



# Molecular Evolution of *rbcL* in Orthotrichales (Bryophyta): Site Variation, Adaptive Evolution, and Coevolutionary Patterns of Amino Acid Replacements

Moisés Bernabeu<sup>1</sup> · Josep A. Rosselló<sup>2</sup>

Received: 17 April 2020 / Accepted: 31 January 2021 / Published online: 20 February 2021  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

Molecular evolution of the large subunit of the RuBisCO enzyme is understudied in early diverging land plants. These groups show morphological and eco-physiological adaptations to the uneven and intermittent distribution of water in the terrestrial environment. This might have prompted a continuous fine-tuning of RuBisCO under a selective pressure modifying the species-specific optima for photosynthesis in contrasting microdistributions and environmental niches. To gain a better insight into the molecular evolution of RuBisCO large subunits, the aim of this study was to assess the pattern of evolutionary change in the amino acid residues in a monophyletic group of Bryophyta (Orthotrichaceae). Tests for positive, neutral, or purifying selection at the amino acid level were assessed by comparing rates ( $\omega$ ) of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) nucleotide substitutions along a Maximum Likelihood phylogenetic tree. Molecular adaptation tests using likelihood ratio tests, reconstruction of ancestral amino acid sites, and intra-protein coevolution analyses were performed. Variable amino acid sites (39) were unevenly distributed across the LSU. The residues are located on *rbcL* sites that are highly variable in higher plants and close to key regions implying dimer–dimer ( $L_2L_2$ ), RuBisCO-activase interactions, and conformational functions during catalysis. Ten *rbcL* sites (32, 33, 91, 230, 247, 251, 255, 424, 449 and 475) have been identified by the Bayesian Empirical Bayes inference to be under positive selection and under adaptive evolution under the M8 model. The pattern of amino acid variation suggests that it is not lineage specific, but rather representative of a case of convergent evolution, suggesting recurrent changes that potentially favor the same amino acid substitutions that are likely optimized the RuBisCO activity.

**Keywords** Early diverging land plants · Mosses · Orthotrichaceae · Ribulose-1,5-bisphosphate carboxylase/oxygenase

## Introduction

Photosynthesis is a key chloroplast process by which photoautotroph organisms transform light energy into chemical energy. The starting reaction is catalyzed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO, E.C. 4.1.1.39), which incorporates atmospheric CO<sub>2</sub> to the Calvin cycle, thus integrating inorganic carbon

into the biosphere (Erb and Zarzycki 2018). In green algae, early diverging land plants (liverworts, mosses and hornworts), and vascular plants (fern like, ferns, seed plants), the RuBisCO Form I holoenzyme is composed by eight large subunits ( $L_8$ ) encoded by the chloroplast *rbcL* gene and eight small subunits ( $S_8$ ) encoded by the nuclear *rbcS* gene family, assembled into a hexadecamer ( $L_8S_8$ ) (Spreitzer and Salvucci 2002).

It was early suggested that *rbcL* evolution is strongly constrained by function (Albert et al. 1994; Kellogg and Juliano 1997). Nucleotide sequences from the *rbcL* gene have been one of the most preferred plastid DNA locus for reconstructing land plant phylogenies, both at deep and lower evolutionary nodes (Chase et al. 1993; Manhart 1994; Gastony and Rollo 1995; Hasebe et al. 1995; Cameron et al. 1999; Tsubota et al. 2004; Masuzaki et al. 2010; Liu et al. 2012). Nevertheless, plant phylogeneticists using *rbcL* nucleotides

Handling editor: Dorothee Huchon.

✉ Josep A. Rosselló  
rossello@uv.es

<sup>1</sup> Departament de Genètica, Universitat de València, c/ Doctor Moliner 50, Burjassot, 46100, València, Spain

<sup>2</sup> Jardín Botánico, ICBiBE, Universitat de València, c/ Quart 80, 46008 València, Spain

have usually ignored functional constraints and have treated DNA sequences as if it were a string of anonymous nucleotides devoid of function as aptly pointed out by Kellogg and Juliano (1997). Neglection of the functional constraints have gone as far as to consider the *rbcL* gene as a universal DNA barcode (and mini-barcode) for plants (Group CPW 2009; Erickson et al. 2017).

At the amino acid level, the identification of sites subjected to adaptive evolution is essential in order to understand the RuBisCo kinetics variability. In addition, the assessment of coevolutionary replacements may reveal clues about the cytonuclear molecular evolution processes of the LSU and SSU *rbc* subunits that may be mediated by intergenomic gene conversion and altered transcription of duplicated, homoeologous nuclear genes (Gong et al. 2014).

In comparative analysis of protein-coding DNA sequences, the non-synonymous–synonymous rate ratio ( $d_N/d_S$ , denoted  $\omega$ ) has been usually used as a measure of selective pressure. In RuBisCO, most *rbcL* sites are likely to be functionally constrained and are under purifying selection ( $\omega < 1$ ), while only a relative small number of amino acid residues might tolerate modification and are under neutral evolution ( $\omega = 0$ ) or positive Darwinian selection ( $\omega > 1$ ). This is not surprising since RuBisCO large subunits possess the catalytic site and other amino acid residues involved in the functionality of the protein. This implies a proper structure folding, interactions with the small subunits of the RuBisCO holoenzyme, as well as RuBisCO-activase, a catalytic chaperone engaged in the RuBisCO activation (Andersson 2007). Several studies have revealed adaptive evolution of RuBisCO in all lineages of green plants (Kapralov and Filatov 2006). This approach has been rarely conducted in land plants and most of the results have been obtained from the analysis of seed plants (Kapralov and Filatov 2006; Sen et al. 2011; Kapralov et al. 2012; Galmés et al. 2014; Hermita-Carrera et al. 2016, 2017, 2020).

In addition, complex evolutionary processes modeling RuBisCO fitness have been inferred by identifying conserved residue sites whose mutations may be deleterious, and sites where amino acid replacements and coevolutionary substitutions are likely to improve the adaptive enzyme performance (Andersson 2007; Sen et al. 2011).

Extant representatives of early diverging plants, i.e., the three basal-most land plant lineages, liverworts (Marchantiophyta), mosses (Bryophyta), and hornworts (Anthocerotophyta), accumulate thousands of *rbcL* sequences in GenBank databases. However, a few reports related the genetic and kinetic variability of the LSU. In fact, only Miwa et al. (2009) and Kapralov and Filatov (2006) have examined *rbcL* sequence variation to assess selective pressure at the protein level.

This scanty knowledge is unfortunate since ancient land plant lineages may provide suitable case studies to gain

insights into the molecular evolution of RuBisCO large subunits. First, their fossil record traces back to the early Paleozoic Era (Tomescu et al. 2018), suggesting a long history of genome evolution and gene expression linked to lineage diversification. Moreover, the colonization of terrestrial habitats where atmospheric CO<sub>2</sub> concentrations differed from those present in the primeval aquatic environments might have prompted a continuous fine-tuning of RuBisCO under a selective pressure modifying the species-specific optima for photosynthesis (Iida et al. 2009). Finally, the bryophytes do not show neither an efficient cuticle layer nor stomata in the prominent phase of their life cycles (the gametophyte) and their water content is completely dependent on the humidity of the environment (atmosphere and substrate) (Glime 2007). This, in addition to their simple photosynthetic structure (in particular the low ratio of the internal photosynthetic tissues to external surface area), affects the ecophysiology of photosynthesis (Green and Lange 1994).

In the present study, we focus on the *rbcL* gene evolution of a moss family, Orthotrichaceae (Bryophyta), which is the single family encompassing the order Orthotrichales. It is a highly speciose and cosmopolitan bryophyte group of slow-growing plants showing a relative homogeneous plant form of the gametophyte. The family includes mostly epiphyte species showing variation in habitats.

Our main objective was to assess the evolutionary changes in the amino acid residues of the large subunit of the RuBisCO enzyme in a monophyletic group of early diverging land plants showing morphological and eco-physiological adaptations to the uneven and intermittent distribution of water in the terrestrial environment (Proctor 1979; Alpert and Oechel 1987). All these features might have modulated the evolution of RuBisCO under contrasting microdistributions and environmental niches.

