

Evolutionary Switch and Genetic Convergence on *rbcL* following the Evolution of C₄ Photosynthesis

Pascal-Antoine Christin,* Nicolas Salamin,* A. Muthama Muasya,† Eric H. Roalson,‡
Flavien Russier,* and Guillaume Besnard*

*Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland; †Botany Department, University of Cape Town, Rondebosch, South Africa; and ‡School of Biological Sciences and Center for Integrated Biotechnology, Washington State University

Rubisco is responsible for the fixation of CO₂ into organic compounds through photosynthesis and thus has a great agronomic importance. It is well established that this enzyme suffers from a slow catalysis, and its low specificity results into photorespiration, which is considered as an energy waste for the plant. However, natural variations exist, and some Rubisco lineages, such as in C₄ plants, exhibit higher catalytic efficiencies coupled to lower specificities. These C₄ kinetics could have evolved as an adaptation to the higher CO₂ concentration present in C₄ photosynthetic cells. In this study, using phylogenetic analyses on a large data set of C₃ and C₄ monocots, we showed that the *rbcL* gene, which encodes the large subunit of Rubisco, evolved under positive selection in independent C₄ lineages. This confirms that selective pressures on Rubisco have been switched in C₄ plants by the high CO₂ environment prevailing in their photosynthetic cells. Eight *rbcL* codons evolving under positive selection in C₄ clades were involved in parallel changes among the 23 independent monocot C₄ lineages included in this study. These amino acids are potentially responsible for the C₄ kinetics, and their identification opens new roads for human-directed Rubisco engineering. The introgression of C₄-like high-efficiency Rubisco would strongly enhance C₃ crop yields in the future CO₂-enriched atmosphere.

Introduction

The key role played by Rubisco (ribulose-1,5-bisphosphate carboxylase) in most trophic chains, including human consumption, has made this enzyme the subject of extensive research (Spreitzer and Salvucci 2002; Portis and Parry 2007). It is well established that this enzyme is inefficient under many conditions, to the extent that it has been called “sluggish” because its turnover rate, the maximum number of CO₂ molecules fixed per second, is very low (Tcherkez et al. 2006). In order to counteract this low catalytic efficiency, plants have to overexpress this enzyme in photosynthetic cells. As a consequence, Rubisco represents, in some plants, about one-third of all soluble proteins and 20% of the total nitrogen is invested in Rubisco production (Evans and Poorter 2001). The low catalytic efficiency of Rubisco probably corresponds to the evolutionary resolution of a trade-off between Rubisco efficiency and CO₂ specificity (Tcherkez et al. 2006). These two enzyme properties, in the case of Rubisco, are negatively correlated. Thus, an increased turnover rate implies a decrease of the CO₂:O₂ relative affinity and consequently a higher O₂ fixation rate (Tcherkez et al. 2006). This, in turn, results in photorespiration that can represent an unproductive energy sink for the plant (Ludwig and Calvin 1971; Monteith 1977; Douce and Heldt 2000). During the early stages of Rubisco evolutionary history, which started some 2.9–2.7 billion years ago (Bya; Nisbet et al. 2007), O₂ atmospheric concentrations were very low, and photorespiration was probably negligible. After the atmospheric O₂ increase (which started some 2.32 Bya; Bekker et al. 2004), the photorespiration cost forced Rubisco to shift its trade-off toward higher specificity at the expense of catalytic efficiency (Tcherkez et al. 2006). The trade-off optimum varies among photosynthetic

organisms depending on the importance of photorespiration (Jordan and Ogren 1981; Brainbridge et al. 1995; Tcherkez et al. 2006). For instance, with lowering temperature, Rubisco affinity for CO₂ increases faster than for O₂ (von Caemmerer and Quick 2000), and the cost of photorespiration is reduced. As a consequence, plants from cooler habitats have a Rubisco with lower specificity and higher turnover rate than those of warm and dry habitats (Sage 2002; Galmés et al. 2005). To decrease the photorespiration cost, some photosynthetic organisms developed CO₂ concentration mechanisms (CCM). By concentrating CO₂ around Rubisco, these mechanisms almost suppress photorespiration and enable higher Rubisco turnover rate (Tcherkez et al. 2006). The C₄ photosynthesis CCM is of great interest because it occurs in plant families that contain important crops. In particular, the grass family (Poaceae) includes major C₃ and C₄ cereals, such as rice, wheat, barley, maize, or sorghum.

In the C₄ pathway, the initial atmospheric CO₂ fixation is performed by phosphoenolpyruvate carboxylase, an enzyme with a high affinity for CO₂. Four-carbon compounds, which result from this fixation in the mesophyll cells are then transported to the specialized bundle-sheath cells of C₄ plants where CO₂ is released in chloroplasts. Therein, CO₂ is fixed by Rubisco as in C₃ plants. The main difference is that CO₂ concentrations in C₄ bundle-sheath cells are up to 10-fold those found in C₃ plants (von Caemmerer and Furbank 2003). This high CO₂ condition in leaves following C₄ evolution potentially decreased the adaptive value of a Rubisco with high CO₂ affinity as attested by the high *K_m* of C₄ Rubisco for the CO₂ substrate (Yeoh et al. 1980, 1981; Hudson et al. 1990; Kubien et al. 2008; fig. 1). The purifying selection on Rubisco sites that were important for maintenance of high CO₂ affinity could thus have been relaxed in C₄ plants, and the possibility of directional selection acting on an increased catalytic efficiency could have led to the increased turnover rate of C₄ Rubiscos (Seemann et al. 1984; Hudson et al. 1990; Sage 2002; Kubien et al. 2008; fig. 1). This higher catalytic efficiency results in a 60–80% reduction of Rubisco

Key words: Rubisco, molecular evolution, positive selection, C₄ photosynthesis, convergence, monocots.

