

RESEARCH ARTICLE

Open Access



Exceptionally high rates of positive selection on the *rbcl* gene in the genus *Ilex* (Aquifoliaceae)

Xin Yao^{1,2*} , Yun-hong Tan^{1,2,3}, Jun-bo Yang⁴, Yan Wang⁵, Richard T. Corlett^{1,2*} and Jean-François Manen^{6*}

Abstract

Background: The genus *Ilex* (Aquifoliaceae) has a near-cosmopolitan distribution in mesic habitats from tropical to temperate lowlands and in alpine forests. It has a high rate of hybridization and plastid capture, and comprises four geographically structured plastid groups. A previous study showed that the plastid *rbcl* gene, coding for the large subunit of Rubisco, has a particularly high rate of non-synonymous substitutions in *Ilex*, when compared with other plant lineages. This suggests a strong positive selection on *rbcl*, involved in yet unknown adaptations. We therefore investigated positive selection on *rbcl* in 240 *Ilex* sequences from across the global range.

Results: The *rbcl* gene shows a much higher rate of positive selection in *Ilex* than in any other plant lineage studied so far (> 3000 species) by tests in both PAML and SLR. Most positively selected residues are on the surface of the folded large subunit, suggesting interaction with other subunits and associated chaperones, and coevolution between positively selected residues is prevalent, indicating compensatory mutations to recover molecular stability. Coevolution between positively selected sites to restore global stability is common.

Conclusions: This study has confirmed the predicted high incidence of positively selected residues in *rbcl* in *Ilex*, and shown that this is higher than in any other plant lineage studied so far. The causes and consequences of this high incidence are unclear, but it is probably associated with the similarly high incidence of hybridization and introgression in *Ilex*, even between distantly related lineages, resulting in large cytonuclear discordance in the phylogenies.

Keywords: Rubisco, Positive selection, PALM, Environmental adaptation, Molecular adaptation

Background

The genus *Ilex* (hollies) constitutes the monogeneric family Aquifoliaceae in the campanulid subclade of the asterids [2]. *Ilex* species are deciduous or evergreen dioecious trees or shrubs of mesic temperate, subtropical, tropical, and tropical-montane climates. At least 644 species have been described. *Ilex* has a near-cosmopolitan distribution, with the highest diversities in East-Asia, and in Central and South America, moderate numbers of species in Southeast Asia and North America, and a few in Africa, Australia,

and Europe, and on most remote volcanic islands in the Atlantic and Pacific, including the Canary Islands, Madeira, Azores, Caribbean, Polynesia, Hawaii, New Caledonia, and Fiji [36]. The lineage is old, dating from the end of Cretaceous [36, 38], with the oldest accepted fossil record of 69 Million Years Ago (MYA, [31]). However, there is a large gap between this date and the estimated Miocene age of the most recent common ancestor of all extant *Ilex*, based on plastid and nuclear markers [36]. Many pre-Miocene fossils have been attributed to *Ilex* ([34]; [33]), suggesting extensive lineage extinctions, so that the extant genus is now phylogenetically isolated.

Abundant production of small, fleshy fruits, consumed primarily by birds in most species, can account for its wide dispersal, but the exceptional adaptive capacity in this morphologically relatively uniform genus is surprising. There is evidence for a high frequency of hybridization and

* Correspondence: yaoxin@xtbg.org.cn; corlett@xtbg.org.cn; jean-francois.manen@wanadoo.fr

¹Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China

⁶Laboratoire de Systématique Végétale et Biodiversité, University of Geneva (retired), Chemin de l'Impératrice 1, CH-1292 Chambésy, Switzerland
Full list of author information is available at the end of the article



introgression in *Ilex*, leading to a reticulate pattern of evolution. Phylogenetic trees obtained from nuclear and plastid data are very different [36] and Shi et al. [49] suggest that many of the 150–200 narrow-range species recognized in China may be a result of hybridization, which may also be true elsewhere. Widespread hybridization is supported by anatomical and morphological studies [3], and documented by molecular phylogenies [36–38, 47, 49].

A feature of the genus *Ilex* is that the *rbcl* gene, encoding the large subunit of Rubisco and a widely-used plastid sequence in phylogeny, appears to have a particularly high rate of non-synonymous substitutions compared with other genera [39]. This high rate of non-synonymous substitution suggests a positive selection on the plastid-encoded Rubisco large subunit in *Ilex*. Rubisco catalyzes the first steps of photosynthetic fixation of carbon and of photorespiration in all photosynthetic organisms (see [4, 10, 44], for recent reviews). The most prevalent form of Rubisco (form I) is an hexadecamer of 8 catalytic large subunits (L, harboring the active site) encoded by the *rbcl* gene in the plastome, and 8 regulatory small subunits (S), encoded in the nucleosome by a small family of *rbcS* genes [6]. The assembly and enzymatic function of Rubisco are, in part, guaranteed by specific interactions of the L subunit with the S subunits, assisted by several nuclear-encoded specific chaperones [22, 23].

Several studies provide evidence that positive selection on particular amino-acid residues of the Rubisco large subunit is involved in plant adaptation to various environmental stresses. This was proposed for the endemic Hawaiian genus *Schiedea* adapting to dry or wet habitats [28], in the heterophyllous aquatic plant *Potamogeton* [26], in switches from C3 to C4 photosynthesis [7, 30], in movements from dry to wet habitats in the Balearic Islands [15], in adaptation of *Cardamine resedifolia* to high altitudes [25], and associated with leaf traits and climate characteristics in the genus *Quercus* [24]. Moreover, studies based on a large sample of *rbcl* sequences show that positive selection is widespread in most lineages of land plants [21, 29].

A consequence of *rbcl* positive selection is a frequently observed pattern of coevolution between positively selected sites. Indeed, the multiple roles of amino acids in enzymes (folding, stability, enzymatic activity) imply that the modification of one parameter can affect other properties. Stability and activity of the enzyme are likely to be negatively correlated, and compensatory mutations are then needed to restore global stability [50]. Thus, when positive selection occurs on a residue, a compensatory co-evolving mutation on another residue of the molecule is likely to be detected. This has been confirmed in land plants [24, 25, 46, 53]. Rubisco has evolved CO₂/O₂ specificity ratios according to the environmental situation, allowing a perpetual fine-tuning of its performance [50]. Although the plastid *rbcl*

gene sequence is highly conserved, this enzyme has played a role in plant adaptation to changing environments across geological time and continues to do so across recent geographical space.

Another feature of adaptive mutations in *rbcl* is that during introgression, plastid-encoded maternal L subunits of one species become associated with nuclear-encoded paternal S subunits of another species, potentially creating compatibility problems [48]. Similar compatibility problems could also arise from the numerous nuclear-encoded chaperones intimately linked to Rubisco subunits ([52]; [18]). The impacts and repair of such incompatibilities on Rubisco function should be most significant in lineages experiencing high rates of hybridization or/and introgression, as it is the case in *Ilex* [36].

In this study, we looked for evidence of positive selection on *rbcl* in the genus *Ilex*. We first investigate *rbcl* sequences (truncated after codon position 436) of 116 species from across the global range (“Worldwide dataset”) and then untruncated *rbcl* sequences of 124 accessions collected in natural habitats in China (“Chinese dataset”).