Specifically, we were interested in addressing the following questions: (i) which amino acid sites appear to be positively selected and under adaptive evolution? (ii) are hot spots of amino acid variation located at specific protein locations or interfaces? (iii) are amino acid replacements lineage specific? and (iv) does variation in the amino acid residues follow any significant coevolutionary pattern?

## Methods

### Species Selection and Retrieval of *rbcL* Sequences

This study is focused on family Orthotrichaceae, excluding Erpodiaceae and Rhachiteciaceae (Goffinet et al. 1998), as delimited by Norris et al. (2004) which forms a distinct monophyletic group using plastid DNA sequences (Tsubota et al. 2004). Generic delimitation is somewhat controversial (Goffinet et al. 1998, 2004) and

contradictory hypotheses have recently resulted in the splitting (Matcham and O’Shea 2005; Lara et al. 2016; Mizia et al. 2019) or lumping of genera (Calabrese and Muñoz 2006). Nucleotide and protein accessions available from GenBank databases were downloaded (accessed 5th June 2019). *Takakia* and *Sphagnum* representatives (the basal-most clade of Bryophyta), and *Tetraphis pelucida*, were used for comparative (outgroup) purposes. Partial *rbcL* sequences shorter than 435 amino acid residues were excluded from the analysis. GenBank accession numbers for the nucleotide and amino acid sequences from 45 ingroup accessions and three outgroups analyzed in this study are shown in Online resource 1.

### Phylogenetic Analysis

Sequences were aligned with MAFFT v7.427 software (Kato et al. 2017) resulting in 1425 nucleotide and 475 amino acid site alignments (Online resource 2). The sequence from *Sphagnum rigescens* was used as a reference for delimiting the 5’ and 3’ ends of the *rbcL* gene. The initial alignment was manually inspected and missing positions and indeterminations were noted as ‘NNN’ codons. Then, the alignment was trimmed using TrimAl v1.2 (Capella-Gutiérrez et al. 2009) with the *-gappycout* option. The optimal nucleotide substitution model was determined using jModelTest v.2.1.10 (Darriba et al. 2012), by comparing available models using the Akaike Information Criterion, Bayesian Information Criterion and Decision Theory Criterion. We inferred the phylogenetic relationships using both Maximum Likelihood and Bayesian approaches. Maximum Likelihood trees were generated in PhyML v3.1 (Guindon et al. 2010). The best fit model of nucleotidic evolution was the GTR model with a *p*-inv = 0.504 value for the proportion of invariant sites and a substitution rate distributed according to a discrete gamma with four categories and an inferred shape of  $\alpha = 0.871$  ( $\Gamma_4$ ). In addition, a phylogenetic Bayesian analysis with the same alignment was performed using BEAST v1.10.4 (Suchard et al. 2018) with a Markov Chain Monte Carlo (MCMC) process using the same models and parameters established for the Maximum Likelihood analysis.  $1 \times 10^6$  generations and trees were sampled every 1000 generations. The output trees were subjected to a phylogenetic signal analysis using the PhyloSignal package v.1.2.1 (Keck et al. 2016), running in R v.3.6.0, to estimate *p*-values for the methods  $C_{\text{mean}}$  and  $\lambda$ . The trend model was defined by the fastBM function of PhyTools v.0.6–99 package (Revell 2012) with  $\mu = 0$  that implies a Brownian motion with a trend. The random model was inferred generating a normal distribution with  $n = 48$  and  $\alpha = 10$ .

### Adaptability Analysis of Amino Acid Substitution

Tests for positive, neutral or purifying selection at the molecular level can be assessed by comparing rates ( $\omega$ ) of non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) nucleotide substitutions (Yang et al. 2005) along a phylogenetic tree. The measure of selection pressure ( $\omega = d_N/d_S$ ) is expected to be equal to 1 under neutral selection. Departures from this value are indicative of either purifying ( $0 < \omega < 1$ ) and positive selection ( $\omega > 1$ ). Estimates of  $\omega$  were carried out with the CodeML program in the PAML v.4.9 software (Yang 2007) and the SLR v.1.3 program (Massingham and Goldman 2005). SLR incorporates a Nielsen–Yang model-based distribution of  $\omega$  at each site in the alignment and allows every site to be under a different level of evolutionary constraint. The SLR software tests whether a particular site is evolving in a non-neutral fashion using likelihood ratio tests (LRT) between  $\omega_i = 1$  vs.  $\omega_i \neq 1$  sites and is a measure of the strength of the evidence for selection.

Molecular adaptation tests (Nielsen and Yang 1998; Yang et al. 2000; Wong et al. 2004) were used to assess selection on a codon-by-codon basis, allowing for variation in  $d_N/d_S$  across the alignment and performed on the ML phylogenetic tree obtained previously. LRT were used to compare the distribution of a null model of codon substitution that does not allow for any codon to be  $d_N/d_S > 1$  against a model that does. Seven different models proposed by Yang et al. (2000, 2005) were compared: M0 (one ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and  $\omega$ ) and M8a (beta and  $\omega = 1$ ).

Briefly, the model 0 (M0) allows for a single  $\omega$  value for all sites and branches from the ML phylogenetic tree. Model 1 (M1a), considers a proportion of conserved sites ( $p_o$ ) with  $\omega_o = 0$  and a proportion of neutral sites ( $p_1$ ) with  $\omega_o = 1$ . Model 2a (M2a), the modified M2 model, adds the proportion of  $p_2$  sites with  $\omega_2 > 1$  estimated from the data. In models 7 (M7) and 8 (M8 and M8a)  $\omega$  was estimated from a beta distribution  $B(p, q)$  for a proportion ( $p_o$ ) of sites. However, M7 does not allow for the presence of positively selected sites ( $\omega > 1$ ), in contrast with M8 models. The difference between M8 and M8a models is that under the more general M8 model  $\omega_s > 1$ , whereas in the M8a model  $\omega_s = 1$ . When the LTR indicates that models accounting for positive selection are preferred, the Bayes Empirical Bayes (BEB) inference was used to calculate posterior probabilities for the sites with  $\omega > 1$  (Yang et al. 2005). Positively selected sites were considered as supported when posterior probability values were greater than 99%.

### Evolutionary Trends in Amino Acid Substitutions

Reconstruction of ancestral amino acid sites showing  $\omega > 1$  with posterior probability higher than 99% in the

M8 model was carried out using the maximum likelihood (ML) approach in Mesquite version 3.04 (Maddison and Maddison 2019), which assigns to each internal node the character state that maximizes the probability of obtaining the observed character states in the terminal taxa under the specified model of evolution. The ML reconstructions were conducted using the Mk1 model of evolution (Schluter et al. 1997; Pagel 1999). The Mk1 (Markov k-state 1 parameter model) is a k-state generalization of the Jukes–Cantor model, corresponding to Lewis (2001) Mk model, which assigns equal probability to changes between any two character states.

### Intra-protein Coevolution Analyses

To assess the evolutionary dependency among amino acid sites of the RuBisCO large subunit we used the CAPS software v.2 (Fares and McNally 2006) at the program server (<http://caps.tcd.ie/caps/home.html>). This program provides an automatization of a designed pipeline that allows comparison of a correlated variance of the evolutionary rates among pairwise sites and their estimated divergence times. The amount of amino acid replacements is used as a relative measure of time. Groups of amino acid sites were considered to be significantly coevolving when correlation values were higher than 0.5 and bootstrap values were larger than 0.95 (Yao et al. 2019).

### Structural Analysis of RuBisCO

A model for the LSU structure of *Macromitrium japonicum* (GenBank accession BAD98508) was generated at the Swiss-Model web site (Waterhouse et al. 2018), a fully automated protein structure homology-modeling server. The protein structure was annotated with PyMol v.2.3.2 (<https://pymol.org/2/>), a molecular visualization tool. Structural motifs of the large subunit of RuBisCO were obtained from Kellogg and Juliano (1997) and Spreitzer and Salvucci (2002).

## Results

### *rbcl* Variability

Complete sequences of the *rbcl* gene (475 amino acid residues) were available for 16 ingroup accessions (14 species). The shortest sequence (437 sites) was present in one of the *Pulviger a lyellii* accessions (Online resource 1). There were 39 variable sites (Table 1), of which 10 were restricted to a single accession and 29 were shared by at least two sequences. Three or four amino acid states could occur at codons 32 (lysine, serine, threonine, leucine), 33 (aspartic

acid, glutamic acid, glutamine), 221 (valine, cysteine, isoleucine), 230 (alanine, glycine, serine), 255 (glutamine, alanine, glutamic acid), 256 (phenylalanine, cysteine, alanine), 424 (valine, alanine, leucine), 449 (alanine, threonine, serine) and 475 (valine, leucine, isoleucine). Overall, we identified 43 unique sequences (amino acid haplotypes) in the ingroup dataset (45 accessions, 43 species), indicating that all species could be identified by a unique amino acid profile of the *rbcl* gene.