E-mail: pascal-antoine.christin@unil.ch; guillaume.besnard@unil.ch.

Mol. Biol. Evol. 25(11):2361–2368. 2008

doi:10.1093/molbev/msn178

Advance Access publication August 11, 2008

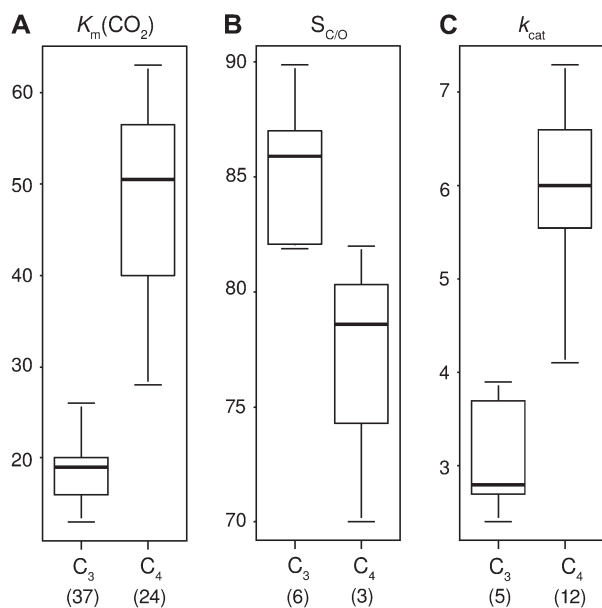


FIG. 1.—Comparison of Rubisco kinetic parameters between C₃ and C₄ plants. (A) Rubisco affinity for the CO₂ substrate [$K_m(\text{CO}_2)$]. Values come from Yeoh et al. (1980) and are all from grasses. (B) Specificity for CO₂ relative to O₂ (S_{c/O_2}). Data comes mainly from Kane et al. (1994). Values for *Atriplex glabriuscula* (C₃), *Sorghum bicolor* (C₄), and *Amaranthus hybridus* (C₄) were added according to Tcherkez et al. (2006). (C) Catalytic efficiency (k_{cat}). Values from Seemann et al. (1984). The number of species for each category is indicated in brackets. To avoid differences due to experimental protocol variations, kinetic values were usually not mixed between different studies.

abundance in C₄ plants (Long 1999) and thus a lower nitrogen requirement (Ghannoum et al. 2005). Engineering of C₃ plants with C₄ more efficient Rubisco would potentially increase yield of C₃ crops, such as rice, wheat, or barley (Brainbridge et al. 1995; Parry et al. 2003, 2007; Zhu et al. 2004; Long et al. 2006). However, the determinants of Rubisco kinetic properties are not precisely known yet. The occurrence of Rubisco with high turnover rate in plant lineages that evolved the C₄ trait independently (Sage 2004; Christin et al. 2008) is of prime importance in this context as it will enable direct comparisons of C₃ and C₄ genes from species sharing long evolutionary history.

Rubisco of land plants and green algae is composed of eight large subunits, encoded by the plastid gene *rbcL* and eight small subunits encoded by a family of nuclear genes (*rbcS* [Spreitzer and Salvucci 2002]). In crosses between C₃ and C₄ species from both *Flaveria* and *Atriplex*, the C₄ catalytic specificities of Rubisco were maternally inherited, demonstrating that genetic modifications conferring the characteristics of a C₄ Rubisco mostly lie in the plastid *rbcL* gene (Hudson et al. 1990).

This study aims at testing whether the evolution of C₄ photosynthesis switched the selective pressures on *rbcL*. C₄-linked positive selection on this gene is tested using phylogenetic analyses of commelinoid monocots including all known major C₄ lineages in the Poaceae and Cyperaceae families (Sage 2004; Bruhl and Wilson 2007; Christin et al. 2008). Sites that underwent adaptive changes in C₄ plants are identified and compared among C₄ lineages. This evolutionary analysis of Rubisco-encoding gene can shed new

light on the putative genetic determinants of its kinetic variations.

Materials and Methods

Phylogenetic Analyses

Sequences used in this study come mainly from a data set covering the main commelinoid flowering plant lineages (monocots; Christin et al. 2008), with an especially dense sampling on the Poaceae (grass family) that contains at least 17 independent C₄ lineages (Christin et al. 2008) and about 60% of C₄ plants (Sage 2004). This data set is composed of *rbcL* and *ndhF* sequences, which are both plastid markers. It was completed by sequencing *rbcL* and *ndhF* genes from 100 members of the Cyperaceae (sedge family), which contains about 20% of C₄ plant species (Sage 2004; supplementary table 2, Supplementary Material online). The sampled species were chosen to represent the different Cyperaceae C₄ lineages as well as their C₃ relatives (based on Bruhl and Wilson [2007]). The polymerase chain reaction amplification protocol from Christin et al. (2008) was used for both *rbcL* and *ndhF*, except that a new set of primers was developed to amplify *ndhF* genes of Cyperaceae in two overlapping segments (supplementary table 3, Supplementary Material online).