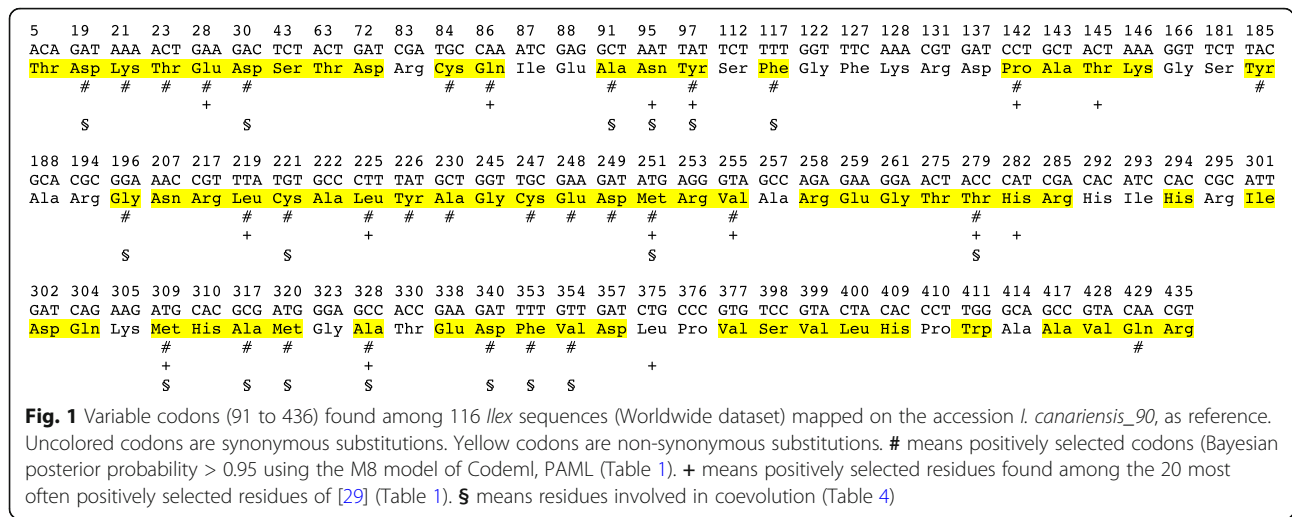
Results

Evolution of the gene *rbcl* of *Ilex* includes all the features of positive selection

Positive selection on the *rbcl* gene of *Ilex*

The Worldwide dataset of 116 *rbcl* sequences (see [Materials and Methods](#) and Additional file 1: Table S1) included the first 436 codons. There were 91 variable codons, of which 24 had synonymous substitutions and 67 had non-synonymous substitutions (Fig. 1). In order to speed up the analyses, PAML calculations were conducted on *rbcl* sequences of the 4 *Ilex* plastid lineages [36] separately: the Asian/North-American Group 1 (14 sequences), the Asian/North-American Group 2 (36 sequences), the American Group 3 (31 sequences), and the Eurasian Group 4 (35 sequences). A 5th “20-*Ilex*” *rbcl* DNA matrix consisted of 20 selected species of *Ilex* representing the 4 plastid groups and a wide geographic distribution (see Fig. 3).

Positively selected sites were detected (at a Bayesian posterior probability of 0.95) by the program Codeml of PAML using model M8 [54]. Altogether for the five alignments studied, 32 positively selected residues were detected (Table 1). However, the SLR (Stepwise likelihood Ratio) test only confirmed 17 of these (Table 1). Using PAML, likelihood ratio tests (LRT) were done for each of the 4 alignments representing the 4 previously defined phylogenetic plastid lineages. Likelihood ratio tests of different models (model M2a allowing $\omega > 1$ versus models M1a not allowing $\omega > 1$, as well as model M8 allowing $\omega > 1$ versus models M7 or M8a not allowing $\omega > 1$) indicated that positive selection was highly probable for the 4 plastid groups at a significance level of < 0.001 :



- Group 1, 2Δl = 42.88 (M2a/M1a), 49.68 (M8/M7), and 40.92 (M8/M8a).
- Group 2, 2Δl = 113.94 (M2a/M1a), 111.20 (M8/M7), and 111.04 (M8/M8a).
- Group 3, 2Δl = 52.14 (M2a/M1a), 52.16 (M8/M7), and 52.12 (M8/M8a).
- Group 4, 2Δl = 175.96 (M2a/M1a), 60.00 (M8/M7), and 42.74 (M8/M8a).

Positively selected sites were mainly found in the Asian/North American Group 2 (25 sites by PAML, 11 confirmed by SLR test), followed by the American Group 3 (12 sites, 2 confirmed), the Asian/North American Group 1 (7 sites, 3 confirmed), and then Eurasian Group 4 (7 sites, 1 confirmed) (Table 1).

A total of 25 positively selected residues were detected by PAML in the *rbcl* coding sequences of the Chinese dataset of 124 sequences (see materials and methods and Additional file 2: Table S2), six of which were new, of which residues 461 and 470 were unavailable in the Worldwide dataset truncated at residue 436 (Table 2). The SLR test confirmed 9 of these residues as positively selected. The Asian/North-American Group 2 again had the most positively selected residues (Table 2).

Most positively selected sites are located at the surface of the folded *rbcl*

The 38 positively selected sites in *rbcl* detected by PAML (32 in the Worldwide dataset and 6 additional ones in the Chinese dataset) were mapped onto the 3-D structure of the modeled Rubisco large subunit of *Ilex canariensis* (see Materials and Methods). A large proportion of positively selected residues are located at the surface of the folded large subunit, with only 6 residues (53, 84, 196, 221, 317 and 328) being buried (Fig. 2a). Exposed residues could be candidates for interactions with other large subunits (intra-dimer and dimer-dimer

interaction), with small subunits, and with Rubisco activase and other chaperones, as suggested by Kapralov and Filatov [29]. Extrapolating from the *Spinacia oleracea* Rubisco 3D-structure (1RCX, [51]), potential implications of *Ilex* positively selected residues in quaternary interactions were examined. Positively selected residues located within 6 Å (the van der Waals contact) of the surface of other subunits were selected using Swiss-Pdb Viewer (<http://swissmodel.expasy.org>). Some of these (13 of 38) were in van der Waals contact with another large subunit intra-dimer interface, a dimer-dimer interface, and/or with a small subunit (Table 3). Many positively selected residues of the equatorial surface of Rubisco were still visible (and potentially accessible by chaperones) even when the 4 interacting large subunits (B, E, R and V) and the 3 interacting small subunits (s, f and w) were represented in contact with the large L subunit (Fig. 2b). Positively selected residues 219, 225, 226, 429, and 461 of the large subunits were in contact with the small subunits (Table 3). The same was true for the numerous nuclear-encoded Rubisco chaperones that might interact with the many exposed positively selected residues on the equatorial surface of Rubisco (15, 19, 21, 23, 28, 30, 84, 86, 91, 95, 97, 340, 353, 354 and 470, see Fig. 2b), that were not in contact with any other L or S subunits (see for instance [22, 23]).

