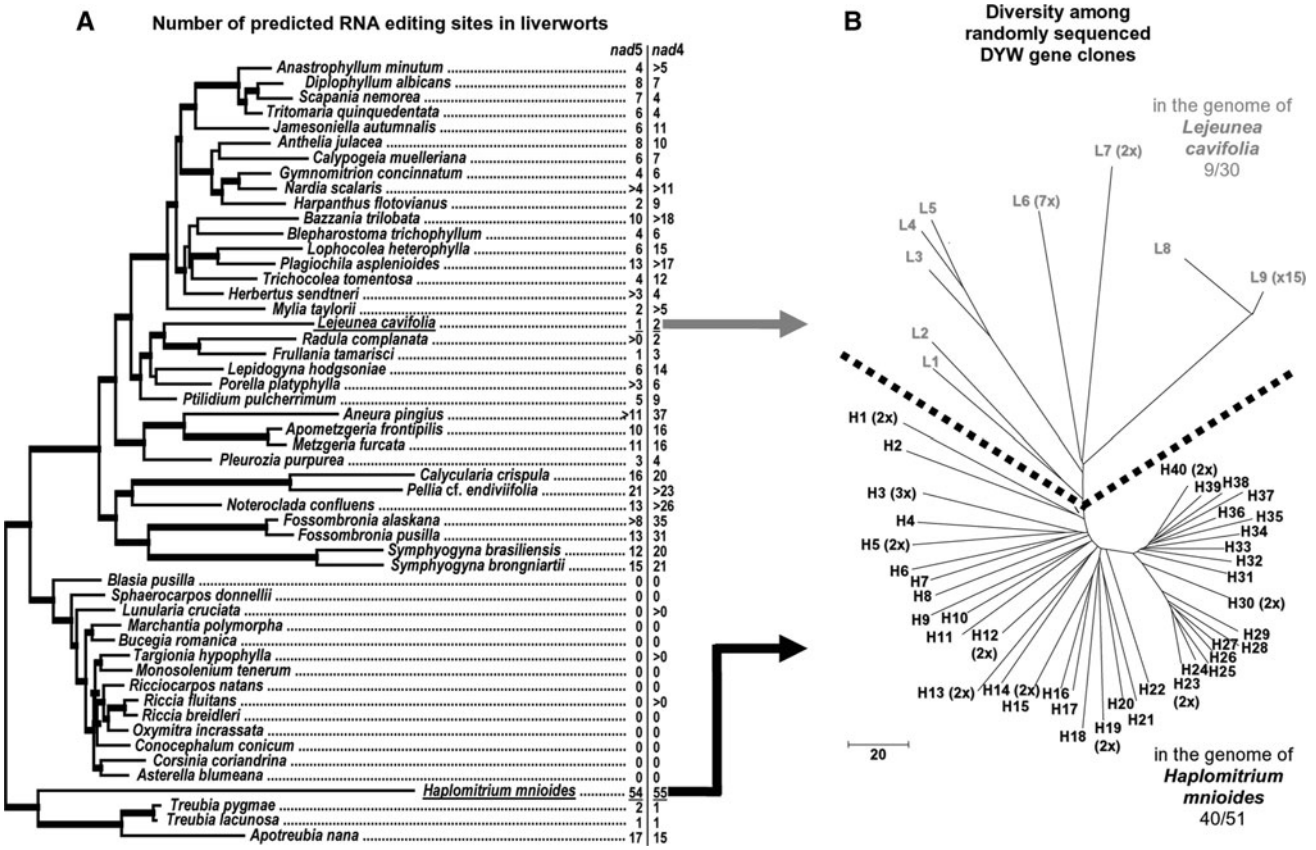


Fig. 2 **a** Liverwort phylogeny based on a concatenated data set of *nad*5, *nad*4, *rbc*L, and *rps*4. *Thickened internode lines* indicate significant Bayesian probability support (>0.96). *Numbers next to taxa* indicate putative editing sites in the investigated gene regions (*nad*5 331–1434 bp, *nad*4 130–1440 bp of the coding regions), here shown with *Marchantia polymorpha* used as reference taxon. The '>' symbol is added to numbers of RNA editing sites when derived from a smaller region than the regular amplicon. **b** For comparison of DYW gene diversity of *Haplomitrium mnioides* (highest predicted editing frequency) and *Lejeunea cavifolia* (low predicted editing frequency) 51 and 30 clones were sequenced, respectively. Nine different E–DYW domain sequences were identified in *Lejeunea* and 40 different sequences in *Haplomitrium*, which all cluster species-specifically in a simple Neighbor-Joining tree using uncorrected (*p*) distances