The final data set contained 338 species (supplementary table 2, Supplementary Material online). For 15 species, no data were available for *rbcL*, and for 3 species, no data were available for *ndhF*. A phylogenetic tree was obtained by Bayesian inference using MrBayes 3.1 (Ronquist and Huelsenbeck 2003). The general time-reversible nucleotide substitution model, with gamma shape parameter and proportion of invariant sites, was used after determining the best-fit model through hierarchical likelihood ratio tests (LRT). All model parameters were optimized independently for each data set. Two independent analyses, each with four parallel chains, were run for 10,000,000 generations sampling trees every 1,000 generations after a burn-in period of 3,000,000 generations.

Positive Selection Tests

Codon models were used to test for the presence of positive selection during the evolutionary history of *rbcL*. To avoid missing data, only species for which the *rbcL* sequence was complete between codons 10 and 456 (positions numbered following the complete sequence of *Zea mays*, NC_001666; the total sequences is 476 codons long) were considered for these analyses. This resulted in 240 *rbcL* sequences (supplementary table 2, Supplementary Material online). Grass C₄ lineages number 14 (represented by *Leptocoryphium*) and 5 (*Eriachne* genus) were absent from this data set. The species whose *rbcL* gene was not considered were manually pruned from the Bayesian phylogenetic tree inferred from all 338 species, which was then used as the reference topology for the codon model analyses.

Four different codon models were optimized using the codeml software in PAML version 4 (Yang 2007), with F3×4 empirical codon frequency estimates. This software

implements the codon substitution model of Yang et al. (1998), a simplified version of the model of Goldman and Yang (1994). The continuous-time Markov process uses a 61×61 rate matrix (transitions from each codon to each other, excluding the three stop codons). Codon substitution rates are multiplied by a factor ω if the encoded amino acid is changed (Yang and Bielawski 2000). This ω is the ratio of nonsynonymous mutations rate per synonymous mutations rate (d_N/d_S ratio). It represents the selective pressures, either neutral ($\omega = 1$), purifying ($\omega < 1$), or positive ($\omega > 1$) selection. The first model, site model M1a, assumes that the selection pressures are the same on all branches of the phylogenetic tree (Yang et al. 2000). In this model, codons can evolve either under neutral selection or under purifying selection. The second model, model A, is a branch site model, which means that the selective pressure can vary not only among sites (similarly to model M1a) but also among branches of the phylogenetic tree (Yang and Nielsen 2002). The logic behind this model is that positive selection concerns only some sites at some points of the evolutionary history, such as ecological or physiological transitions. It requires an a priori identification of the branches on which selective pressures are expected to differ (foreground branches). In our case, we tested whether positive selection could be detected along all branches inside the C₄ clades as defined in Christin et al. (2008) for the grasses and all branches inside monophyletic C₄ clades in Cyperaceae. Two monocots using the Crassulacean acid metabolism (CAM) are also present in the data set (*Ananas comosus* and *Tillandsia bergeri*). The CAM pathway also increases the CO₂ concentration in leaves (Lüttge 2004), and for this reason, branches leading to these two species were also considered as foreground branches for the selection tests. Codons were then assigned to four selective pressure categories: 1) purifying selection, 2) neutral selection on the whole tree, 3) purifying selection on every branches except on foreground branches that experience positive selection, and 4) neutral selection on every branches except on foreground branches that experience positive selection. Posterior probabilities of belonging to the different classes were estimated by the Bayes empirical Bayes procedure (Yang et al. 2005). The third model, model A', is a particular case of model A where positive selection in foreground branches is replaced by neutral selection (Zhang et al. 2005). Model M1a and A' are nested in model A, which is nested in model M1a. These three models can thus be compared through LRT. The fourth model M2a is a branch model with a proportion of codons evolving under positive selection in all branches of the phylogenetic tree. This model is similar to model M1a but implements a third category of sites with $\omega > 1$. Models M1a and M2 are nested and can be compared through LRT. However, LRT cannot be used to compare models M2a and A. These two models were thus compared using the Akaike information criterion (AIC).

The codons identified as evolving under positive selection in C₄ branches (in model A) with a posterior probability, estimated by the Bayes empirical Bayes procedure (Yang et al. 2005), greater than 0.95 were then compared among all the 323 *rbcl* sequences for which the encoded amino acid was known, even if the sequence was not included in selection tests.

Table 1
Proportion of Sites Attributed to Each Class and ω (d_N/d_S) Values Optimized in the Branch-Site Model A

Codon Class	Proportion of Sites	ω in C ₃ Clades	ω in C ₄ Clades
0	0.88	0.014	0.014
1	0.10	1.0 ^a	1.0 ^a
2a	0.02	0.014	5.652
2b	< 0.01	1.0 ^a	5.652

^a ω values defined in the model and not optimized.