As expected, positively selected sites of the *rbcl* gene are extensively coevolving

Selecting only pairwise correlations above 0.5, CAPS detected 17 (out of 32) positively selected residues of the *rbcl* Worldwide dataset that were involved in coevolution (Table 4). Except for residue 95, all the others were detected as positively selected residues in the Worldwide dataset. However, residue 95 was detected as positively selected in the Chinese dataset (Table 2). The mean

Table 1 List of positively selected sites of *rbcL* of *Ilex* with posterior probabilities (PAML, Codeml, model M8) > 0.95 in the 5 subsets of the Worldwide dataset. Groups 1, 2, 3 and 4 represent the 4 phylogenetic plastid groups of [36] (Asian/North-American Group 1, Asian/North/American Group 2, American Group 3, and Eurasian Group 4). The 20-species sample represents a selection of the 116 specimens of [36], reflecting the 4 groups and a wide geographic distribution (see Fig. 3). In red, selected *Ilex* residues belonging to the list of the 20 most often positively selected sites detected in the 151 lineages of Kapralov and Filatov [29]

Sites	Group 1	Group 2	Group 3	Group 4	20-species
(a)	14	36	31	35	20
19		1.00			1.00
21		0.99			
23		1.00*			1.00*
28		1.00*	0.95		1.00
30		1.00*	0.95	0.98	1.00*
84		0.99			1.00
86		0.99			0.99
91		0.99			
97		1.00	1.00	0.98	1.00
117		0.99			1.00
142					1.00
185	0.99				1.00
196	0.98	1.00			
219		1.00*			1.00
221	1.00*	0.99			0.99
225		1.00*			1.00
226			1.00*	0.98	1.00*
230			1.00		0.99
247				0.99*	
248			0.99		
249	1.00*	1.00			1.00
251	1.00*	0.99	0.96		1.00
255		1.00*	0.95		1.00*
279		1.00*	0.98		1.00
309	0.98	1.00*			0.99
317					1.00
320		1.00*	1.00		
328		1.00*		0.98	1.00*
340	0.98	1.00*	1.00*		1.00*
353		0.99	1.00		1.00
					20
354		1.00		0.98	1.00*
429		1.00		0.98	1.00

(*) indicates residues having $\geq 95\%$ SLR support for positive selection

(a) Number of sequences

Table 2 List of positively selected residues of *rbcl* with posterior probabilities (PAML, Codeml, Model M8) > 0.95 in the Chinese dataset. This dataset represents group 1, 2, and 4 (The American Group 3 is not present in China). In red, selected *Ilex* residues belonging to the list of 20 most often positively selected sites detected in the 151 lineages of Kapralov and Filatov [29]. (*) indicates that the residues have $\geq 95\%$ SLR support for positive selection. (a) indicates that residues are not positively selected in the Worldwide dataset or that residues (461 and 470) were not available in the truncated Worldwide dataset. (b) Number of residues

Residues (b)	Group 1 39	Group 2 40	Group 4 45
15 ^a		0.98	
19		0.98	
28		1.00*	
30	0.98	0.98	0.98
53 ^a			0.98
86	1.00*	1.00	
95 ^a		0.97	
97		0.98	
142		0.98	
219		0.99	
225		0.98	
226			0.97
247			0.99*
249	1.00*	1.00	
251	1.00*	0.99	
255		1.00	
279	0.98	0.99	
281 ^a	0.98		
309	1.00*	0.99	
328	0.99*	1.00*	
340		1.00*	
354		0.98	
429		1.00	
461 ^a		0.99	
470 ^a		1.00	

pairwise distance of coevolving pairs in the 3-D structure of the Rubisco large subunit was 22.9 Å.

Compared with other lineages, *Ilex* has the highest rate of positive selection

Using the M8 model of PAML, positive selection on the *rbcl* gene of *Ilex* was compared with 8 lineages studied by Kapralov and Filatov [29] using recalculations based

on their data. Eight of their lineages potentially have high rates of positive selection and were chosen for a comparison with *Ilex* (see their additional file 2). Three (out of 151) lineages (eurosids I-26, euasterids I-9 and Coniferales 6) had dN/dS values (in the “eleventh class” of the M8 model) higher than 20.5 (the value found in *Ilex*). Five (out of 151) other lineages (commelinids-7, commelinids-15, eurosids I-24, euasterids I-8 and Gnetales-1) had a higher proportion of detected positive

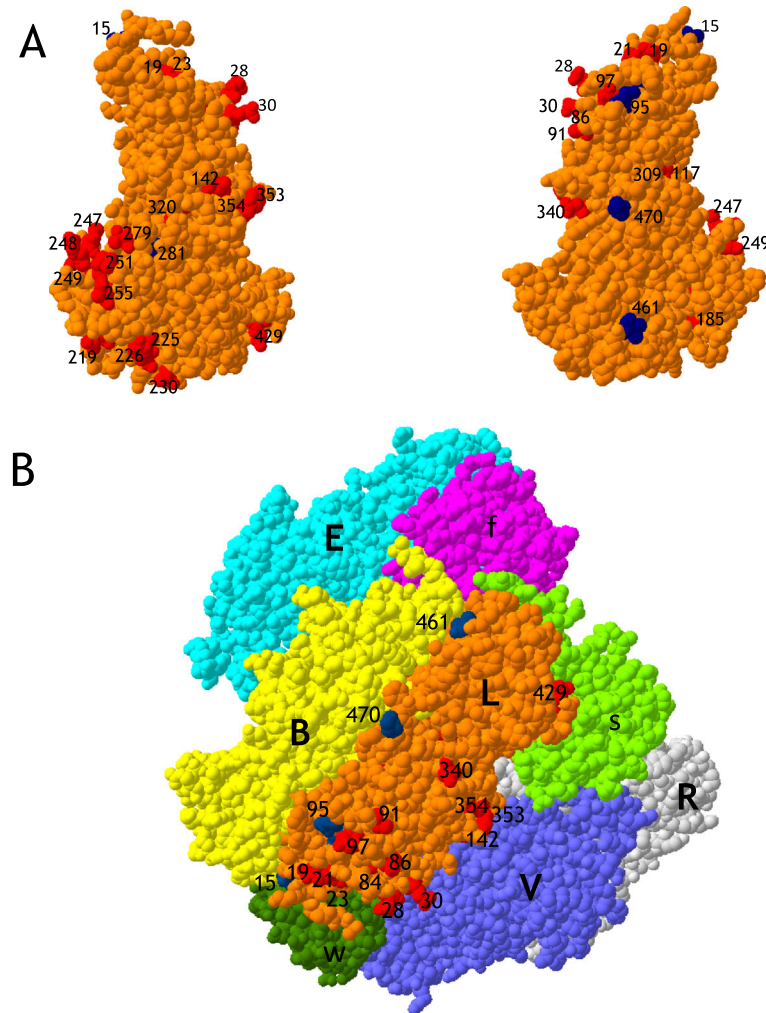


Fig. 2 Positively selected residues of *Ilex* represented on the surface of the Rubisco large subunit. **a** Front and back of folded large subunit of the modeled *Ilex canariensis* 3-D structure. Visible positively selected residues from the Worldwide dataset are in red, while additional positively selected residues from the Chinese dataset are in blue. **b** Extrapolation on the Rubisco *Spinacia oleracea* 3-D structure (1RCX). Presented are the 4 large subunits (E,B,V and R) and 3 small subunits (f, s and w) that are in contact with one large subunit (here L, in orange). Positively selected residues are represented with the same colors as in A. Most of these residues (represented here on the L subunit) are exposed on the equatorial surface the L_8S_8 Rubisco hexadecamer. Sites 142, 429 and 461 are involved in intra-dimer, dimer-dimer and L/S interactions (see Table 3). All other sites (15, 19, 21, 23, 28, 30, 84, 86, 91, 95, 97, 340, 353, 354 and 470) represent positively selected residues potentially involved in interactions with chaperones

selection sites (in the “eleventh class” of the M8 model) than 8.2% (the proportion found in *Ilex*).

PAML calculations indicated that the *rbcl* gene of the 20-species dataset of *Ilex* had 26 positively selected residues, while the largest number in the other 8 lineages was 15 in the commelinids-15 lineage (Table 5). All the others had < 10. The SLR test confirmed 7 positively selected residues in the *Ilex* 20-species dataset, while the highest number found in the other 8 lineages was 2 in the Coniferales-6 lineage. A comparison with other angiosperm lineages (Table 5), shows that the number of positively selected residues of *rbcl* in the Worldwide dataset of *Ilex* (32) (Bayesian posterior probability: 0.95)

was never reached in any lineage studied by Kapralov and Filatov [29]. In their 151 studied datasets, the number of positively selected residues never passed 16 (see their additional file 3).