Seven species with two accessions were present in GenBank. The two sequences of both *Macromitrium gymnostomum* and *M. ferriei* showed the same haplotype. However, intraspecific variation was detected in *Macromitrium incurvifolium* (five amino acid sites), *Pulviger a lyellii* (four sites), *Ulota crispa* (three sites), *Nyholmia obtusifolia* (two sites), and *Macrocoma tenuis* (one site).

### Phylogenetic Analysis

The statistics values and associated *p*-values of the phylogenetic signal analysis on the ML and Bayesian trees are indicated in Online resource 3. The best significance of the trend model values compared to the random model indicated the presence of phylogenetic signal in both ML and Bayesian trees. Since the phylogenetic trees obtained by both approaches showed similar topologies, we selected the ML tree for the analyses, since the *p*-value of the random model and trend model test showed a better adjustment to the trend model. The selected ML tree used to assess the adaptability analysis of amino acid substitution and to infer the ancestral amino acid sites is shown in Fig. 1.

### Tests of Positive Selection in the *rbcl* Gene

Table 2 lists the parameters-estimated values for seven codon substitution models of molecular evolution of the RuBisCO LSU. The significance of nested site-specific models were tested by LRT and the results are shown in Table 3. Model comparisons indicated that the *rbcl* gene has evolved in a non-neutral fashion since the null model assuming neutrality (M0) was rejected in favor of the alternative M1a model. Heterogeneity in  $\omega$  ratios was shown in the comparison of the M3 and M0 codon models, where the former model was favored. Finally, the M2a and M8 models of positive selection were preferred over the null M1a and M7 models, suggesting the presence of positive selection. The amino acid residues 32, 33, 91, 230, 247, 251, 255, 424, 449 and 475 were identified as evolving under positive selection under the M8 model, and a subset of them (33, 91, 230, 475) under the M2a model, computed with PAML by the BEB analysis with a Bayesian posterior probability larger than 0.99. For each analyzed genus, the ten amino acid residues selected by

**Table 1** Amino acid substitutions in *Orthotrichaceae rbcL*

Codon number	Location of residue	Amino acid change		Type of change
11	N-terminal domain below $\beta$ -strand A	S	VA	Variable amino acids
<b>23</b>	N-terminal domain below $\beta$ -strand A	TN	TN	Variable amino acids
<b>28</b>	Between $\beta$ -strand A and $\beta$ -strand B	ED	ED	Variable amino acids
32	Between $\beta$ -strand A and $\beta$ -strand B	K	KSTL	Variable amino acids
33	Between $\beta$ -strand A and $\beta$ -strand B	D	DEQ	Variable amino acids
87	$\beta$ -strand C	I	IL	HN $\rightarrow$ HN
91	Between $\beta$ -strand C and $\beta$ -strand D	A	AP	HN $\rightarrow$ HN
94	Between $\beta$ -strand C and $\beta$ -strand D	E	ED	UA $\rightarrow$ UA
<b>139</b>	$\beta$ -strand E	R	RS	UB $\rightarrow$ UP
200	$\beta$ -strand 2	T	TP	UP $\rightarrow$ HN
221	$\alpha$ -helix 2	V	VCI	Variable amino acids
230	$\alpha$ -helix 2	A	AGS	Variable amino acids
247	$\alpha$ -helix 3	C	CS	HP $\rightarrow$ UP
251	$\alpha$ -helix 3	M	ML	HN $\rightarrow$ HN
255	$\alpha$ -helix 3	QA	QAE	Variable amino acids
256	$\alpha$ -helix 3	F	FCA	Variable amino acids
279	$\alpha$ -helix 4	ST	ST	Variable amino acids
<b>281</b>	$\alpha$ -helix 4	A	AS	HN $\rightarrow$ UP
<b>282</b>	$\alpha$ -helix 4	H	HR	UB $\rightarrow$ UB
288	Between $\alpha$ -helix 4 and $\beta$ -strand 5	G	GA	UN $\rightarrow$ HN
292	$\beta$ -strand 5	H	HY	UB $\rightarrow$ UR
301	$\alpha$ -helix F	Y	YL	UR $\rightarrow$ HN
<b>302</b>	$\alpha$ -helix F	D	DE	UA $\rightarrow$ UA
306	Between $\alpha$ -helix F and $\beta$ -strand F	IN	IN	Variable amino acids
309	$\beta$ -strand F	M	MI	HN $\rightarrow$ HN
326	$\beta$ -strand 6	I	IV	HN $\rightarrow$ HN
328	Between $\beta$ -strand 6 and $\alpha$ -helix 6	AS	AS	Variable amino acids
340	$\alpha$ -helix 6	QE	QE	Variable amino acids
<b>348</b>	$\alpha$ -helix 6	L	LP	HN $\rightarrow$ HN
<b>352</b>	Between $\alpha$ -helix 6 and $\beta$ -strand G	D	DN	UA $\rightarrow$ UP
374	Between $\beta$ -strand H and $\beta$ -strand 7	V	VI	HN $\rightarrow$ HN
<b>400</b>	$\beta$ -strand 8	L	LS	HN $\rightarrow$ UP
424	$\alpha$ -helix 8	I	VAL	Variable amino acids
<b>429</b>	$\alpha$ -helix 8	Q	QL	UP $\rightarrow$ HN
443	$\alpha$ -helix G	NED	ED	Variable amino acids
444	$\alpha$ -helix G	I	IV	HN $\rightarrow$ HN
449	$\alpha$ -helix G	A	ATS	Variable amino acids
470	C-terminal domain	ED	ED	Variable amino acids
475	C-terminal domain	V	VLI	Variable amino acids

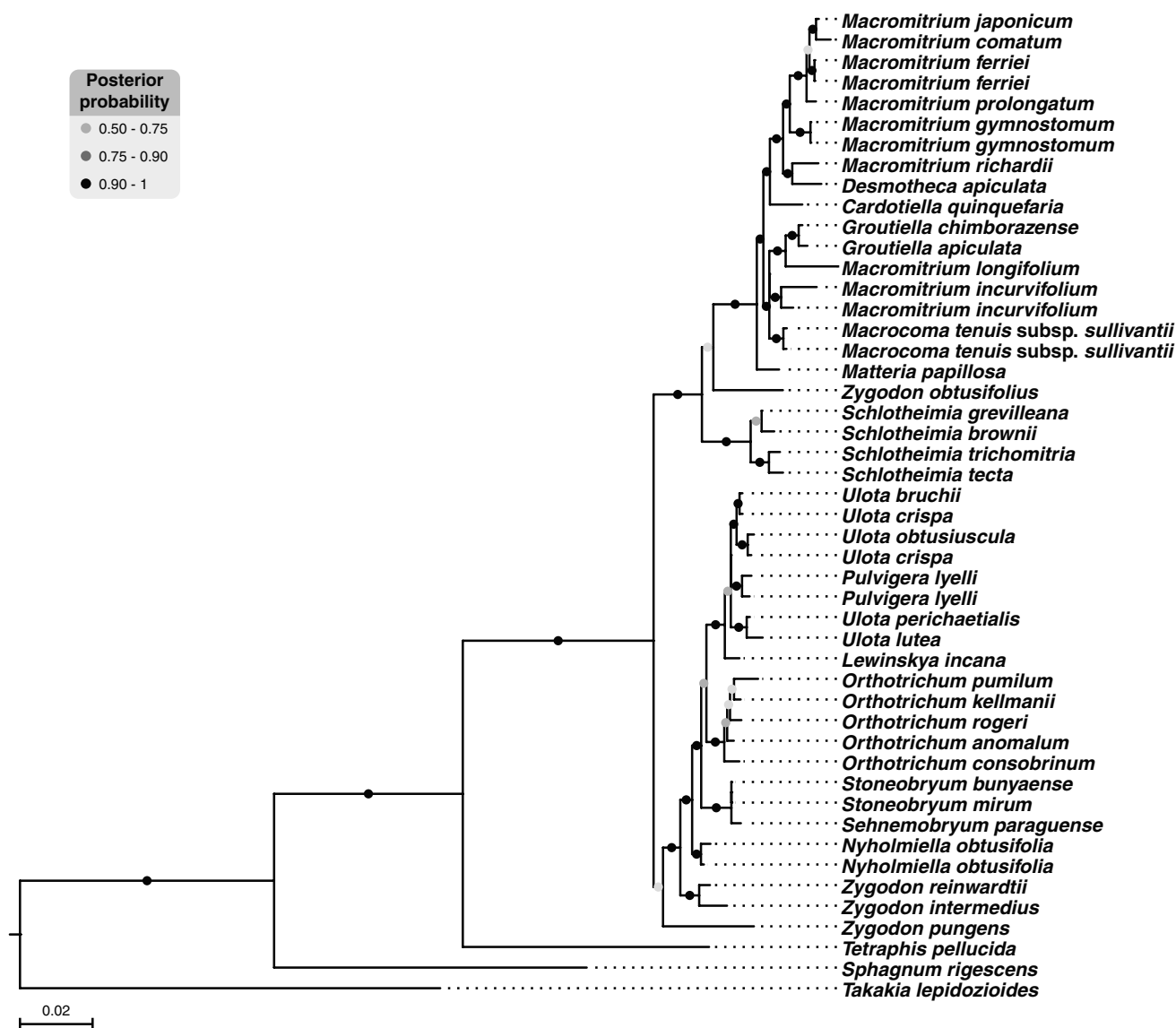
The location of the residues in the secondary structure of *rbcL* (Kellogg and Juliano 1997), the inferred substitutions using *Sphagnum* species as the outgroup, and the side chain type changes are indicated. Variable residues present only in a single accession are indicated in bold