Results

Phylogenetic Tree

The phylogenetic tree inferred from *rbcl* and *ndhF* plastid markers (supplementary fig. 1, Supplementary Material online) was largely congruent with a phylogeny previously obtained with a smaller species sampling (Christin et al. 2008). Recognized families were all monophyletic. In grasses, species clustered by subfamilies (as defined by Duvall et al. [2007]) and formed 18 monophyletic C₄ clades, in subfamilies Aristidoideae (lineages 1 and 2), Chloridoideae (lineages 3 and 4), Micrairoideae (lineage 5), and Panicoideae (lineages 6–17). These were shown to correspond to 17 or 18 independent acquisitions of the C₄ trait (Christin et al. 2008). C₄ sedges analyzed formed six distinct phylogenetic lineages; that is, one subgroup of *Rhynchospora* (belonging to subgenus *Haplostyleae*; lineage 18), two distinct clades in tribe Abildgaardieae (named *Fimbristylis* [lineage 19] and *Bulbostylis* [lineage 20] clades), two *Eleocharis* species (*Eleocharis badwinii* [lineage 21] and *Eleocharis vivipara* [lineage 22]), and the chlorocyperoid species of tribe Cypereae (lineage 23). At least twenty-three independent C₄ lineages (in Cyperaceae and Poaceae) were thus represented in our data set.

Positive Selection Tests

The model allowing positive selection in branches inside the C₄ clades, but not in C₃ branches, was significantly better than the model with constant selective pressures across the phylogenetic tree (models M1a vs. A; chi-squared = 235, degrees of freedom [df] = 2, *P* value < 0.0001) and the model assuming relaxed selection in C₄ branches (models A' vs. A; chi-squared = 150, df = 1, *P* value < 0.0001). In C₄ clades, the estimated ratio of nonsynonymous versus synonymous mutations on some *rbcl* codons was 5.65 (table 1), demonstrating that this gene evolved under strong positive selection in C₄ clades.

Eight codons were identified as evolving under positive selection in C₄ clades (fig. 2). The posterior probabilities of being under positive selection (calculated as the probability to belong to class 2a or 2b; table 1) were greater than 0.999 for all these sites (table 2), whereas they were below 0.2 for all the remaining *rbcl* codons (fig. 2). At the eight positively selected positions, the same amino acids recurrently appeared in independent C₄ lineages (fig. 3, supplementary fig. 1 and table 1, Supplementary Material online). For example, the codon at position 309 is a methionine in almost all considered C₃ monocots, whereas an

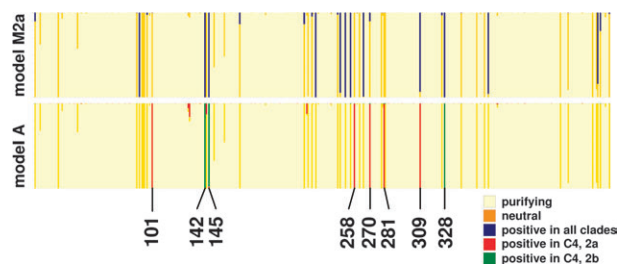


FIG. 2.—Selective pressures variation across sites of *rbcL*. The probability of assignment to the different codon classes, estimated by the Bayes empirical Bayes procedure (Yang et al. 2005), is presented for each *rbcL* site. Results of site model M2a are at the top, whereas those of branch site model A are at the bottom. Assignment to purifying selection in all branches is in pale yellow, to relaxed selection in all branches in orange, to positive selection in all branches in blue (specific to model M2a), to purifying selection in C_3 clades but positive selection in C_4 clades in red (class 2a of model A), and to neutral selection in C_3 clades but positive selection in C_4 clades in green (class 2b of model A). Detailed values for the different selection pressures are shown in table 1, and exact assignments are given in table 2. Positions of sites under positive selection in C_4 clades are numbered following *Zea mays* complete sequence, NC_001666.

isoleucine was present in 58 C_4 species belonging to 13 independent C_4 lineages (fig. 3; supplementary table 1, Supplementary Material online). The codon at position 281 is the one that underwent most parallel genetic changes, with an alanine in almost all C_3 species recurrently replaced by a serine in 80 C_4 species belonging to 16 independent C_4 lineages. One of the two CAM monocots (*T. bergeri*) also exhibits a serine at this position. Recurrent appearances of the same amino acids are also observed for the other positions (fig. 3, supplementary fig. 1, Supplementary Material online), and adaptive mutations have occurred several times independently within several of the C_4 lineages (supplementary fig. 1 and table 1, Supplementary Material online).

Table 2
Posterior Probabilities (PP) of Evolving under Positive Selection of Codons Identified in Models M2a and A

Position	Branch Site Model A		Site Model M2a
	PP, Class 2a ^a	PP, Class 2b ^b	PP, Positive Selection (2a + 2b)
91	0.0000	0.0000	0.9966
101	1.0000	0.0000	0.0012
142	0.0000	0.9993	0.9992
145	0.0242	0.9758	0.9990
228	0.0000	0.0000	1.0000
251	0.0000	0.0001	1.0000
255	0.0000	0.0047	1.0000
258	1.0000	0.0003	0.0003
265	0.0000	0.0014	0.9916
270	0.9924	0.0074	0.9998
281	0.9945	0.0055	0.0007
309	1.0000	0.0000	0.9345
328	0.0000	1.0000	1.0000
362	0.0000	0.0001	0.9573

^a Purifying selection in C_3 clades and positive selection in C_4 ones (see table 1).

^b Relaxed selection in C_3 clades and positive selection in C_4 ones (see table 1).

^c Posterior probabilities greater than 0.95 are highlighted in bold.