The program CodeML of PAML was used in the above comparisons. It should be noted, however, that using the Site Likelihood-Ratio (SLR) test, a measure of the strength of evidence for selection, the number of positively selected residues remaining significant at 95% after the SLR adjustment was 17 (among the 32 residues found by PAML in the Worldwide dataset of *Ilex*, see Table 1). A similar decrease was generally observed in other studies when PAML results were adjusted by SLR

Table 3 Positively selected residues of the large subunit of *Ilex* Rubisco that are in contact with another L or S subunit, extrapolated from the *Spinacia oleracea* Rubisco 3D-structure (see Fig. 2). Thirteen positively selected residues of the Rubisco large subunit of *Ilex* are at the interface ($\leq 6 \text{ \AA}$) of intra-dimer (ID), of dimer-dimer (DD), and of large/small subunits (SSU). Two of them (219 and 461) share a contact with another large and with a small subunit. In brackets, the total number of residues of the Rubisco large subunit that are in contact with another large subunit or the small subunit in ID, DD and SSU interactions

Residues	Molecular interactions		
	ID (96)	DD (39)	SSU (76)
117	X		
142		X	
219		X	X
225			X
226			X
247	X		
248	X		
249	X		
251	X		
279	X		
309	X		
429			X
461	X		X

Table 4 Pairs of coevolving residues detected by CAPS (Coevolution Analysis using Protein Sequences) in 116 *rbcl* sequences of the *Ilex* Worldwide dataset. Parameters used in CAPS are: alpha threshold 0.001, number of simulated alignments 1000, bootstrap value 0.95, await convergence and time correction on

Codon 1	Codon 2	Correlation	Bootstrap	Distance in \AA
D 19	G 196	0.51	0.99	68.6
D 30	G 196	0.50	0.99	57.2
A 91	N 95	0.99	0.97	8.2
Y 97	M 309	0.70	1.00	12.8
F 117	T 279	0.67	0.98	13.2
C 221	M 251	0.51	1.00	20.9
M 309	A 328	0.70	1.00	19.7
A 317	M 320	0.62	0.98	6.7
A 328	D 340	0.78	1.00	16.9
A 328	V 354	0.70	1.00	23.9
D 340	V 354	0.70	1.00	21.0
F 353	V 354	0.73	1.00	5.5

(see for instance [27]). SLR adjustment did not change the major results of this study. The 17 positively selected residues in the Rubisco large subunit of *Ilex* detected by SLR still exceeded the maximum of 16 found in a broad survey of green plants by Kapralov and Filatov [29] using PAML.

Seven additional lineages (Table 5 and Additional file 3: Table S3) were also compared because they are trees (as are most *Ilex*) that have experienced frequent hybridization-introgression (*Nothofagus*, *Populus*, *Quercus*, *Salix*) or because of reported positive selection (*Flaveria*, *Potamogeton*, *Schiedea*). Positively selected sites are much more frequent in *Ilex* than in any other lineage studied (Table 5).

Discussion

A very high rate of positive selection on the *rbcl* gene of *Ilex*

This study has confirmed that the plastid gene *rbcl*, encoding the large subunit of Rubisco, has experienced an exceptionally high rate of positive selection in the genus *Ilex*. Kapralov and Filatov [29] listed the 20 most often positively selected *rbcl* residues, which accounted for > 70% of the cases of positive selection in their broad survey of green plants. Twelve positively selected residues detected in *Ilex* belonged to this top-20 list. Several residues detected in other lineages were also identified as positively selected in *Ilex* (Table 6). Thus, a large proportion of the *rbcl* sites positively selected in the green plants as a whole were also detected in the *Ilex*. This implies that the *rbcl* gene has a limited number of residue positions that can be positively selected for the fine-tuning of Rubisco [16], and that many of these have been selected in *Ilex*.

This particularly high rate of positive selection may affect the molecular clock of *rbcl* in *Ilex*, compared with other lineages. A high proportion of non-synonymous substitutions (such as positive selection) is indeed linked to a high rate of nucleotide substitution at the 1st and 2nd codon position compared to the rate of nucleotide substitution at the 3rd position. This last feature explains why the dating of the crown age of the extant genus *Ilex* using a fossil-calibrated molecular clock was at least two times overestimated using 1st and 2nd codon positions than using the 3rd position of *rbcl* [36]. As positive selection seems to be widespread in *rbcl*, a codon effect on age calculation using *rbcl* may be frequent ([45]; [35]), and this is particularly true in the genus *Ilex*.

Positive selection and environmental traits in *Ilex*

Positive selection on *rbcl* is frequently observed in terrestrial plants but not in aquatic plants, algae, or bacteria [29]. It has been proposed that the thermal and water regime of the aquatic habitat is more stable, while land

Table 5 Comparison of positive selection in *rbcl* of different lineages detected by PAML (M8 model) and by SLR (probability ≥ 0.95)

Lineages	N. seq ^a	dN/dS ^b	%p ^c	PALM ^d	SLR ^d
<i>Ilex</i> , the 20-species alignment	20	20.5	8.2	26	7
eurosids I-26 (Fagaceae) ^e	21	62.0 (61.0)	0.4 (0.4)	2 (2)	1
euasterids I-9 (Lamiaceae) ^e	19	27.9 (26.7)	6.3 (6.0)	3 (2)	0
Coniferales 6 (<i>Pinus</i>) ^e	35	13.1 (24.8)	1.6 (1.1)	6 (6)	2
commelinids-7 (Cyperaceae) ^e	27	1.1 (1.2)	12.7 (12.9)	0 (0)	0
Commelinids-15 (Poaceae) ^e	28	2.4 (2.4)	10.0 (9.9)	15 (15)	0
eurosids I-24 (Fagales) ^e	20	1.7 (1.6)	7.4 (8.5)	0 (0)	0
euasterids I-8 (Lamiaceae) ^e	20	2.7 (2.8)	10.7 (10.3)	2 (2)	0
Gnetales-1 (<i>Gnetum</i>) ^e	23	3.7 (3.5)	8.5 (8.6)	3 (3)	0
<i>Flaveria</i>	16	19.2	1.8	5	1
<i>Nothofagus</i>	24	6.2	5.4	7	1
<i>Populus</i>	14	5.1	3.9	6	0
<i>Potamogeton</i>	18	11.6	1.2	1	1
<i>Quercus</i>	30	18.4	0.6	2	1
<i>Salix</i>	28	2.5	9.7	1	0
<i>Schiedea</i>	27	26.1	0.6	0	0

^anumber of sequences^bmean dN/dS, "eleventh class" of the M8 model (PAML)^cproportion of positively selected residues, "eleventh class" of the M8 model (PAML)^dnumber of positively selected sites^ePAML re-calculations from data of Kapralov and Filatov [29], with their published results between brackets

plants must adapt to the variability of their habitat, requiring a tuning of the activity of Rubisco by positive selection. The CO₂/O₂ specificity factor of Rubisco τ (determining the relative rate of photosynthesis and photorespiration) is very sensitive to cellular water availability, CO₂ concentration, and temperature [14]. A relation between *rbcl* positively selected substitution at particular sites and morphological, biological, or climate characteristics of different lineages has been suggested (Table 6).