Type of change: H=Hydrophobic, U=Hydrophilic, A=Acidic, B=Basic, N=Non-polar, P=Polar, R=Aromatic

the M8 model are indicated in Online resource 4. The average  $\omega$  ratio of 0.053 was calculated by SLR with a log-likelihood of  $-5878.420$ , indicating a conservative nature of the whole *rbcL* gene. The single rate homogeneity was discarded in two positions, 251 and 475, which are a subset of those previously inferred by the M8 model.

### Structural Distribution and Features of Amino Acid Replacement Sites

The amino acid substitutions identified in *Orthotrichaceae*, the location of the residues in the structure regions of the RuBisCO large subunit, their evolutionary polarity (using



**Fig. 1** Maximum Likelihood tree of Orthotrichaceae inferred from *rbcL* sequences under the GTR model with a  $p\text{-inv}=0.504$  value for the proportion of invariant sites and  $\alpha=0.871$  as the gamma shape

value for four rate categories. Posterior probability values are indicated by circles in different gray tones according to the inset label

*Sphagnum* species as reference), and the type of physical properties of the changes are presented in Table 1. Seven substitutions were located on the N-terminal and the C-carboxyl terminus, 16 in  $\alpha$ -helices, six in  $\beta$ -strands, two between  $\alpha$ -helices, six between  $\beta$ -strands, one between  $\alpha$ -helix and  $\beta$ -strand, and one between a  $\beta$ -strand and a loop. The placement of the variable sites in the three-dimensional structure of the RuBisCO large subunit is shown in Fig. 2. Inferring the ancestral state of several amino acid substitutions, and hence assessing the direction of change in their physical properties, was not always possible using outgroup comparisons. In some cases, sites were variable in the outgroup (23, 28, 255, 279, 306, 328, 340, 443, 470). In others,

sites were monomorphic in the outgroup (11, 32, 33, 221, 230, 256, 449, 475) but showed three or four amino acid substitutions in variable sites of Orthotrichaceae. This precluded a straightforward explanation for multistate characters and require theories of character evolution, or to assume that any transformation series between different states is always possible and could be established with certainty. The inspection of the 21 sites for which the direction of change could be indicated, showed that 11 substitutions did not involve changes in both the hydrophobicity and polarity of the amino acids, three changed only the hydrophobicity of the residues, two affected their polarity, and five were implicated in changes in both their hydrophobicity and polarity (Table 1).

**Table 2** Parameter estimates and log-likelihood values for Orthotrichaceae under seven codon substitution models

Model	$N_p$	Log-likelihood	Parameters	Positively selected sites (BEB)
M0 (one ratio)	96	−5850.366	$\omega = 0.061$	Not allowed
M1a (nearly neutral)	97	−5619.013	$p_0 = 0.916$ $\omega_0 = 0.084$ $p_1 = 0.017$ $\omega_1 = 1.000$	Not allowed
M2a (positive selection)	98	−5847.117	$p_0 = 0.984$ $\omega_0 = 0.016$ $p_1 = 0.014$ $\omega_1 = 1.000$ $p_2 = 0.002$ $\omega_2 = 3.134$	33E, 91P, 230A, 475I
M3 (discrete)	100	−5606.112	$p_0 = 0.839$ $\omega_0 = 0.007$ $p_1 = 0.115$ $\omega_1 = 0.257$ $p_2 = 0.046$ $\omega_2 = 1.265$	Not allowed
M7 (beta)	97	−5614.558	$p = 0.0610$ $q = 0.475$	Not allowed
M8 (beta and $\omega > 1$ )	98	−5845.503	$p_0 = 0.995$ $p = 0.061$ $q = 1.900$ $p_1 = 0.005$ $\omega_1 = 2.335$	32T, 33E, 91P, 230A, 247S, 251M, 255A, 424V, 449T, 475I
M8a (beta and $\omega = 1$ )	98	−5606.952	$p_0 = 0.942$ $p = 0.111$ $q = 2.863$ $p_1 = 0.058$ $\omega_1 = 1.000$	Not allowed

$N_p$  is the number of *used* parameters. Positively selected amino acid sites with posterior probability higher than 99% are indicated. Amino acids positions refer to the *Macromitrium japonicum* sequence (GenBank accession BAD98508)

**Table 3** Test for selection models by LRT

Model (alternative-null)	$2\Delta l$	d.f.	$p$ -value
M2a-M1a	456.208	2	< 0.0001
M3-M0	488.508	4	< 0.0001
M8-M7	461.890	2	< 0.0001
M8-M8a	477.102	1	< 0.0001

## Ancestral Amino Acid Sequences

The ML reconstruction of ten ancestral amino acid sites (32, 33, 91, 230, 247, 251, 255, 424, 449 and 475) showing positive selection signatures ( $\omega > 1$ ) by BEB inferred by the M8 codon model is shown in Online resources 6–15. Overall, most of the site changes in the ancestral sequences were located at the more derived nodes of the tree. Independent changes in the different states of the ten amino acids are present throughout the phylogenetic tree.

## Inter-dependent Evolution of Amino Acid Sites in the LSU Subunit

Coevolutionary interactions between pairwise amino acid sites in LSU were assessed using the CAPS software. The most significant coevolving amino acid pairs (i.e., those showing correlation and bootstrap values higher than 0.5 and 0.95, respectively) are shown in Table 4. The relevant associations involved nine coevolutionary pairs involving 13