The model assuming positive selection on all branches was also significantly better than the model without positive selection (models M1a vs. M2a; chi-squared = 161, df = 2, P value < 0.0001). Thus, models M2a and A both point to positive selection on *rbcL* but on all branches for M2a and on C_4 branches only for A. These two models have the same parameters number and are not nested, disabling comparison through LRT. However, the use of the AIC clearly points to a better fit of the A model (differences in AIC = 73.42), demonstrating that positive selection only on C_4 branches better explains *rbcL* evolution. In the M2a model, nine sites were attributed to positive selection with a posterior probability greater than 0.95 (positions 91, 142, 145, 228, 251, 255, 265, 328, and 362; fig. 2; table 2). Of these, only three were also identified in branch site model A (sites 142, 145, and 328; table 2). Note that in A model, these were the sites belonging to class 2b, supposed to have evolved under relaxed selection in C_3 species and under positive selection in C_4 ones (fig. 2).

Discussion

Selective Switch and C_4 Adaptive Mutations

We showed that selective pressures changed after the evolution of C_4 photosynthesis, an event that first occurred 25–32 MYA (Christin et al. 2008). Some Rubisco sites that were under strong purifying or relaxed selection in C_3 ancestral species evolved under positive selection in C_4 plants (table 1; fig. 2). This result confirms that C_4 photosynthesis switched the optimal resolution of Rubisco specificity–efficiency trade-off. The different catalytic properties of C_4 Rubisco demonstrated in different C_4 lineages (Yeoh et al. 1980, 1981; Seemann et al. 1984; Hudson et al. 1990; Sage 2002; Kubien et al. 2008; fig. 1) are likely the consequence of mutations at those sites presenting higher rates of fixed nonsynonymous mutations (i.e., evolving under positive selection).

Other selective pressures could have driven Rubisco molecular evolution in non- C_4 plant lineages (Kapralov and Filatov 2006, 2007). For instance, C_3 plants exhibit different Rubisco catalytic properties following the mean temperature that they encounter, that is, C_3 plants from cooler habitats having a Rubisco with a higher turnover rate, like C_4 plants (Sage 2002). Specificity factors of Rubiscos of C_3 plants also vary according to the environmental xericity, that is, C_3 plants from more arid habitats having a Rubisco with a higher CO_2 specificity (Galmés et al. 2005). This could explain the good fit of codon model implementing positive selection in all clades (already shown by Kapralov and Filatov [2007]) as well as the occurrence of mutations supposed to lead to different catalytic properties in non- C_4 lineages (fig. 3; supplementary fig. 1 and table 1, Supplementary Material online). Sites 142, 145, and 328 underwent many nonsynonymous changes in C_3 plants (supplementary fig. 1 and table 1, Supplementary Material online). At position 328, many C_3 species exhibit the C_4 -dominant serine residue (supplementary fig. 1 and table 1, Supplementary Material online). This suggests that these three sites are not directly linked to the C_4 pathway but could be involved in the more general adaptation of

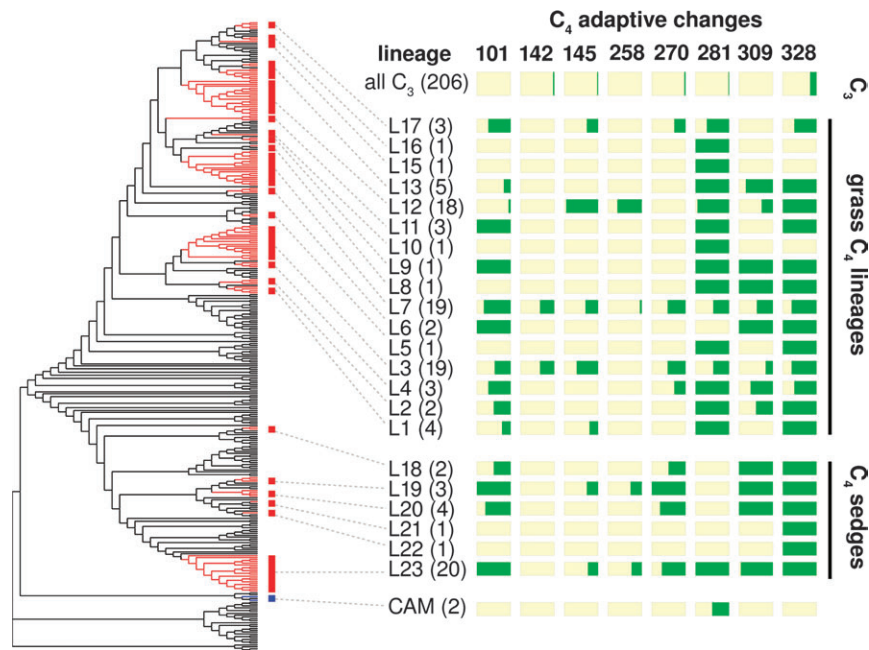


FIG. 3.—Codons of *rbcL* that converged in the different C₄ lineages. The phylogenetic tree obtained through Bayesian inference from *rbcL* and *ndhF* genes is shown on the left. C₄ lineages are compressed and indicated with a red square. Grass C₄ lineages are numbered as in Christin et al. (2008) and can be found, together with the clade numbering for Cyperaceae in supplementary figure 1 (Supplementary Material online). The number of sequences considered for each group is indicated into brackets. On the right, the green boxes represent the proportion of the sequences with the C₄-prevailing residues for each independent C₄ lineage, for the two CAM species, and for all C₃ plants grouped together (details available in supplementary table 1, Supplementary Material online). This proportion is given for the eight positions that evolved under positive selection in C₄ monocots (numbered following *Zea mays* complete sequence, NC_001666). Complete phylogenetic tree and amino acid under positive selection for each species are provided in supplementary figure 1 (Supplementary Material online).