Ilex has two independent hotspots of speciation, in South America and East Asia, each comprising around 200 *Ilex* species. Both hotspots show high rates of recent speciation associated with, respectively, the final uplift of the Andes (from 5 to 2 MYA [19];, [1] [34];) and the Hengduan Mountains (4 to 3 MYA, [17]). In both of these hotspots, the recent development of high mountains and deep valleys may have accelerated the diversification of *Ilex* species through local vicariance, dispersal, isolation, secondary contact, and ecological speciation events [32, 42]. Asian/North-American Group 2 had the highest rate of positive selection on *rbcl*, but the species of this lineage included in this study do not occupy a wider range of environments than those of other lineages. This suggests that there is no simple relationship between positive selection and adaptability. Indeed, although *Ilex* occupies a wide range of environments, it is absent from areas without year-round soil moisture or with exposure to prolonged winter cold, which may have reduced the pressure for tuning Rubisco activity.

The *rbcl* gene has a limited number of sites that can be positively selected. However, they are enough to suggest that cooperation of several different selected substitution sites can mediate a similar effect on Rubisco conformation. Conversely, one unique selected substitution can have different effects on *rbcl* folding and on Rubisco conformation, depending on the other sites of mutation [50]. Figures 3 shows that a particular selected mutation site of *rbcl* can be detected in different branches of the *Ilex* tree, indicating convergence in different clades and reversion in other, making interpretation rather complex. The fine-tuning of Rubisco is the result of a complex synergy of multiple coevolving substitutions sites. For instance, several positively selected residues (P142T, M309I and A328S) that were often associated to C3-C4 transition (Table 6) were also positively selected in *Ilex*, which is entirely C3. Most studies associating *rbcl* positive selection with a particularly biological or climatic trait involved only 2 or 3 selected mutations that were often linked by coevolution (for instance the pair M309I and A328S, [50]). The large number of positively selected *rbcl* sites found in *Ilex* suggests that other drivers are involved.

Introgression may play a major role in the high rate of *rbcl* positive selection

Hybridization and introgression are pervasive in *Ilex* [36]. Advantageous positively selected residues in *rbcl* might contribute to the survival and spread of individuals and species with introgression. Adaptive positive

Table 6 Positively selected residues detected in *Ilex* (in both the Worldwide and Chinese datasets) also described in other lineages and associated with biological and environment traits

Residues	<i>Ilex</i> ^a	Literature ^b	Biological and environmental traits	Authors
86*	Q to E,G	H,D to E,G	Rubisco kinetics	[16]
	Q to E,G	H to Y	Wet to dry environment	[28]
95*	N to S	N,T to D,S	Leaf traits and climate characteristics	[24]
	N to S	N.A.	Rubisco kinetics	[16]
142*	P to T	P to I,T,V	Specificity O ₂ /CO ₂	[16]
	P to T	P to T	C3 to C4 plants	[50]
	P to T	P to A,T	C4 to C4 plants	[50]
	P to T	N.A.	C3 to C4 plants	[7]
219*	L to V	N.A.	Leaf traits and climate characteristics	[24]
225*	L to I	I to L	<i>Potamogeton</i> submerged, floating and terrestrial leaves	[26]
230	A to T	A to T	Wet to dry environment	[28]
281*	A to S	S to A	<i>Potamogeton</i> submerged, floating and terrestrial leaves	[26]
	A to S	A to S	Rubisco kinetics	[16]
309*	M to I	M to I	C3 to C4 plants	[27]
	M to I	M to I	Specificity O ₂ /CO ₂	[15]
	M to I	M to I	C3 to C4 plants	[50]
	M to I	M to I	C4 to C4 plants	[50]
	M to I	N.A.	C3 to C4 plants	[7]
320	M to L	M to L	Rubisco kinetics	[16]
328*	A to S,G	A to S	Rubisco kinetics	[16]
	A to S,G	A to S	C3 to C4 plants	[50]
	A to S,G	A to S	C4 to C4 plants	[50]
	A to S,G	N.A.	C3 to C4 plants	[7]
	A to S,G	N.A.	Leaf traits and climate characteristics	[24]

**Ilex* positively selected residues belonging to the 20-most often selected residues of Kapralov and Filatov [29]

^aAmino acid substitution in *Ilex*

^bAmino acid substitution in other lineages recorded in the literature

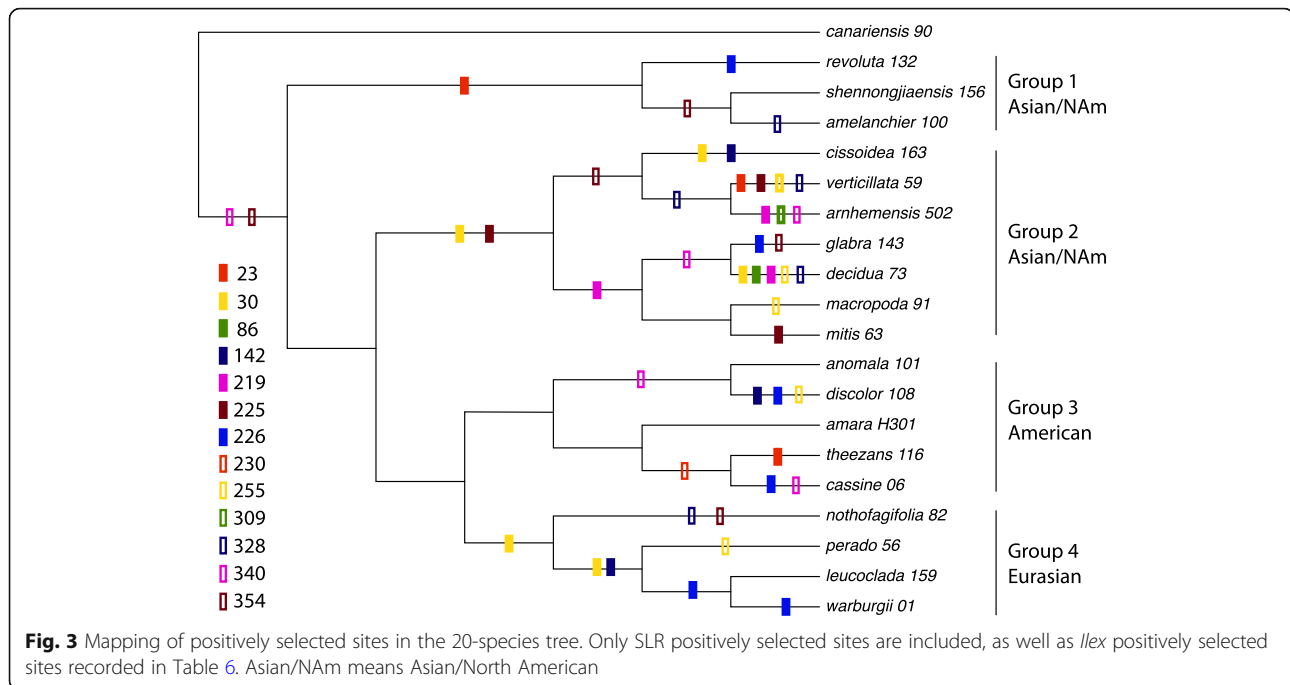
selection on the Rubisco large subunit encoded by a non-recombining plastid DNA may promote, among other mechanisms (demographic events, fixation by chance), advantageous plastid captures [28]. This could be the case in *Ilex*. Huge cytonuclear discordances in the phylogeny of this genus have been reported: the plastid phylogeny is more geographically structured (with 4 geographically-determined phylogenetic clades) than the nuclear phylogeny, which was closer to the morphological classification of the genus [36]. Since the genus *Ilex* probably experienced extensive adaptation to changing habitats, plastid capture by adaptive introgression could well be promoted by a Darwinian selection on Rubisco.