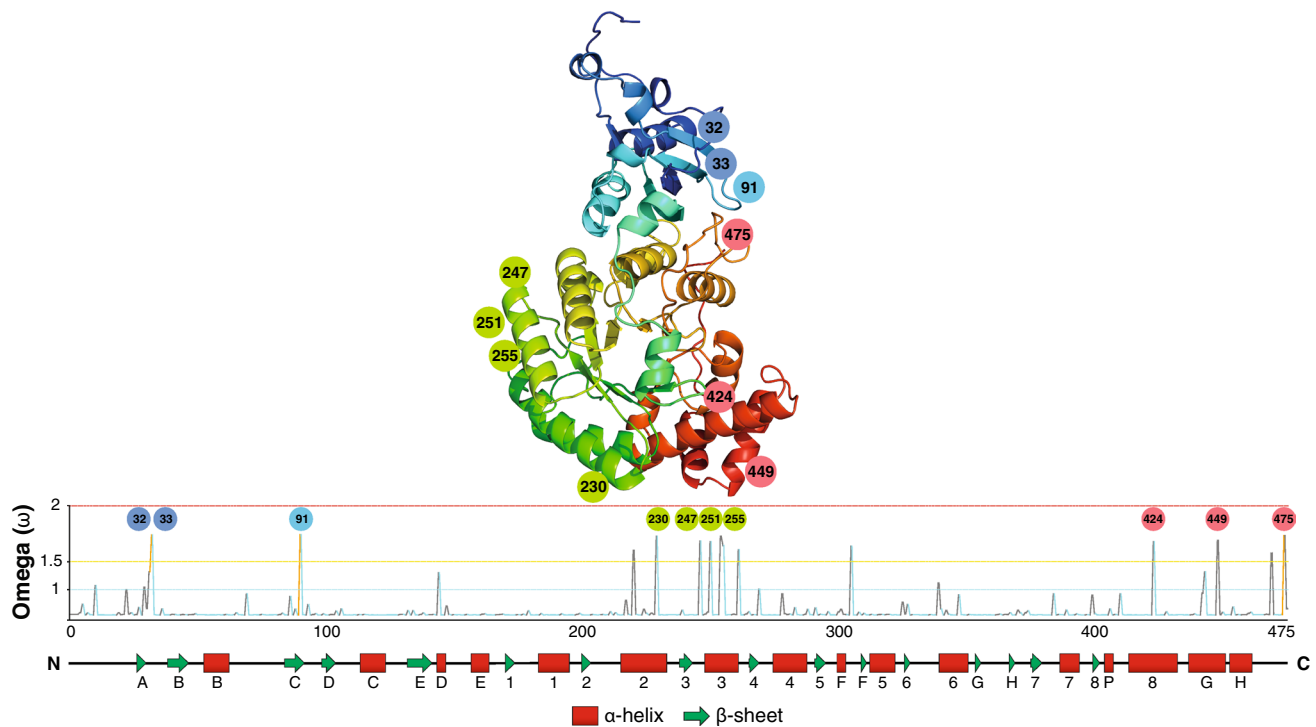
sites (11, 30, 50, 89, 262, 270, 348, 387, 404, 443, 449, 470 and 475). The most interacting site (50) involved four amino acid residues and was located on the N-terminal domain ( $\alpha$ -strand B). In turn, sites 89 ( $\beta$ -strand C) and 262 ( $\beta$ -strand 4) (C-terminal domain) were the most frequent significant coevolving residues involving two sites. Interestingly, only two sites involved in the coevolutionary interactions (449, 475) were identified as subjected to adaptive selection (see above).

## Discussion

A relative large amount of *rbcL* gene sequences (about 9000) are currently available in public databases for the three extant groups of early land plants (GenBank accessed 7th August 2019). However, most of these nucleotide data have been obtained for phylogenetic reconstruction, taxonomic purposes or molecular species identification (Tsubota et al. 2004). In fact, only two studies, specifically assessing adaptive evolution in *rbcL* have used protein sequences to analyze the molecular evolution of RuBisCO in early diverging plants (Kapralov and Filatov 2007; Miwa et al. 2009).

## *rbcL* Residue Variation in Orthotrichaceae

We have identified 39 amino acid replacements representing 8.4% of variable sites in the LSU. Certainly, this number is conservative and probably does not account for the real extent of variable amino acid sites present in



**Fig. 2** Estimated  $\omega$  values under the M8 model for Orthotrichaceae *rbcL* residues (bottom) using the ETE Toolkit v.3.1.1 pipeline (Huerta-Cepas et al. 2016). The position of  $\alpha$ -helices and  $\beta$ -sheets is indicated according to the secondary structure given in Kellogg and Juliano (1997). Positive selected sites are indicated (top) in the ter-

tiary structure model for the LSU of *Macromitrium japonicum* (GenBank accession BAD98508) generated at the Swiss-Model web site (Waterhouse et al. 2018). Alignment for the ten indicated positively selected sites is shown in Online resource 5

**Table 4** Coevolving pairs of amino acids with a correlation higher than 0.50 and a bootstrap value higher than 95% obtained with CAPS v2 (Fares and McNally 2006)

AA 1	AA 2	Correlation	Bootstrap value (%)
89	262	0.994426	100
470	475	0.889498	100
30	50	0.770563	100
50	89	0.688643	97.6
348	449	0.685013	95.3
50	262	0.666142	97.7
50	270	0.656543	97.0
11	443	0.560222	96.9
387	404	0.542261	96.5

Orthotrichaceae. On the one hand, few accessions were available for the largest genera of the family, including *Macromitrium* (about 370 species; Guo et al. 2012), *Orthotrichum* (103 species; Lara et al. 2016), *Zygodon* (93 species; Calabrese and Muñoz 2006), *Ulota* (69 species; Garilleti et al. 2015) and *Lewinskya* (66 species; Lara et al. 2016), and sampling of additional unrelated species are required to substantiate the current data. In addition, most *rbcL* sequences

were incomplete and usually lack about 35 residues of the complete gene as a consequence of the close location of the amplifying and sequencing primers at both the 5' and 3' ends, where variable sites have been identified in plants (Kellogg and Juliano 1997) and in our Orthotrichaceae dataset.

Despite these shortcomings some conclusions can be drawn from a comparison of the sequences. Inspecting the GenBank sequence data obtained by Miwa et al. (2009) and Shaw et al. (2016) in two early diverging genera from liverworts (*Conocephalum*) and mosses (*Sphagnum*), 5.29% and 4.84% of variable amino acid sites, respectively, were identified. The roughly double value present in Orthotrichaceae was perhaps not unexpected as it was obtained from a more diverse phylogenetic sample (involving deeper nodes; 15 genera). However, when we compare the range of variable sites in these three groups of early land plants with other plant vascular lineages for which data are available, it appears that bryophytes may show more RuBisCO variability per phylogenetic unit. For instance, the number of amino acid replacements reported for Brassicaceae (3.93% across 45 genera and 33 species; Liu et al. 2012), and Fagaceae plus Nothofagaceae families (4.00% across 6 genera and 190 species; Hermida-Carrera et al. 2017). There is not a direct



association between rates of evolution and the number of variable sites in genes. In fact, it has been estimated that bryophytes show slow molecular evolution compared with vascular plants (ferns and seed plants) for ribosomal and protein-coding genes from the three molecular genomes (Stenøien 2008). However, a larger proportion of variable sites implies an evolutionary plasticity for the molecular adaptation of the RuBisCO protein to changing selective forces. The dataset used for comparison is extremely reduced that it can only stimulate future research in this specific topic.

### Amino Acid Site Variations are Located at Different Structural and Functional Domains

Variable amino acid sites detected in Orthotrichaceae were unevenly distributed across the LSU, which is in accordance with its complex structure and functional domains. Eight amino acid substitutions are located at the hydrophobic cores of the *rbcL* monomer, which are critical for the stabilization of the folded state. The hydrophobic cores of the N-terminal domain and that in the interface region between the N-terminal and C-terminal domains did not show any amino acid residue variation. In addition, only a site (292) was variable at the first hydrophobic core in the C-terminal domain, at the  $\alpha/\beta$  barrel that is formed by those residues from the  $\beta$ -strand 5 that point to the interior of the barrel. In contrast, most of the variation was placed at the second hydrophobic core of the C-terminal domain (200, 221, 326, 328, 374, 400, and 424). Interestingly, amino acid variation at sites 200, 328, and 400 involved changes of hydrophobicity.

Another set of amino acid substitutions were at key positions involved in the correct assembly of the RuBisCO holoenzyme, mainly at the dimer interface ( $L_2$  dimer). Thus, sites 247, 279, 301, 306, and 309 were at the C-terminal domain of two  $L_2$  subunits, site 470 was at the C-terminal domain of one  $L$  subunit interacting with the N-terminal domain of a second  $L$  unit, whereas site 288 was involved at the dimer-dimer interface. This last residue, together with sites 288 and 429, interacts between a large subunit ( $L$  subunit B) and a small subunit ( $S1$ ) of the  $L_8S_8$  holoenzyme. However, all these sites are not conserved across the large subunit of RuBisCO across seed plants. In fact, three alternative residues have been reported to be present at sites 279, 288, 301, and 309, four at site 306, five at site 247, six at site 429, whereas nine and ten amino acids have been indicated at sites 470 and 230, respectively (Kellogg and Juliano 1997).