specificity and/or catalytic efficiency to the variations of environmental xericity and warmness (e.g., Sage 2002; Galmés et al. 2005). In the vicinity of site 328, the histidine 327 is involved in the catalytic site (Kellogg and Juliano 1997; Andersson 2008). Residues 331–338, including active site residues 334 and 335, form the loop 6, whose conformation influences the CO₂/O₂ specificity (Andersson 2008). Mutation of site 328, by affecting the movement of loop 6, could alter the interaction with the six-carbon intermediates and thus change the CO₂/O₂ specificity of the Rubisco.

Codons 101, 258, 270, 281, and 309 had a high probability of being under positive selection in C₄ clades (in branch site model A) but not in C₃ ones (in site model M2a). In addition, these five sites evolved under purifying selection in C₃ clades (fig. 2), where they underwent almost no changes (fig. 3; supplementary fig. 1 and table 1, Supplementary Material online), strongly suggesting that amino acid replacements at these positions were specifically driven by selective switches toward higher catalytic efficiency that followed the evolution of the C₄ trait. These sites are thus excellent candidates for genetic engineering of Rubisco.

Out of the five codons that shifted from purifying to positive selection in C₄ plants, three positions (101, 258, and 309) exhibit the same residues in many C₄ plants, cyanobacteria, and anaerobic bacteria (organisms that all have Rubiscos with higher catalytic efficiency; Tcherkez et al. 2006) but not in C₃ monocots. At position 309, the replacement of the isoleucine (as in most C₄ plants; supplementary

table 1, Supplementary Material online) by a methionine (present in all studied C₃ plants; supplementary table 1, Supplementary Material online) in the *Synechococcus* cyanobacterium had no apparent effect on the catalytic efficiency (Morell et al. 1972), which indicates that this mutation is not responsible per se for the Rubisco kinetic variations. Mutation in the Rubisco of the green alga *Chlamydomonas* of site 258, in combination with sites 256 and 265, led to a decrease in both catalytic efficiency and CO₂/O₂ specificity (Du et al. 2003). However, when additional sites were mutated, kinetics typical of land plants Rubisco were obtained (Spreitzer et al. 2005). Site 258, which participates to the interactions between the large subunit and the small subunit (Kellogg and Juliano 1997; Spreitzer et al. 2005), could thus play a role in the determinism of Rubisco kinetics (Yu et al. 2005). These Rubisco experimental mutations showed that the kinetic variations are not caused by single amino acid changes but the joint effects of different mutations (as seen in the case of *Chlamydomonas* mutants; Du et al. 2003; Spreitzer et al. 2005). Similarly, the C₄-specific amino acids highlighted in this study are found alone only in few species. This suggests that they could strongly interact with each other to cause the C₄-specific Rubisco kinetics.

The exact effect of mutations on sites highlighted in this study as under C₄-specific positive selection remains largely unknown. Rubisco has a very complex structure and minimal changes, even far from the catalytic sites, can strongly alter the kinetic properties (Spreitzer et al. 2005; Karhehabadi et al. 2007; Andersson 2008). The

recurrent and numerous appearances of the same residues in independent C₄ lineages highlight the evolutionary significance of these sites. Their effects have to be depicted through structural analyses, and these sites should be mutated, both alone and in combinations.

Genetic Convergence

The recurrence of some amino acid changes in independent C₄ lineages highlights the limited ways by which natural selection can act to improve a specific enzyme (Weinreich et al. 2006) and adds to other evidence for the occurrence of genetic convergences during evolution (Wood et al. 2005; Christin et al. 2007). After billions years of evolution, the largest part of the amino acids that compose an enzyme such as Rubisco has been optimized, and most of the potential amino acid changes are detrimental because of a reduction or suppression of enzyme function (Kellogg and Juliano 1997). This accounts for 88% of Rubisco sites evolving under strong purifying selection, with 70-fold less amino acid replacements than expected by chance (table 1). The spectrum of amino acid mutations susceptible to improve an enzyme of a species facing a new environment, such as the CO₂-rich bundle-sheath cells of C₄ plants, is thus very limited. The C₄-specific amino acids that we observe today thus represent the very low proportion of all the past mutations that were beneficial for C₄ plants and persisted through numerous generations.