On the other hand, such adaptive introgressions require that the surface of plastid-encoded Rubisco large subunits of one species fit with the surface of nuclear-encoded small subunits and chaperones of another species. This “molecular adaptation” could be essential for introgression between distantly related species, requiring a subsequent

modification of molecular interactions. In *Ilex*, species from different lineages can hybridize when they come into secondary contact, helped by rather weak reproductive barriers and by many repeated intercontinental dispersals [36].

In this context, selection pressure on the Rubisco large subunit of one species might be, in part, the consequences such molecular incompatibility with the small subunit of another species. The Rubisco activase chaperone was suggested to interact with the region 89–94 of *rbcL* [43]. In *Ilex*, the exposed positively selected residues 91 and 95 may interfere with Rubisco activase (see Fig. 2b). Chaperones *rbcX* interacted with the C-terminal tail of one L subunit and the N-terminal region of another L subunit, to form 2 clamps that stabilized the dimer L₂ [23]. Residues 461 and 470 (C-terminal tail) were positively selected. The same was true for positively selected residues 15 and 19 in the N-terminal region (see Fig. 2b).

It has to be noted that the majority of exposed positively selected residues of the *Ilex* Rubisco, found on the



equatorial surface of the L_8S_8 hexadecamer shown in Fig. 2b, were not in the list of the 20 most often positively selected residues detected by Kapralov and Filatov [29]. These were residues 15, 19, 21, 23, 30, 84, 91, 340, 353 and 354 (see Fig. 2b and Tables 1, 2). Such uncommon positively selected sites, potentially interacting with chaperones, may represent a particularity of the genus *Ilex*. Thus, it can be postulated that the *rbcL* gene of taxa experiencing speciation by adaptive introgressions was expected to be under compensatory selective pressure to make the interface with nuclear-encoded partners compatible. In *Flaveria*, specific residue substitutions in the small subunit gene (*rbcS*) were correlated with the kinetic properties of Rubisco [30], but this observation was not further examined. In addition, Rubisco chaperones might have some effects on evolvability of *rbcL* [8].

Cytosynuclear phylogenetic discordance is pervasive in *Ilex*, except in the well-defined Eurasian plastid group 4 which perfectly matches the nuclear clade Aquifolium [36]. Some hybridization events were noticeable between closely related members of this group. However, it appears that, for an unknown reason, introgression never occurred between species of the Eurasian group 4 and distantly related species belonging to the other groups 2, 3 and 4. This may explain the fewer positive selections *rbcL* in this group compared with other groups (Tables 1 and 2). Thus, in lineages experiencing introgression, a high rate of positive selection on the Rubisco large subunit may be expected. A large part of it would represent molecular adaptations resulting from introgression and plastid capture, a current situation in *Ilex*. Therefore, in future studies,

molecular adaptations need to be discriminated from environmental adaptations of Rubisco (see [41]).

Conclusions

In *Ilex*, the plastid *rbcL* gene coding for the large subunit of Rubisco had a higher rate of positive selection than in other lineages studied so far. Most of these positively selected residues are located at the surface of the large subunit and are likely to be involved in interactions with other plastid-encoded L or nuclear-encoded S subunits, as well as with the numerous nuclear-encoded Rubisco chaperones. Coevolution between positively selected sites to restore global stability is common. The high rate of positive selection in *Ilex* is probably linked with the pervasive hybridization and introgression in the genus, although the precise mechanisms behind this link are currently unclear.

Materials and methods

We first investigated *rbcL* sequences of 116 specimens [36] representing most of the global distribution range of the genus (including 4 continents and 8 Atlantic and Pacific archipelagos); the “Worldwide dataset”. Then we examined *rbcL* sequences of 124 accessions of Chinese *Ilex* species recently collected in their natural habitats, mainly in Yunnan, SW China; the “Chinese dataset”. SW China is the center of *Ilex* diversity in East Asia and species occupy an exceptionally wide range of habitats, from lowland tropical rainforest to subtropical tree lines. Altogether, this study was based on 240 specimens.

The worldwide dataset

For the detection of positively selected sites and site coevolution in *rbcl*, the data of Manen et al. [36], comprising 116 specimens and 108 species, was used. These 116 *rbcl* sequences were truncated after codon position 436. DNA-bank accession numbers are listed in Additional file 1: Table S1. To speed up PAML and SLR calculations (see below), the *rbcl* data was divided into five DNA alignments. Four of them represent previously recognized plastid groups ([36]; see Additional file 1: Table S1): Asian/North-American Group 1 (14 sequences), Asian/North-American Group 2 (36 sequences), American Group 3 (31 sequences), and Eurasian Group 4 (35 sequences). The fifth consisted of 20 representative *Ilex* species from the 4 phylogenetic plastid groups (species names in bold in Additional file 1: Table S1). All *Ilex* trees used in this study had the topology of the plastid tree of Manen et al. [36]. The *rbcl* sequence of the rather phylogenetically isolated *I. canariensis* [36] was chosen as a reference in alignments and trees because it didn't belong to a particular plastid group. All alignments were done well using MUSCLE [9] because of the high conservatism in *rbcl* sequences. The five *Ilex* alignments (14 to 36 *rbcl* sequences) were in the size range of alignments used by Kapralov and Filatov [29] in their broad study based on 151 lineages of photosynthetic organisms, so their results could be compared with the results obtained in this study. For CAPS calculation (detection of coevolving residues, see below), the *rbcl* alignment of the 116 *Ilex* specimens was used.

The Chinese dataset

New *rbcl* sequences from China consist of 124 accessions, including 77 identified and 35 as yet unidentified species. Most *rbcl* sequences were amplified and sequenced from DNA-extracts of fresh or dried material, using amplifying and sequencing primers of Fay et al. [13]. Some others were extracted from total plastid genomes of Yao et al. [55]. Additional file 2: Table S2 gives information on specimens and the DNA-bank accession numbers of their *rbcl* sequences. These sequences were not truncated after codon position 436. Several accessions that have not yet been morphologically determined were kept in the analysis because *rbcl* sequence comparisons with BLAST (Basic Local Alignment Search Tool, www.ncbi.nlm.nih.gov/BLAST) confirmed that they indeed represent *Ilex* and phylogenetic comparison with the sequences of Manen et al. [36] allowed the determination of the phylogenetic groups to which they belong: North-America/East-Asian Group 1 (39 sequences), Asian/North-American Group 2 (40 sequences), and Eurasian Group 4 (45 sequences). PAML and SLR calculations (detection of positive selection, see below) were done on these three plastid lineages separately and results compared with calculations on the Worldwide dataset.

Detection of positive selection on *rbcl*

The nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) is an indicator of selective pressure at the protein level, with $\omega = 1$ meaning neutral mutations, $\omega < 1$ purifying selection, and $\omega > 1$ diversifying positive selection. Amino acid sites in a protein are expected to be under different selective pressures and to have different underlying ω ratios. In the program Codeml of the PAML package (Phylogenetic Analysis by Maximum Likelihood [54]), different models that account for heterogeneous ω ratios among amino acid sites were tested by maximum likelihood on aligned protein-coding DNA sequences and the corresponding phylogenetic tree. The program is useful for testing adaptive molecular evolution and identifying amino acid sites under positive selection. The likelihoods obtained by the models M7 and M8a (allowing $\omega \leq 1$, the null hypothesis) and by the model M8 (allowing $\omega > 1$) were compared by the likelihood ratio test (LRT). For the model M8, a Bayesian empirical procedure was used to estimate the mean value for each codon and the posterior probability that it is under positive selection. LRT of models Ma1 (the null hypothesis) and Ma2 (allowing $\omega > 1$) were also calculated.