The residue present at the active site (201) needs to be carbamylated for operating functional RuBisCO (Kannappan and Gready 2008). As expected, this residue was strictly conserved in Orthotrichaceae. This lack of variation was also observed in key residues at sites 175, 204, 294, and 334, which are involved in the multistep catalytic reactions of

RuBisCO (Andersson et al. 1989; Andersson 2007; Kannappan and Gready 2008). Likewise, residues 331–338, forming Loop 6, which are associated to cover the opening of the  $\alpha/\beta$  barrel, serving to close the active site and influencing the  $CO_2/O_2$  specificity (Andersson 2007), were conserved.

Two variable sites in Orthotrichaceae, 91 and 94, were present in a region, the  $\beta C$ – $\beta D$  loop (residues between 89 and 94), which has been previously identified as critical for species specificity of RuBisCO-activase interaction (Ott et al. 2000). The  $\beta C$ – $\beta D$  loop is located on the surface of the RuBisCO holoenzyme close to the conserved Loop 6 (Andersson 2007). The binding of RuBisCO-activase (site 311) at residue 94, together with the likely steric interactions of residues 314 and 312 on RuBisCO sites 89 and 93, respectively (Portis et al. 2007), promotes conformational changes in RuBisCO releasing the inhibitor sugar phosphates from the active site (Portis 2003). It has been also hypothesized that RuBisCO amino acids at sites 89 and 93 are likely associated to steric interactions with RuBisCO-activase residues 314 and 312, respectively (Portis et al. 2007). Given the specific critical residues for activase recognition in the large subunit of RuBisCO, it appears that residues at the  $\beta C$ – $\beta D$  loop should be highly conserved. This is not the case in Orthotrichaceae nor in seed plants. Thus, sites 91 and 94 are embedded within the hypervariable region from residue 86 through 95 and constitute one of the main mutational hotspots in the amino acid sequences of the large subunit of RuBisCO (Kellogg and Juliano 1997; Larson et al. 1997). This marked variability contradicts observations, reporting reversal in RuBisCO/RuBisCO-activase activity, when mutations at the  $\beta C$ – $\beta D$  loop occurred (Larson et al. 1997). It seems plausible to argue that mutations at the  $\beta C$ – $\beta D$  loop might be functionally neutral if compensatory changes at this activase region do not cause substantial differences in conformation at this surface domain (Ott et al. 2000) that may disturb critical interactions for the binding of both enzymes.

### Positively Selected Amino Acid Replacements and Codependent Evolution of Amino Acid Sites in the LSU Subunit

The evolution of the RuBisCO LSU subunit has been hypothesized to be under strong biophysical constraints and is modulated by recurrent tradeoffs between activity and stability (Studer et al. 2014). Thus, the acquisition of enhanced activity has been mediated by destabilizing structural mutations that are followed by compensatory mutations that restore global stability.

Coevolution of residues is common in RuBisCO and about half of the amino acid sites were detected as coevolving in green algae and land plants (Wang et al. 2011). Furthermore, it has been detected an overall overlap between

coevolving and positively selected residues (Wang et al. 2011).

Our results in Orthotrichales show that matches of coevolving (13 amino acid sites) and positively selected residues (ten sites) are few (two sites). The reasons underlying this low overlap merit discussion. Several algorithms involving different assumptions and methods have been developed to identify compensating alterations during protein evolution (see Juan et al. 2013 for a comprehensive account). Dunn et al. (2008) reported that a high background composed of random noise and phylogenetic components may interfere with the identification of evolving positions. In addition, most of the studies on protein evolution are performed over the linear sequence, ignoring the predicted atomic interactions between amino acid sites. The fact that not only amino acid interactions and function, but also phylogeny and stochastic components account for co-variation, caution about the correct identification of the coevolution components (Fares and Travers 2006; Codoñer and Fares 2008). We have reanalyzed our *rbcL* alignment using the MIP software (Dunn et al. 2008) that does not consider tree topology nor protein structure. This method is based on information theory and accurately estimates the level of background mutual information for each pair of positions and then the algorithm corrects the information formula to remove the influence entropy. Interestingly, MIP has identified only three coevolving sites (32, 255, 256) which are different from those identified by the CAPS software (11, 30, 50, 89, 262, 270, 348, 387, 404, 443, 449, 470 and 475). More important, two of the three coevolving sites selected by MIP (32 and 255) were also identified as evolving under positive selection under the M8 model in this study. Clearly, additional studies are needed to identify consensus coevolutionary amino acid sites in bryophytes using several contrasting assumptions and methods.

### Positively Selected Amino Acid Replacements

Ten *rbcL* sites have been identified by the Bayesian Empirical Bayes inference using the M8 model to be under positive selection in Orthotrichaceae. However, using the SLR test only two out of the ten residues (251 and 475) remained significant (95%) as positively selected sites. This decrease is not restricted to Orthotrichaceae since similar trends have been generally observed in other studies when PAML and SLR results were compared (e.g., Yao et al. 2019).

Kapralov and Filatov (2007) searched for positive selection in over 3000 *rbcL* sequences from species representing all lineages of green plants and other photosynthetic organisms. Their study analyzed 88 species of Bryophyta and included 12 *rbcL* accessions from Orthotrichales which have also been used in our study. Two out of the four *rbcL* residues identified to evolve under positive selection in

Orthotrichales by Kapralov and Filatov (2007), sites 251 and 255, have been detected in our expanded dataset by the M8 model. Both sites are among the most frequently reported residues to be under selection by Kapralov and Filatov (2007) across all lineages of green plants. However, the sites 32T, 33E, 91P, 230A, 247S, 424V, 449T, and 475I were not detected by these authors neither in Orthotrichales nor in the whole Bryophyta dataset. These diverging results are intriguing since their analysis and ours used the same (a) parameter estimates from M8 model, (b) Bayesian Empirical Bayes approaches, (c) Bayesian posterior probability of positive selection larger than 0.99, and (d) identical PALM package. Maximum likelihood estimates to reveal diversifying selection at amino acid sites relies on the phylogenetic relationships among sequences (organisms) (Yang et al. 2000). Due to computational time constraints, Kapralov and Filatov (2007) divided all of the *rbcL* sequences they analyzed into 151 small monophyletic groups by manual dissection of trees constructed, using the neighbor-joining algorithm. They built the phylogenetic trees using rather crude distance estimates (a homogeneous pattern of molecular evolution among lineages and uniform rates among sites) and no measure of clade reliability (as bootstrap values) was reported to be applied to their phylogenetic results. The four monophyletic groups presented in their Additional file 3 (Kapralov and Filatov 2007) showed spurious results which clearly conflicts with currently supported phylogenetic relationships in Bryophyta (Tsubota et al. 2004; Liu et al. 2019). More importantly, Kapralov and Filatov (2007) results indicated that Orthotrichales is polyphyletic; an inference that contradicts all phylogenetic hypotheses based on coding and non-coding sequences from the nuclear and both organellar genomes (Goffinet and Vitt 1998; Goffinet et al. 1998, 2001, 2004; Tsubota et al. 2004; Liu et al. 2019). It is likely that the use of non-supported phylogenetic trees in Bryophyta and the analysis of *rbcL* sequences from polyphyletic assemblages led to artefactual results concerning the identification of positively selected sites in Bryophyta in general, and in Orthotrichaceae in particular. It has been stated that the inference of sites under positive selection does not appear to be sensitive to the tree topology using ML trees (Yang et al. 2000). However, these authors also emphasized that a reasonably good phylogeny is necessary to perform LTRs of positive selection (Pie 2006). The odd phylogenetic results obtained by Kapralov and Filatov (2007) in *rbcL* sequences from Bryophyta may have influenced the detection of overall positive selection in RuBisCO, and could have affected the number of sites identified. Further research is necessary on the subject, since the RuBisCO sites shown to be under adaptive selection (117, 169, 247, 279, 309, 340) in another lineage of Early Land Plants (Marchantiophyta) by Miwa et al. (2009) are in disagreement with those reported by Kapralov and Filatov (2007) and this study.