Implications for Bioengineering

The repeated mutation of the same amino acids in independent C₄ lineages (fig. 3) emphasizes the importance of these mutations for the C₄-specific Rubisco kinetics. This is of greatest interest for the current attempts to improve C₃ crops yield through the introgression of C₄ characteristics (Long et al. 2006; Raines 2006; Hibberd et al. 2008), efforts that particularly focus on the improvement of Rubisco (Parry et al. 2003, 2007; Zhu et al. 2004). However, if the increase in catalytic efficiency induces a decrease in affinity of the CO₂ substrate, introgressing the C₄ Rubisco gene copy in C₃ plants would lead to higher photorespiration rates. Thus, in current atmospheric conditions, a Rubisco with C₄ characteristic would decrease C₃ plants yield instead of the desired improvement of C₃ crops productivity (Zhu et al. 2004). However, the predicted future increase of atmospheric CO₂ concentrations would drastically change the efficiency/specificity balance. Atmospheric CO₂ is predicted to reach 2- to 4-fold higher concentrations than in pre-Industrial climates within 50–100 years (Swart et al. 2002; IPCC 2007). In such a CO₂-enriched atmosphere, photorespiration would decrease thanks to the higher CO₂:O₂ ratio. A decreased Rubisco affinity for CO₂ would thus not hamper C₃ plants, enabling an increase of their yield through an improvement of Rubisco catalytic efficiency (Brainbridge et al. 1995; Zhu et al. 2004; Ainsworth et al. 2008). Bioengineering of C₃ Rubisco based on our knowledge of C₄ *rbcL* changes appears as a promising alternative to the conversion of C₃ crops to C₄ photosynthesis

(Long et al. 2006). The comparative phylogenetic approach adopted here was powerful for identifying genetic adaptive changes that are shared between distant species facing the same environmental pressures. The convergent nature of C₄ photosynthesis provides natural replicates of enzyme adaptation under high CO₂ concentrations. The identification of key amino acids putatively responsible for the kinetics variation between C₃ crops and closely related C₄ species could be a first step on the road to human-directed improvement of plant species for the predicted future CO₂-rich atmosphere.

Supplementary Material

Supplementary figure 1 and tables 1–3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

This work was funded by the Swiss National Science Foundation (SNSF) grant 3100AO-105886. It partially results from a Switzerland–South Africa joint meeting, funded by the NSFS grant IB32A0-112804. We thank the Swiss Institute of Bioinformatics for access to the Vital-IT computers cluster. P. Danton, F. Dessimoz, N. Fumeaux, B. Gehrke, F. Jacquemoud, and O. Ueno provided very useful samples. L. Büchi, E.E. Farmer, and E. Samaritani made useful comments on the earlier versions of the manuscript.

Literature Cited

- Ainsworth EA, Rogers A, Leakey ADB. 2008. Targets for crop biotechnology in a future high-CO₂ and high-O₃ world. *Plant Physiol.* 147:13–19.
- Andersson I. 2008. Catalysis and regulation in Rubisco. *J Exp Bot.* 59:1555–1568.
- Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL, Coetzee LL, Beukes NJ. 2004. Dating the rise of atmospheric oxygen. *Nature.* 427:117–120.
- Brainbridge G, Madgwick P, Parmar S, Mitchell R, Paul M, Pitts J, Keys AJ, Parry MAJ. 1995. Engineering Rubisco to change its catalytic properties. *J Exp Bot.* 46:1269–1276.
- Bruhl JJ, Wilson KL. 2007. Towards a comprehensive survey of C₃ and C₄ photosynthetic pathways in Cyperaceae. *Aliso.* 23:99–148.
- Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N. 2008. Oligocene CO₂ decline promoted C₄ photosynthesis in grasses. *Curr Biol.* 18:37–43.
- Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C₄ photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol.* 17:1241–1247.
- Douce R, Heldt HW. 2000. Photorespiration. In: Leegood RC, Sharkey TD, von Caemmerer S, editors. *Photosynthesis: physiology and metabolism*. Dordrecht (The Netherlands): Kluwer Academic Publishers. p. 115–136.
- Du YC, Peddi SR, Spreitzer RJ. 2003. Assessment of structural and functional divergence far from the large subunit active site of ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Biol Chem.* 278:49401–49405.