The use of PAML allows a comparison with earlier studies on *rbcl* positive selection, particularly Kapralov and Filatov [29]. The Sitewise Likelihood-Ratio test (SLR test, <http://www.ebi.ac.uk/goldman-srv/SLR/#download>), a measure of the strength of the evidence for selection ([40]), was used for adjustment. It distinguishes potential false-positive results that have been reported with PAML. Data on eight lineages from Kapralov and Filatov [29] were re-calculated using both PAML and the SLR test for comparison.

Structural analysis of the large subunit of Rubisco

The protein structure of the large subunit of Rubisco for *Ilex canariensis* was modeled by homology using the Swiss Model server (<http://swissmodel.expasy.org> [5]). This program looks for the closest sequence from which a 3-D structure is available in the Protein Data Base (PDB) and uses it as model. The “magic fit” option of Swiss-Pdb Viewer shows that the three-dimensional structure of the large subunit of *Ilex canariensis* fits well with the three-dimensional structure of the large subunit of *Spinacia oleracea* (Additional file 4: Figure S1). Swiss-Pdb Viewer (<http://www.expasy.org/spdbv>; [20]) was used to localize positively selected residues on this modeled *Ilex canariensis* structure. Using the “Select” commands of Swiss-Pdb Viewer, the vicinity of positively selected residues at a 6 Å distance (the van der Waals contact) can be investigated at the interface between other Rubisco large subunits and between the large subunit and small subunits. As *rbcl* is not yet available in *Ilex*, an extrapolation

was based on the *Spinacia oleracea* Rubisco three-dimensional structure (PDB accessions 1RCX, [51]).

Detecting coevolving sites

CAPS (Coevolution Analysis using Protein Sequences, <http://caps.tcd.ie> [11, 12];) was used for identifying coevolution between amino acid sites. Blossum-corrected amino acid distances were used to identify amino acid covariation. Phylogenetic relationships were used to remove the phylogenetic and stochastic dependencies between sites. The 3D protein structure was used to identify the nature of the dependencies between coevolving amino acid sites [11]. Times of divergence were estimated as the mean number of substitutions per synonymous site. DNA alignment of *rbcl* sequences of the worldwide dataset of 116 *Ilex* specimens, the corresponding plastid ML-optimized tree of Manen et al. [36], and the previously determined 3-D structure of the large subunit of *Ilex canariensis* Rubisco, were used as inputs. Parameters were: alpha threshold 0.001, number of simulated alignments 1000, bootstrap value 0.95, await convergence and time correction on.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12862-019-1521-1>.

Additional file 1: Table S1. List of world-wide *Ilex* species [32]. In bold, species used in the 20-species analyses.

Additional file 2: Table S2. List of Chinese *Ilex* species studied here.

Additional file 3: Table S3. List of taxa used in comparison with *Ilex*.

Additional file 4: Figure S1. Comparison of the 3-dimensional structure of the Rubisco large subunit of *Spinacia oleracea* (in pink) with modeled 3-dimensional structure of the Rubisco large subunit of *Ilex canariensis* (in orange) using the "magic fit" option of Swiss-Pdb Viewer.

Abbreviations

CAPS: Coevolution Analysis using Protein Sequences; GBIF: Global Biodiversity Information Facility; MYA: Million years ago; PAML: Phylogenetic analysis by maximum likelihood; PCA: Principal component analysis; PDB: Protein Data Base; Rubisco: Ribulose-1,5-bisphosphate carboxylase-oxygenase; SLR: Sitewise Likelihood-Ratio tests

Acknowledgements

We thank Juan-hong Zhang, Chun-yan Lin and Ji-xiong Yang from the Germplasm Bank of Wild Species of Kunming Institute of Botany, Chinese Academy of Sciences, for their help with experiments.

Authors' contributions

XY, RTC and JFM designed the study. XY, YHT and JFM collected the samples. XY, JBY and JFM performed the experiments. XY, YW and JFM analyzed the data. XY, RTC and JFM wrote and edited the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the CAS "Light of West China" Program [Y9XB071B01 to X.Y.], the CAS 135 program [2017XTBG-T03 to X.Y.] and the 1000 Talents Program [WQ20110491035 to R.T.C.]. Y.H.T. was supported by grants from the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences for collecting some samples.

Availability of data and materials

All data included in the manuscript has been deposited in GenBank and their accession numbers can be found in supplementary files.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China. ²Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla 666303, China. ³Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw, Myanmar. ⁴Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China. ⁵CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, Yunnan, China. ⁶Laboratoire de Systématique Végétale et Biodiversité, University of Geneva (retired), Chemin de l'Impératrice 1, CH-1292 Chambésy, Switzerland.

Received: 18 March 2019 Accepted: 27 September 2019

Published online: 21 October 2019

References

- Antonelli A. Higher level phylogeny and evolutionary trends in Campanulaceae subfam. Lobelioideae: Molecular signal overshadows morphology. *Molecular phylogenetics and evolution*. 2008;46(1):1–18.
- APG IV. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants. *Bot J Linn Soc*. 2016;181:1–20.
- Baas P. Inheritance of foliar and nodal anatomical characters in some *Ilex* hybrids. *Bot J Linn Soc*. 1978;77:41–52.
- Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. Rubisco is not really so bad. *Plant Cell Environ*. 2018;41:705–16.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42:W252–8.
- Bracher A, Hauser T, Liu C, Hartl FU, Hayer-Hartl M. Structural analysis of the Rubisco-assembly chaperone *RbcX-II* from *Chlamydomonas reinhardtii*. *PLoS One*. 2015;10:e0135448.
- Christin PA, Salamin N, Muasya AM, Roalson EH, Russier F, Besnard G. Evolutionary switch and genetic convergence on *rbcl* following the evolution of C4 photosynthesis. *Mol Biol Evol*. 2008;25:2361–8.
- Durão P, Aigner H, Nagy P, Mueller-Cajjar O, Hartl FU, Hayer-Hartl M. Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat Chem Biol*. 2015;11:148–55.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
- Erb JT, Zarzycki J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO₂ fixing enzyme. *Curr Opin Biotechnol*. 2018;49:100–7.
- Fares MA, McNally D. CAPS: coevolution analysis using protein sequences. *Bioinformatics*. 2006;22:2821–2.
- Fares MA, Travers SA. A novel method for detecting intramolecular coevolution: adding a further dimension to selective constraints analyses. *Genetics*. 2006;173:9–23.
- Fay MF, Bayer C, Alverson WS, de Bruijn AY, Chase MW. Plastid *rbcl* sequence data indicate a close affinity between *Diegodendron* and *Bixa*. *Taxon*. 1998;47:43–50.
- Galmés J, Flexas J, Keys AJ, Cifre J, Mitchell RA, Madgwick PJ, Haslam RP, Medrano H, Parry MA. Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ*. 2005;28:571–9.
- Galmés J, Andralojc PJ, Kapralov MV, Flexas J, Keys AJ, Molins A, Parry MAJ, Conesa MA. Environmentally driven evolution of Rubisco and improved