## Concluding Remarks

Most members of Orthotrichaceae are epiphytes and usually species from the same and different genera grow intermingled. We have shown that, despite this apparent environmental homogeneity, all species analyzed to date show unique L subunit protein haplotypes. Ten *rbcL* sites (32, 33, 91, 230, 247, 251, 255, 424, 449 and 475) have been strongly supported to be positively selected and under adaptive evolution. The pattern of amino acid variation suggests that it is not lineage specific, but represents a case of convergent evolution suggesting recurrent changes that potentially favor the same amino acid substitutions and likely optimized the RuBisCO activity. The selected residues are located on *rbcL* sites that are highly variable in higher plants and close to key regions implying dimer–dimer ( $L_2L_2$ ), RuBisCO-activase interactions, and conformational functions during catalysis. Our results prompted future research in RuBisCO kinetics in co-growing species of Orthotrichaceae to assess to what extent catalytic properties of the holoenzyme are significantly different and which are the eco-physiological niche microenvironments driving RuBisCO variability.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00239-021-09998-w>.

**Acknowledgements** We thank the subject editor, two anonymous reviewers, and J. Martínez-Abaigar for providing insightful comments and constructive criticisms that greatly improved the manuscript.

**Author Contributions** MB and JAR conceived the study, MB and JAR designed the analysis performance, MB performed the analysis, and MB and JAR wrote the manuscript. All authors read and approved the manuscript.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Consent for Publication** All authors declare that they have read and approved the manuscript version to be published.

## References

- Albert VA, Backlund A, Bremer K et al (1994) Functional constraints and *rbcL* evidence for land plant phylogeny. *Ann Missouri Bot Gard* 81:534–567. <https://doi.org/10.2307/2399902>
- Alpert P, Oechel WC (1987) Comparative patterns of net photosynthesis in an assemblage of mosses with contrasting microdistributions. *Am J Bot* 74:1787. <https://doi.org/10.2307/2443962>
- Andersson I (2007) Catalysis and regulation in Rubisco. *J Exp Bot* 59:1555–1568. <https://doi.org/10.1093/jxb/ern091>
- Andersson I, Knight S, Schneider G et al (1989) Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature* 337:229–234. <https://doi.org/10.1038/33729a0>
- Calabrese GM, Muñoz J (2006) *Zygodon bistratus* sp. nov. (Orthotrichaceae) from the Iberian Peninsula. *Bryologist* 109:38–42. [https://doi.org/10.1639/0007-2745\(2006\)109\[0038:ZBSNO\]2.0.CO;2](https://doi.org/10.1639/0007-2745(2006)109[0038:ZBSNO]2.0.CO;2)
- Cameron KM, Chase MW, Whitten WM et al (1999) A phylogenetic analysis of the Orchidaceae: evidence from *rbcL* nucleotide sequences. *Am J Bot* 86:208–224. <https://doi.org/10.2307/2656938>
- Capella-Gutierrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Chase MW, Soltis DE, Olmstead RG et al (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann Missouri Bot Gard* 80:528–580. <https://doi.org/10.2307/2399846>
- Codoñer FM, Fares MA (2008) Why should we care about molecular coevolution? *Evol Bioinform* 4:117693430800400000. <https://doi.org/10.1177/117693430800400003>
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772–772. <https://doi.org/10.1038/nmeth.2109>
- de Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. *Nat Rev Genet* 14:249–261. <https://doi.org/10.1038/nrg3414>
- Dunn SD, Wahl LM, Gloor GB (2008) Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. *Bioinformatics* 24:333–340. <https://doi.org/10.1093/bioinformatics/btm604>
- Erb TJ, Zarzycki J (2018) A short history of RubisCO: the rise and fall (?) of Nature's predominant CO<sub>2</sub> fixing enzyme. *Curr Opin Biotechnol* 49:100–107. <https://doi.org/10.1016/j.copbio.2017.07.017>
- Erickson DL, Reed E, Ramachandran P et al (2017) Reconstructing a herbivore's diet using a novel *rbcL* DNA mini-barcode for plants. *AoB Plants* 9:plx015. <https://doi.org/10.1093/aobpla/plx015>
- Fares MA, McNally D (2006) CAPS: coevolution analysis using protein sequences. *Bioinformatics* 22:2821–2822. <https://doi.org/10.1093/bioinformatics/btl493>
- Fares MA, Travers SA (2006) A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics* 173:9–23. <https://doi.org/10.1534/genetics.105.053249>
- Galmés J, Andralojc PJ, Kapralov MV et al (2014) Environmentally driven evolution of Rubisco and improved photosynthesis and growth within the C3 genus *Limonium* (Plumbaginaceae). *New Phytol* 203:989–999. <https://doi.org/10.1111/nph.12858>
- Garilleti R, Mazimpaka V, Lara F (2015) *Ulota larrainii* (Orthotrichoideae, Orthotrichaceae, Bryophyta) a new species from Chile, with comments on the worldwide diversification of the genus. *Phytotaxa* 217:133. <https://doi.org/10.11646/phytotaxa.217.2.3>
- Gastony GJ, Rollo DR (1995) Phylogeny and generic circumscriptions of cheilanthoid ferns (Pteridaceae: Cheilanthesoideae) inferred from *rbcL* nucleotide sequences. *Am Fern J* 85:341–360. <https://doi.org/10.2307/1547814>
- Glime JM (2007) Physiological ecology. In: *Bryophyte ecology*, vol 1. Michigan Technological University and the International Association of Bryologists, Houghton
- Goffinet B, Vitt DH (1998) Revised generic classification of the Orthotrichaceae based on a molecular phylogeny and comparative morphology. In: Wates JW, Ashton NW, Duckett JG (eds) *Bryology for the twenty-first century*, 1st edn. British Bryological Society, London, pp 143–159