Discussion

Considering the variability of RNA editing frequencies among the basal-most land plant clades and across the wide plant phylogeny at large, the mechanism to correct genetic information in organellar genomes on RNA level seems to be a highly dynamic evolutionary process. The most obvious and suggestive idea explaining the origin of RNA editing in land plants in the first place is the adaptation to higher exposure of mutational stress such as UV light in the new terrestrial environment (Fujii and Small 2011; Tillich et al. 2006). Strikingly, our analyses found extraordinarily high amounts of RNA editing in the respective early-branching taxa, *Takakia* and *Haplomitrium*, within their respective clades. A tendency to reduce RNA editing is then obvious both in the diversification of mosses and liverworts. It is certainly tempting to speculate that a potential reduction in RNA editing may reflect reductions of mutational pressure after other adaptations to land plant life have come into existence, such as improved protective plant surfaces or DNA repair mechanisms in the organelles (Maier et al. 2008). This would all the more emphasize the "living fossil" roles of the two enigmatic genera *Takakia* and *Haplomitrium* in their phylogenetically isolated positions, which have only recently been elucidated with molecular data (Crandall-Stotler et al. 2005; Davis 2004; Volkmar and Knoop 2010). However, it should not be overlooked that other early-branching taxa both in the mosses (e.g., *Sphagnum*) and in the liverworts (e.g., *Apotreubia* and *Treubia*) show massively lower or even extraordinarily low RNA editing rates leaving open that RNA editing in *Takakia* and *Haplomitrium* has instead increased independently in frequency. One clear point of evidence, however, in support for loss over independent gain scenarios is the striking overall absence of RNA editing in marchantiid liverworts, which we now find strongly supported from the much extended sequence and taxon sampling reported here. While the high numbers of