- photosynthesis and growth within the C3 genus *Limonium* (Plumbaginaceae). *New Phytol.* 2014a;203:989–99.
16. Galmés J, Kapralov MV, Andralojc PJ, Conesa MA, Keys AJ, Martin AJ, Parry MAJ, Flexas J. Expanding knowledge of the Rubisco kinetics variability in plant species: environmental and evolutionary trends. *Plant Cell Environ.* 2014b;37:1989–2001.
 17. Gao YD, Harris AJ, Zhou SD, He XJ. Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the Q–T plateau and the Hengduan Mountains. *Mol Phylogenet Evol.* 2013;68:443–60.
 18. Gruber AV, Feiz L. Rubisco assembly in the chloroplast. 2018; *Frontiers in Molecular Biosciences.* 5:24.
 19. Gregory-Wodzicki KM. Uplift history of the central and northern Andes: a review. *Geol Soc Am Bull.* 2000;112:1091–105.
 20. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-Pdb viewer: an environment for comparative protein modeling. *Electrophoresis.* 1997;18:2714–23.
 21. Hao DC, Mu J, Xiao PG. Molecular evolution and positive Darwinian selection of the gymnosperm photosynthetic Rubisco enzyme. *Bot Stud.* 2010;51:491–510.
 22. Hauser T, Bhat JY, Miličić G, Wendler P, Hartl FU, Bracher A, Hayer-Hartl M. Structure and mechanism of the Rubisco-assembly chaperone *Raf1*. *Nat Struct Mol Biol.* 2015a;22:720–8.
 23. Hauser T, Popilka L, Hartl FU, Hayer-Hartl M. Role of auxiliary proteins in Rubisco biogenesis and function. *Nat Plants.* 2015b;1:15065.
 24. Hermida-Carrera C, Fares MA, Fernández Á, Gil-Pelegrín E, Kapralov MV, Mir A, Molins A, Peguero-Pina JJ, Rocha J, Sancho-Knapik D, Galmés J. Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. *PLoS One.* 2017;12:e0183970.
 25. Hu S, Sablok G, Wang B, Qu D, Barbaro E, Viola R, Li M, Varotto C. Plastome organization and evolution of chloroplast genes in *Cardamine* species adapted to contrasting habitats. *BMC Genomics.* 2015;16:306.
 26. Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K. Molecular adaptation of *rbcl* in the heterophyllous aquatic plant *Potamogeton*. *PLoS One.* 2009;4:e4633.
 27. Kapralov MV, Smith JAC, Filatov DA. Rubisco evolution in C4 Eudicots: an analysis of *Amaranthaceae sensu lato*. *PLoS One.* 2012;1:e52974.
 28. Kapralov MV, Filatov DA. Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS One.* 2006;1:e8.
 29. Kapralov MV, Filatov DA. Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol.* 2007;7:73.
 30. Kapralov MV, Kubien DS, Andersson I, Filatov DA. Changes in Rubisco kinetics during the evolution of C4 photosynthesis in *Flaveria* (Asteraceae) are associated with positive selection on genes encoding the enzyme. *Mol Biol Evol.* 2011;28:1491–503.
 31. Knobloch E, Mai DH. 1986. Monographie der Früchte und Samen in der Kreide von Mitteleuropa. *Rozpr. Ustredni Ustav Geologicky.* 1986;47:5–219.
 32. Liu J-Q, Wang Y-J, Wang A-L, Hideaki O, Abbott RJ. Radiation and diversification within the *Ligularia-Cremnathodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan plateau. *Mol Phylogenet Evol.* 2006;38:31–49.
 33. Li X, Sun B, Xiao L, Wu J, Lin Z. Leaf macrofossils of *Ilex* protocornuta sp. nov. (Aquifoliaceae) from the Late Miocene of East China: Implications for palaeoecology. *Review of Palaeobotany and Palynology.* 2010;161(1-2):87–103.
 34. Loizeau P-A, Barriera G, Manen J-F, Broennimann O. Towards an understanding of the distribution of *Ilex* L. (Aquifoliaceae) on a world-wide scale. *Biol Skr.* 2005;55:501–20.
 35. Magalon and Sanderson, 2005" should be "Magallóan SA, Sanderson MJ. Angiosperm divergence times: the effect of genes, codon positions, and time constraints. *Evolution.* 2005;59(8):1653–1670.
 36. Manen J-F, Barriera G, Loizeau PA, Naciri Y. The history of extant *Ilex* species (Aquifoliaceae): evidence of hybridization within a Miocene radiation. *Mol Phylogenet Evol.* 2010;57:961–77.
 37. Manen J-F. Are both sympatric species *Ilex perado* and *Ilex canariensis* secretly hybridizing? Indication from nuclear markers collected in Tenerife. *BMC Evol Biol.* 2004;4:1.
 38. Manen J-F, Boulter MC, Naciri-Graven Y. The complex history of the genus *Ilex* L. (Aquifoliaceae): evidence from the comparison of plastid and nuclear DNA sequences and from fossil data. *Plant Syst Evol.* 2002;235:79–98.
 39. Manen J-F, Cuénoud P, Martinez MD. Intralineage variation in the pattern of *rbcl* nucleotide substitution. *Plant Syst Evol.* 1998;211:103–12.
 40. Massingham T, Goldman N. Detecting amino acid sites under positive selection and purifying selection. *Genetics.* 2005;169(3):1753–1762.
 41. Parto S, Lartillot N. Molecular adaptation in Rubisco: discriminating between convergent evolution and positive selection using mechanistic and classical codon models. *PLoS One.* 2018; <https://doi.org/10.1371/journal.pone.0192697>.
 42. Pennington RT, Lavin M, Särkinen T, Lewis GP, Klitgaard BB, Hughes C. Contrasting plant diversification histories within the Andean biodiversity hotspot. *PNAS.* 2010;107:13783–7.
 43. Portis AR, Li C, Wang D, Salvucci ME. Regulation of Rubisco activase and its interaction with Rubisco. *J Exp Bot.* 2008;59:1597–604.
 44. Pottier M, Gilis D, Boutry M. The hidden face of Rubisco. *Trends Ecol Evol.* 2018; <https://doi.org/10.1016/j.tplants.2018.02.006>.
 45. Sanderson MJ, Doyle JA. Sources of error and confidence intervals in estimating the age of angiosperms from *rbcl* and 18S rDNA data. *Am J Bot.* 2001;88:1499–516.
 46. Sen L, Fares MA, Liang B, Gao L, Wang B, Wang T, Su YJ. Molecular evolution of *rbcl* in three gymnosperm families: identifying adaptive and coevolutionary patterns. *Biol Direct.* 2011;6:29.
 47. Setoguchi H, Watanabe I. Intersectional gene flow between insular endemics of *Ilex* (Aquifoliaceae) on the Bonin Islands and the Ryukyu Islands. *Am J Bot.* 2000;87:793–810.
 48. Sharwood RE, von Caemmerer S, Maliga P, Whitney SM. The catalytic properties of hybrid Rubisco comprising tobacco small and sunflower large subunits mirror the kinetically equivalent source Rubiscos and can support tobacco growth. *Plant Physiol.* 2008;146:83–96.
 49. Shi L, Li N, Wang S, Zhou Y, Huang W, Yang Y, Ma Y, Zhou R. Molecular evidence for the hybrid origin of *Ilex dabieshanensis* (Aquifoliaceae). *PLoS One.* 2016;11:e0147825.
 50. Studer RA, Christin PA, Williams MA, Orengo CA. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *PNAS.* 2014;111:2223–8.
 51. Taylor TC, Andersson I. The structure of the complex between rubisco and its natural substrate ribulose 1,5-bisphosphate. *J Mol Biol.* 1997;265:432–44.
 52. Wachter RM, Salvucci ME, Carmo-Silva AE, Barta C, Genkov T, Spreitzer R. Activation of interspecies-hybrid Rubisco enzymes to assess different models for the Rubisco–Rubisco activase interaction. *Photosynth Res.* 2013;117:557–66.
 53. Wang M, Kapralov MV, Anisimova M. Coevolution of amino acid residues in the key photosynthetic enzyme Rubisco. *BMC Evol Biol.* 2011;11:266.
 54. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91.
 55. Yao X, Tan Y-H, Liu Y-Y, Song Y, Yang J-B, Corlett RT. Chloroplast genome structure in *Ilex* (Aquifoliaceae). *Sci Rep.* 2016;6:28559.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