- Goffinet B, Bayer RJ, Vitt DH (1998) Circumscription and phylogeny of the Orthotrichales (Bryopsida) inferred from *rbcL* sequence analyses. *Am J Bot* 85:1324–1337. <https://doi.org/10.2307/2446642>
- Goffinet B, Cox CJ, Shaw AJ, Hedderson TAJ (2001) The Bryophyta (mosses): systematic and evolutionary inferences from an *rps4* gene (cpDNA) phylogeny. *Ann Bot* 87:191–208. <https://doi.org/10.1006/ANBO.2000.1318>
- Goffinet B, Shaw AJ, Cox CJ et al (2004) Phylogenetic inferences in the Orthotrichoideae (Orthotrichaceae: Bryophyta) based on variation in four loci from all genomes. *Monogr Syst Bot Missouri Bot Gard* 98:270–289
- Gong L, Olson M, Wendel J (2014) Cytonuclear evolution of Rubisco in four allopolyploid lineages. *Mol Biol Evol* 31:2624–2636. <https://doi.org/10.1093/molbev/msu207>
- Green TGA, Lange OL (1994) Photosynthesis in poikilohydric plants: a comparison of Lichens and Bryophytes. In: Schulze ED, Caldwell MM (eds) *Ecophysiology of photosynthesis*. Springer, Berlin, pp 319–341
- Group CPW (2009) A DNA barcode for land plants. *Proc Natl Acad Sci* 106:12794–12797
- Guindon S, Dufayard JF, Lefort V et al (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
- Guo SL, Ma YH, Cao T, Lou YX (2012) A synopsis of *Macromitrium* (Orthotrichaceae) in China. *Cryptogam Bryol* 33:341–355. <https://doi.org/10.7872/cryb.v33.iss4.2012.341>
- Hasebe M, Wolf PG, Pryer KM et al (1995) Fern phylogeny based on *rbcL* nucleotide sequences. *Am Fern J* 85:134–181. <https://doi.org/10.2307/1547807>
- Hermida-Carrera C, Kapralov MV, Galmés J (2016) Rubisco catalytic properties and temperature response in crops. *Plant Physiol* 171:2549–2561. <https://doi.org/10.1104/pp.16.01846>
- Hermida-Carrera C, Fares MA, Fernández Á et al (2017) Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. *PLoS ONE* 12:e0183970. <https://doi.org/10.1371/journal.pone.0183970>
- Hermida-Carrera C, Fares MA, Font-Carrascosa M et al (2020) Exploring molecular evolution of Rubisco in C3 and CAM Orchidaceae and Bromeliaceae. *BMC Evol Biol* 20:1–17. <https://doi.org/10.1186/s12862-019-1551-8>
- Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol* 33:1635–1638. <https://doi.org/10.1093/molbev/msw046>
- Iida S, Miyagi A, Aoki S et al (2009) Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. *PLoS ONE* 4:e4633. <https://doi.org/10.1371/journal.pone.0004633>
- Kannappan B, Gready JE (2008) Redefinition of Rubisco carboxylase reaction reveals origin of water for hydration and new roles for active-site residues. *J Am Chem Soc* 130:15063–15080. <https://doi.org/10.1021/ja803464a>
- Kapralov MV, Filatov DA (2006) Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS ONE* 1:e8. <https://doi.org/10.1371/journal.pone.0000008>
- Kapralov MV, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol* 7:73. <https://doi.org/10.1186/1471-2148-7-73>
- Kapralov MV, Smith JAC, Filatov DA (2012) Rubisco evolution in C4 eudicots: an analysis of Amaranthaceae sensu lato. *PLoS ONE* 7:e52974. <https://doi.org/10.1371/journal.pone.0052974>
- Katoh K, Rozewicki J, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:bbx108. <https://doi.org/10.1093/bib/bbx108>
- Keck F, Rimet F, Bouchez A, Franc A (2016) phyloSignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol Evol* 6:2774–2780. <https://doi.org/10.1002/ece3.2051>
- Kellogg EA, Juliano ND (1997) The structure and function of RuBisCO and their implications for systematic studies. *Am J Bot* 84:413–428. <https://doi.org/10.2307/2446015>
- Lara F, Garilletti R, Goffinet B et al (2016) *Lewinskya*, a new genus to accommodate the phaneroporous and monoicous taxa of *Orthotrichum* (Bryophyta, Orthotrichaceae). *Cryptogam Bryol* 37:361–382. <https://doi.org/10.7872/cryb/v37.iss4.2016.361>
- Larson EM, O'Brien CM, Zhu G et al (1997) Specificity for activase is changed by a Pro-89 to Arg substitution in the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Biol Chem* 272:17033–17037. <https://doi.org/10.1074/jbc.272.27.17033>
- Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol* 50:913–925. <https://doi.org/10.1080/106351501753462876>
- Liu L, Zhao B, Zhang Y, Wang J (2012) Adaptive evolution of the *rbcL* gene in Brassicaceae. *Biochem Syst Ecol* 44:13–19. <https://doi.org/10.1016/j.bse.2012.04.007>
- Liu Y, Johnson MG, Cox CJ et al (2019) Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nat Commun* 10:1485. <https://doi.org/10.1038/s41467-019-09454-w>
- Maddison WP, Maddison DR (2019) Mesquite: a modular system for evolutionary analysis. Version 3.61 <http://www.mesquiteproject.org>.
- Manhart JR (1994) Phylogenetic analysis of green plant *rbcL* sequences. *Mol Phylogenet Evol* 3:114–127. <https://doi.org/10.1006/mpev.1994.1014>
- Massingham T, Goldman N (2005) Detecting amino acid sites under positive selection and purifying selection. *Genetics* 169:1753–1762. <https://doi.org/10.1534/genetics.104.032144>
- Masuzaki H, Shimamura M, Furuki T et al (2010) Systematic position of the enigmatic liverwort *Mizutania* (Mizutaniaceae, Marchantiophyta) inferred from molecular phylogenetic analyses. *Taxon* 59:448–458. <https://doi.org/10.1002/tax.592010>
- Matcham HW, O'Shea BJ (2005) A review of the genus *Codonoblepharon* Schwgr. (Bryopsida: Orthotrichaceae). *J Bryol* 27:129–135. <https://doi.org/10.1179/037366805X53040>
- Miwa H, Odrzykoski IJ, Matsui A et al (2009) Adaptive evolution of *rbcL* in *Conocephalum* (Hepaticae, Bryophytes). *Gene* 441:169–175. <https://doi.org/10.1016/j.gene.2008.11.020>
- Mizia P, Myszczynski K, Ślipiko M et al (2019) Comparative plastomes analysis reveals the first infragenetic evolutionary hotspots of *Orthotrichum* s.l. (Orthotrichaceae, Bryophyta). *Turk J Bot* 43:444–457. <https://doi.org/10.3906/bot-1811-13>
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936. <https://doi.org/10.1021/ja803464a>
- Norris DH, Shevock JR, Goffinet B (2004) *Orthotrichum kellmanii* (Bryopsida, Orthotrichaceae), a remarkable new species from the central coast of California. *Bryologist* 107:209–214. [https://doi.org/10.1639/0007-2745\(2004\)107](https://doi.org/10.1639/0007-2745(2004)107)
- Ott CM, Smith BD, Portis AR, Spreitzer RJ (2000) Activase region on chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase: nonconservative substitution in the large subunit alters species specificity of protein interaction. *J Biol Chem* 275:26241–26244. <https://doi.org/10.1074/jbc.M004580200>
- Page M (1999) Inferring the historical patterns of biological evolution. *Nature* 401:877–884. <https://doi.org/10.1038/44766>
- Pie MR (2006) The influence of phylogenetic uncertainty on the detection of positive Darwinian selection. *Mol Biol Evol* 23:2274–2278. <https://doi.org/10.1093/molbev/msl116>
- Portis AR (2003) Rubisco activase—Rubisco's catalytic chaperone. *Photosynth Res* 75:11–27. <https://doi.org/10.1023/A:1022458108678>

- Portis AR, Li C, Wang D, Salvucci ME (2007) Regulation of Rubisco activase and its interaction with Rubisco. *J Exp Bot* 59:1597–1604. <https://doi.org/10.1093/jxb/erm240>
- Proctor MCF (1979) Structural and eco-physiological adaptation in Bryophytes. In: Clarke GC, Duckett J (eds) *Bryophyte systematics*. Academic Press, London, pp 479–509
- Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 3:217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Schluter D, Price T, Mooers AØ, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. *Evolution (NY)* 51:1699–1711. <https://doi.org/10.1111/j.1558-5646.1997.tb05095.x>
- Sen L, Fares MA, Liang B et al (2011) Molecular evolution of *rbcL* in three gymnosperm families: identifying adaptive and coevolutionary patterns. *Biol Direct* 6:29. <https://doi.org/10.1186/1745-6150-6-29>
- Shaw AJ, Devos N, Liu Y et al (2016) Organellar phylogenomics of an emerging model system: *Sphagnum* (peatmoss). *Ann Bot* 118:185–196. <https://doi.org/10.1093/aob/mcw086>
- Spreitzer RJ, Salvucci ME (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol* 53:449–475. <https://doi.org/10.1146/annurev.arplant.53.100301.135233>
- Stenøien HK (2008) Slow molecular evolution in 18S rDNA, *rbcL* and *nad5* genes of mosses compared with higher plants. *J Evol Biol* 21:566–571. <https://doi.org/10.1111/j.1420-9101.2007.01479.x>
- Studer RA, Christin PA, Williams MA, Orengo CA (2014) Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci USA* 111:2223–2228
- Suchard MA, Lemey P, Baele G et al (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* 4:vey016. <https://doi.org/10.1093/ve/vey016>
- Tomescu AMF, Bomfleur B, Bippus AC, Savoretti A (2018) Why are Bryophytes so rare in the fossil record? A spotlight on taphonomy and fossil preservation. In: Krings M, Harper CJ, Cuneo NR, Rothwell GW (eds) *Transformative paleobotany*, 1st edn. Academic Press, London, pp 375–416
- Tsubota H, De Luna E, González D et al (2004) Molecular phylogenetics and ordinal relationships based on analyses of a large-scale data set of 600 *rbcL* sequences of mosses. *Hikobia* 14:149–170
- Wang M, Kapralov MV, Anisimova M (2011) Coevolution of amino acid residues in the key photosynthetic enzyme Rubisco. *BMC Evol Biol* 11(1):266. <https://doi.org/10.1186/1471-2148-11-266>
- Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 46:W29–W303. <https://doi.org/10.1093/nar/gky427>
- Wong SL, Zhang LV, Tong AHY et al (2004) Combining biological networks to predict genetic interactions. *Proc Natl Acad Sci U S A* 101:15682–15687. <https://doi.org/10.1073/pnas.0406614101>
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang Z, Nielsen R, Goldman N, Pedersen A (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107–1118. <https://doi.org/10.1093/molbev/msi097>
- Yao X, Tan Yh, Jb Y et al (2019) Exceptionally high rates of positive selection on the *rbcL* gene in the genus *Ilex* (Aquifoliaceae). *BMC Evol Biol* 1:192. <https://doi.org/10.1186/s12862-019-1521-1>