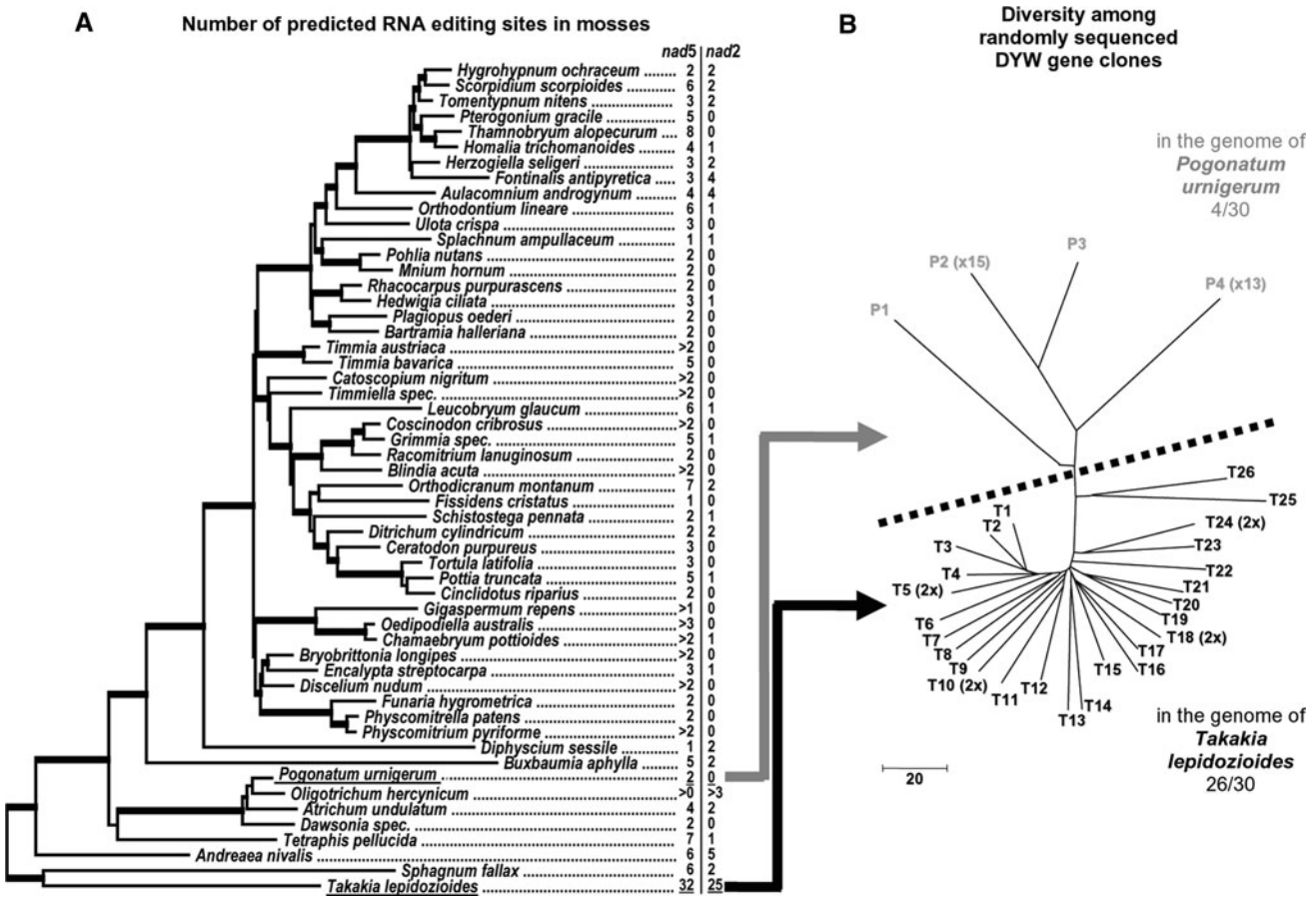


Fig. 3 **a** Moss phylogeny based on a concatenated data set of *nad*5, *nad*2, *nad*5-*nad*4 spacer, cobi420, *rbc*L and *rps*4. *Thickened internode lines* indicate significant Bayesian probability support (>0.96). *Numbers next to taxa* indicate putative editing sites in the investigated gene regions (*nad*5 331–1434 bp, *nad*2 99–1350 bp of the coding regions), here shown with *Marchantia polymorpha* used as reference taxon. The '>' symbol is added to numbers of RNA editing sites when derived from a smaller region than the regular amplicon. **b** For comparison of DYW gene diversity of *Takakia lepidozioides* (highest predicted editing frequency) and *Pogonatum urnigerum* (low predicted editing frequency) 30 clones were sequenced for each species. Four different E–DYW domain sequences were identified in *Pogonatum* and 26 different sequences in *Takakia*, which all cluster species-specifically in a simple Neighbor-Joining tree using uncorrected (*p*) distances

RNA editing events in early-branching lineages both among mosses and the liverworts might argue for highly frequent RNA editing as a plesiomorphic character state among land plants, the detailed RNA editing patterns support the idea only weakly. Out of 71 sites of editing in the *nad*5 gene, only 11 are shared between *Haplomitrium* and *Takakia*. Conceptually, the loss of editing sites can be expected to occur much more easily than the emergence of new sites, which require appropriate novel editing factors. It has indeed been shown conclusively that losses of editing sites occur faster than gains in angiosperm mitochondria (Mower 2007; Shields and Wolfe 1997) and chloroplasts (Tillich et al. 2006). However, a different picture may emerge on larger evolutionary timescales as reflected by the dramatic increase of RNA editing frequencies in ancient clades such as the hornworts and lycophytes (Duff 2006; Grewe et al. 2011; Hecht et al. 2011; Sper-Whitis et al. 1996). In those lineages, and possibly in yet more ancient isolated lineages like *Haplomitrium* and *Takakia*, evolutionary forces reshaping nuclear genomes by massive waves of gene duplications may allow for neo-functionalizations of editing factors addressing new editing sites, which may appear faster than others are lost.

Independent of the above gain or loss scenarios, the diversity of the nuclear DYW genes correlates well with the RNA editing frequencies. Extrapolating our data, hundreds of such genes must be expected in the nuclear genomes of *Takakia* and *Haplomitrium.* We here observed a clustering of the E–DYW domain sequences of one species (Figs. 2b, 3b) instead of a clustering of potential functional orthologues from different taxa. This is in stark contrast to the unequivocal identification of DYW protein orthologue pairs in the two closely related mosses *Physcomitrella* and *Funaria* that was recently observed

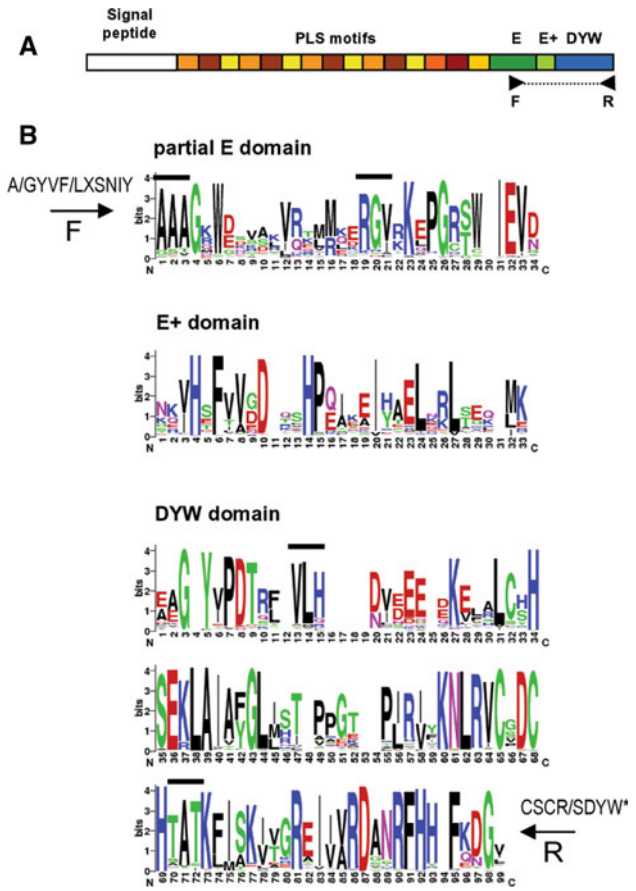


Fig. 4 **a** Typical motif structure of a PLS-type PPR protein, characterized by variable numbers (2–26) of alternating PPR motif repeats: canonical 35 aa P motifs (*orange*) and "long" L (*brown*) and "short" S (*yellow*) variants and optional carboxyterminal domain extensions E, E+, and DYW. Oligonucleotide primers F and R used in this study target the conserved E–DYW domain continuity (*black arrowheads*) are indicated. **b** Weblogo sequence conservation plot for 119 liverwort and 65 moss E–DYW amplicon sequences (conserved sequence motifs of primer binding sites indicated) obtained from their fused alignments using the WEBLOGO service at http://weblogo.berkeley.edu/logo.cgi. *Blank positions* indicate rare insertions of amino acids in individual sequences. Comparison with the corresponding conservation plot of the 87 *Arabidopsis thaliana* E/E+/DYW gene sequences shows largely identical conservations of highly conserved (bit score of at least 2) positions with only four sequence stretches (AAA and RGV in the E domain and VLH and TAT in the DYW domain) showing relaxed conservation in *Arabidopsis thaliana* marked with *bold lines* above (Color figure online)

(Rüdinger et al. 2011b) and obviously related to the much wider phylogenetic divergence between the respective pairs of high and low frequency editing liverworts and mosses investigated here (Figs. 2, 3). Clustering of DYW gene paralogs within one species would certainly be expected as the result of multiple gene duplications concomitant with vastly (and rapidly) increased frequencies of RNA editing in the organelles of taxa such as *Haplomitrium*, *Takakia*, the hornworts or the lycophytes (Duff 2006; Grewe et al. 2011; Hecht et al. 2011; Sper-Whitis et al. 1996).

The here documented conservation of the C-terminal domain extensions E–DYW across the entire land plant phylogeny with the domain sequences of liverworts and mosses being highly similar in conservation patterns to the vascular plant homologues is striking. This consequently suggests highly conserved functions of these protein domains, which, however, are still elusive. The DYW domain in particular has been suggested to play an important role in RNA editing given its weak similarity to cytidine deaminases (Salone et al. 2007) and bioinformatic protein structure analyses have indeed found strong support for this (Iyer et al. 2011). However, this assumption could not be confirmed in vitro or in vivo (Nakamura and Sugita 2008; Okuda et al. 2009) and—besides several DYW-type proteins identified as RNA editing factors (Kim et al. 2009; Ohtani et al. 2010; Okuda et al. 2009; Robbins et al. 2009; Rüdinger et al. 2011b; Tasaki et al. 2010; Verbitskiy et al. 2011; Zehrmann et al. 2009; Zhou et al. 2008), numerous other PLS proteins of the E/E+ type lacking the carboxy-terminal DYW domain were also identified as RNA editing recognition factors in chloroplasts (Chateigner-Boutin et al. 2011; Kotera et al. 2005; Okuda et al. 2007) and mitochondria (Sung et al. 2010; Takenaka 2010; Takenaka et al. 2010; Tang et al. 2010; Verbitskiy et al. 2010, 2011). One DYW-type PPR protein (Phypa_154890; PPR_43) in *Physcomitrella patens* has recently been shown to be relevant for splicing of mitochondrial group II intron cox1i732g2 and not for RNA editing (Ichinose et al. 2011). Interestingly, however, this is the single *P. patens* DYW-type PPR protein lacking highly conserved amino acid positions assumed to be important for deaminase functionality (Rüdinger et al. 2011b). Moreover, the degenerated DYW domain and the preceding E and E+ domains could be deleted without loss of splicing function (Ichinose et al. 2011). Aside from the observations reported here, the link between the DYW domain and the cytidine deamination type of RNA editing is furthermore strengthened from recent findings in the heterolobosean protist *Naegleria gruberi*. After the first discovery of DYW-type PPR protein genes outside of land plants in the nuclear genome of *Naegleria* (Knoop and Rüdinger 2010), we could recently identify two events of C-to-U RNA editing in its mitochondrial transcriptome (Rüdinger et al. 2011a). The huge phylogenetic distance of some 1.5 billion years separating *Naegleria* from land plants raises yet more questions on why and how RNA editing came into existence and is maintained in certain lineages of life.

Acknowledgments We are grateful to Prof. Jan-Peter Frahm and Dr. Boon-Chuan Ho (University Bonn) for their help to collect bryophyte plant material and to Prof. Wolfgang Alt and Martin Bock

(University Bonn) for doing initial MATLAB simulations and help with the Fisher's exact test. Work in the authors' laboratory on RNA editing and DYW proteins is supported by DFG Grant Kn411/7, work on liverwort and moss phylogeny was supported by the DFG Grant Kn411/6. Finally, we would like to thank two anonymous reviewers for their helpful suggestions for changes in the manuscript and Julia Hecht for comments on text and language.

References

Beckert S, Steinhauser S, Muhle H, Knoop V (1998) A molecular phylogeny of bryophytes on nucleotide sequences of the mitochondrial nad5 gene. Plant Syst Evol 218:179–192

Chateigner-Boutin AL, des Francs-Small CC, Delannoy E, Kahlau S, Tanz SK, de Longevialle AF, Fujii S, Small I (2011) OTP70 is a pentatricopeptide repeat protein of the E subgroup involved in splicing of the plastid transcript rpoC1. Plant J 65:532–542

Covello PS, Gray MW (1989) RNA editing in plant mitochondria. Nature 341:662–666

Covello PS, Gray MW (1993) On the evolution of RNA editing. Trends Genet 9:265–268

Crandall-Stotler BJ, Forrest LL, Stotler RE (2005) Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). Taxon 54:299–316

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

Davis CE (2004) A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). In: Goffinet B, Hollowell V, Magill R (eds) Monographs in systematic botany. Missouri Botanical Garden Press, St Louis, pp 61–86

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

Duff RJ (2006) Divergent RNA editing frequencies in hornwort mitochondrial nad5 sequences. Gene 366:285–291

Freyer R, Kiefer-Meyer MC, Kossel H (1997) Occurrence of plastid RNA editing in all major lineages of land plants. Proc Natl Acad Sci USA 94:6285–6290

Fujii S, Small I (2011) The evolution of RNA editing and pentatricopeptide repeat genes. New Phytol 191:37–47

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

Grewe F, Herres S, Viehover P, Polsakiewicz M, Weisshaar B, Knoop V (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte Isoetes engelmannii. Nucleic Acids Res 39:2890–2902

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 24:1068–1074

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

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

Hecht J, Grewe F, Knoop V (2011) 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. Genome Biol Evol 3:344–358

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

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

Hoch B, Maier RM, Appel K, Igloi GL, Kossel H (1991) Editing of a chloroplast messenger-RNA by creation of an initiation codon. Nature 353:178–180

Ichinose M, Tasaki E, Sugita C, Sugita M (2011) A PPR-DYW protein is required for splicing of a group II intron of cox1 pre-mRNA in *Physcomitrella patens*. Plant J. doi:10.1111/j.1365-313X.2011.04869.x

Inada M, Sasaki T, Yukawa M, Tsudzuki T, Sugiura M (2004) A systematic search for RNA editing sites in pea chloroplasts: an editing event causes diversification from the evolutionarily conserved amino acid sequence. Plant Cell Physiol 45:1615–1622

Iyer LM, Zhang D, Rogozin IB, Aravind L (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. Nucleic Acids Res 39:9473–9497

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4:18

Jobson RW, Qiu YL (2008) Did RNA editing in plant organellar genomes originate under natural selection or through genetic drift? Biol Direct 3:43

Kim SR, Yang JI, Moon S, Ryu CH, An K, Kim KM, Yim J, An G (2009) Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. Plant J 59:738–749

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 584:4287–4291

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

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

Lamattina L, Weil JH, Grienenberger JM (1989) RNA editing at a splicing site of NADH dehydrogenase subunit IV gene transcript in wheat mitochondria. FEBS Lett 258:79–83

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

Lurin C, Andres C, Aubourg S, Bellaoui M, Bitton F, Bruyere 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 16:2089–2103

Maier RM, Neckermann K, Hoch B, Akhmedov NB, Kossel H (1992) Identification of editing positions in the ndhB transcript from maize chloroplasts reveals sequence similarities between editing sites of chloroplasts and plant mitochondria. Nucleic Acids Res 20:6189–6194

Maier UG, Bozarth A, Funk HT, Zauner S, Rensing SA, Schmitz-Linneweber C, Borner T, Tillich M (2008) Complex chloroplast RNA metabolism: just debugging the genetic programme? BMC Biol 6:36

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

Mower JP (2007) Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. Mol Biol Evol 25:52–61

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

Nakamura T, Sugita M (2008) A conserved DYW domain of the pentatricopeptide repeat protein possesses a novel endoribonuclease activity. FEBS Lett 582:4163–4168

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 Genomics 268:434–445

Ohtani S, Ichinose M, Tasaki E, Aoki Y, Komura Y, Sugita M (2010) Targeted gene disruption identifies three PPR-DYW proteins involved in RNA editing for five editing sites of the moss mitochondrial transcripts. Plant Cell Physiol 51:1942–1949

Ohyama K, Takemura M, Oda K, Fukuzawa H, Kohchi T, Nakayama S, Ishizaki K, Fujisawa M, Yamato K (2009) Gene content, organization and molecular evolution of plant organellar genomes and sex chromosomes: insights from the case of the liverwort *Marchantia* polymorpha. Proc Jpn Acad Ser B Phys Biol Sci 85:108–124

Okuda K, Myouga F, Motohashi R, Shinozaki K, Shikanai T (2007) Conserved domain structure of pentatricopeptide repeat proteins involved in chloroplast RNA editing. Proc Natl Acad Sci USA 104:8178–8183

Okuda K, Chateigner-Boutin AL, Nakamura T, Delannoy E, Sugita M, Myouga F, Motohashi R, Shinozaki K, Small I, Shikanai T (2009) Pentatricopeptide repeat proteins with the DYW motif have distinct molecular functions in RNA editing and RNA cleavage in *Arabidopsis* chloroplasts. Plant Cell 21:146–156

O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I (2008) On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 25:1120–1128

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Robbins JC, Heller WP, Hanson MR (2009) A comparative genomics approach identifies a PPR-DYW protein that is essential for C-to-U editing of the *Arabidopsis* chloroplast accD transcript. RNA 15:1142–1153

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rüdinger M, Polsakiewicz M, Knoop V (2008) Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. Mol Biol Evol 25:1405–1414

Rüdinger M, Funk HT, Rensing SA, Maier UG, Knoop V (2009) RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. Mol Genet Genomics 281:473–481

Rüdinger M, Fritz-Laylin L, Polsakiewicz M, Knoop V (2011a) Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. RNA 17:2058–2062

Rüdinger M, Szövényi P, Rensing SA, Knoop V (2011b) Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses *Physcomitrella patens* and *Funaria hygrometrica*. Plant J 67:370–380

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 581:4132–4138

Sasaki T, Yukawa Y, Miyamoto T, Obokata J, Sugiura M (2003) Identification of RNA editing sites in chloroplast transcripts from the maternal and paternal progenitors of tobacco (*Nicotiana tabacum*): comparative analysis shows the involvement of distinct trans-factors for ndhB editing. Mol Biol Evol 20: 1028–1035

Shields DC, Wolfe KH (1997) Accelerated evolution of sites undergoing mRNA editing in plant mitochondria and chloroplasts. Mol Biol Evol 14:344–349

Small ID, Peeters N (2000) The PPR motif—a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25: 46–47

Sper-Whitis GL, Russell AL, Vaughn JC (1994) Mitochondrial RNA editing of cytochrome c oxidase subunit II (coxII) in the primitive vascular plant *Psilotum nudum*. Biochim Biophys Acta 1218:218–220

Sper-Whitis GL, Moody JL, Vaughn JC (1996) Universality of mitochondrial RNA editing in cytochrome-c oxidase subunit I (coxI) among the land plants. Biochim Biophys Acta 1307: 301–308

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V (1999) Plant mitochondrial RNA editing. J Mol Evol 48:303–312

Sugita M, Miyata Y, Maruyama K, Sugiura C, Arikawa T, Higuchi M (2006) Extensive RNA editing in transcripts from the PsbB operon and RpoA gene of plastids from the enigmatic moss *Takakia lepidozioides*. Biosci Biotechnol Biochem 70:2268–2274

Sung TY, Tseng CC, Hsieh MH (2010) The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. Plant J 63:499–511

Takenaka M (2010) MEF9, an E-subclass pentatricopeptide repeat protein, is required for an RNA editing event in the nad7 transcript in mitochondria of *Arabidopsis*. Plant Physiol 152: 939–947

Takenaka M, Verbitskiy D, Zehrmann A, Brennicke A (2010) Reverse genetic screening identifies five E-class PPR proteins involved in RNA editing in mitochondria of *Arabidopsis thaliana*. J Biol Chem 285:27122–27129

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tang J, Kobayashi K, Suzuki M, Matsumoto S, Muranaka T (2010) The mitochondrial PPR protein LOVASTATIN INSENSITIVE 1 plays regulatory roles in cytosolic and plastidial isoprenoid biosynthesis through RNA editing. Plant J 61:456–466

Tasaki E, Hattori M, 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

Tillich M, Funk HT, Schmitz-Linneweber C, Poltnigg P, Sabater B, Martin M, Maier RM (2005) Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. Plant J 43:708–715

Tillich M, Lehwark P, Morton BR, Maier UG (2006) The evolution of chloroplast RNA editing. Mol Biol Evol 23:1912–1921

Tsudzuki T, Wakasugi T, Sugiura M (2001) Comparative analysis of RNA editing sites in higher plant chloroplasts. J Mol Evol 53: 327–332

Vangerow S, Teerkorn T, Knoop V (1999) Phylogenetic information in the mitochondrial nad5 gene of pteridophytes: RNA editing and intron sequences. Plant Biol 1:235–243

Verbitskiy D, Zehrmann A, van der Merwe JA, Brennicke A, Takenaka M (2010) The PPR protein encoded by the LOVASTATIN INSENSITIVE 1 gene is involved in RNA editing at three sites in mitochondria of *Arabidopsis thaliana*. Plant J 61: 446–455

Verbitskiy D, Hartel B, Zehrmann A, Brennicke A, Takenaka M (2011) The DYW-E-PPR protein MEF14 is required for RNA editing at site matR-1895 in mitochondria of *Arabidopsis thaliana*. FEBS Lett 585:700–704

Volkmar U, Knoop V (2010) Introducing intron locus cox1i624 for phylogenetic analyses in Bryophytes: on the issue of *Takakia* as sister genus to all other extant mosses. J Mol Evol 70:506–518

Volkmar U, Groth-Malonek M, Heinrichs J, Muhle H, Polsakiewicz M, Knoop V (2011) Exclusive conservation of mitochondrial group II intron nad4i548 among liverworts and its use for phylogenetic studies in this ancient plant clade. Plant Biol. doi: 10.1111/j.1438-8677.2011.00499.x

Wickham H (2009) Ggplot2: elegant graphics for data analysis. Springer, New York

Wolf PG, Rowe CA, 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

Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL (2005) The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350:117–128

Yoshinaga K, Iinuma H, Masuzawa T, 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

Yura K, Miyata Y, Arikawa T, Higuchi M, Sugita M (2008) Characteristics and prediction of RNA editing sites in transcripts of the Moss *Takakia* lepidozioides chloroplast. DNA Res 15: 309–321

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 21:558–567

Zhou W, Cheng Y, Yap A, Chateigner-Boutin AL, Delannoy E, Hammani K, Small I, Huang J (2008) The Arabidopsis gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. Plant J 58:82–96