- Duvall MR, Davis JI, Clark LG, Noll JD, Goldman DH, Sánchez-Ken JG. 2007. Phylogeny of the grasses (Poaceae) revisited. *Aliso*. 23:237–247.
- Evans JR, Poorter H. 2001. Photosynthetic acclimation of plants to growth irradiance: the relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain. *Plant Cell Environ*. 24:755–767.
- Galmés J, Flexas J, Keys AJ, Cifre J, Mitchell RAC, Madgwick PJ, Haslam RP, Medrano H, Parry MAJ. 2005. Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. *Plant Cell Environ*. 28:571–579.
- Ghannoum O, Evans JR, Chow WS, Andrews TJ, Conroy JP, von Caemmerer S. 2005. Faster Rubisco is the key to superior nitrogen-use efficiency in NADP-malic enzyme relative to NAD-malic enzyme C₄ grasses. *Plant Physiol*. 137:638–650.
- Goldman N, Yang ZH. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 11:725–736.
- Hibberd JM, Sheehy JE, Langdale JA. 2008. Using C₄ photosynthesis to increase the yield of rice—rationale and feasibility. *Curr Opin Plant Biol*. 11:228–231.
- Hudson GS, Mahon JD, Anderson PA, Gibbs MJ, Badger MR, Andrews TJ, Whitfeld PR. 1990. Comparisons of *rbcl* genes for the large subunit of ribulose-bisphosphate carboxylase from closely related C₃ and C₄ plant species. *J Biol Chem*. 265:808–814.
- IPCC. 2007. Climate change 2007: the physical basis. Contribution of Working Group I to the Fourth Annual Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge (UK): Cambridge University Press.
- Jordan DB, Ogren WL. 1981. Species variation in the specificity of ribulose-bisphosphate carboxylase-oxygenase. *Nature*. 291:513–515.
- Kane HJ, Viil J, Entsch B, Paul K, Morell MK, Andrews TJ. 1994. An improved method for measuring the CO₂/O₂ specificity of ribulosebisphosphate carb0ylase-oxygenase. *Aust J Plant Physiol*. 21:449–461.
- Kapralov MV, Filatov DA. 2006. Molecular adaptation during adaptive radiation in the Hawaiian endemic genus *Schiedea*. *PLoS ONE*. 1:e8.
- Kapralov MV, Filatov DA. 2007. Widespread selection in the photosynthetic Rubisco enzyme. *BMC Evol Biol*. 7:73.
- Karhehabadi S, Satagopan S, Taylor TC, Spreitzer RJ, Andersson I. 2007. Structural analysis of altered large-subunit loop-6/carboxy-terminus interactions that influence catalytic efficiency and CO₂/O₂ specificity of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Biochemistry*. 46:11080–11089.
- Kellogg EA, Juliano ND. 1997. The structure and function of RuBisCO and their implications for systematic studies. *Am J Bot*. 84:413–428.
- Kubien DS, Whitney SM, Moore PV, Jesson LK. 2008. The biochemistry of Rubisco in *Flaveria*. *J Exp Bot*. 59:1767–1777.
- Long SP. 1999. Environmental responses. In: Sage RF, Monson RK, editors. C₄ plant biology. San Diego (CA): Academic Press. p. 215–249.
- Long SP, Zhu XG, Shawna LN, Ort DR. 2006. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ*. 29:315–330.
- Ludwig LJ, Canvin DT. 1971. The rate of photorespiration during photosynthesis and the relationship of the substrate of light respiration to the products in sunflower leaves. *Plant Physiol*. 48:712–719.
- Lüttge U. 2004. Ecophysiology of Crassulacean acid metabolism (CAM). *Ann Bot*. 93:629–652.
- Monteith JL. 1977. Climate and efficiency of crop production in Britain. *Philos Trans R Soc Lond B Biol Sci*. 281:277–294.
- Morell MK, Kane HJ, Hudson GS, Andrews TJ. 1972. Effects of mutations at residue 309 of the large subunit of ribulosebisphosphate carboxylase from *Synechococcus* PCC 6301. *Arch Biochem Biophys*. 299:295–301.
- Nisbet EG, Grassineau NV, Howe CJ, Abell PI, Regelous M, Nisbet RER. 2007. The age of Rubisco: the evolution of oxygenic photosynthesis. *Geobiology*. 5:311–335.
- Parry MAJ, Andralojc PJ, Mitchell RAC, Madgwick PJ, Keys AJ. 2003. Manipulation of Rubisco: the amount, activity, function and regulation. *J Exp Bot*. 54:1321–1333.
- Parry MAJ, Madgwick PJ, Carvalho JFC, Andralojc PJ. 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J Agric Sci*. 145:31–43.
- Portis AR, Parry MAJ. 2007. Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective. *Photosynth Res*. 94:121–143.
- Raines CA. 2006. Transgenic approaches to manipulate the environmental responses of the C₃ carbon fixation cycle. *Plant Cell Environ*. 29:331–339.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Sage RF. 2002. Variation in the *k_{cat}* of Rubisco in C₃ and C₄ plants and some implications for photosynthetic performance at high and low temperature. *J Exp Bot*. 53:609–620.
- Sage RF. 2004. The evolution of C₄ photosynthesis. *New Phytol*. 161:341–370.
- Seemann JR, Badger MR, Berry JA. 1984. Variations in the specific activity of ribulose-1,5-bisphosphate carboxylase between species utilizing different photosynthetic pathways. *Plant Physiol*. 74:791–794.
- Spreitzer RJ, Peddi SR, Satagopan S. 2005. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proc Natl Acad Sci USA*. 22:17225–17230.
- Spreitzer RJ, Salvucci ME. 2002. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu Rev Plant Biol*. 53:449–475.
- Swart R, Mitchell J, Morita T, Raper S. 2002. Stabilisation scenarios for climate impact assessment. *Glob Environ Change*. 12:155–165.
- Tcherkez GBB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci USA*. 103:7246–7251.
- von Caemmerer S, Furbank RT. 2003. The C₄ pathway: an efficient CO₂ pump. *Photosynth Res*. 77:191–207.
- von Caemmerer S, Quick WP. 2000. Rubisco: physiology in vivo. In: Leegood RC, Sharkey TD, von Caemmerer S, editors. *Physiology and metabolism*. Dordrecht (The Netherlands): Kluwer Academic Publishers. p. 173–211.
- Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*. 312:111–114.
- Wood TE, Burke JM, Rieseberg LH. 2005. Parallel genotypic adaptation: when evolution repeats itself. *Genetica*. 123:157–170.
- Yang ZH. 2007. PAML: a program package for phylogenetic analysis by maximum likelihood. Version 4. *Mol Biol Evol*. 24:1586–1591.
- Yang ZH, Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol*. 15:496–503.
- Yang ZH, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*. 19:908–917.

- Yang ZH, Nielsen R, Goldman N, Krabbe Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155:431–449.
- Yang ZH, Nielsen R, Hasegawa M. 1998. Models of amino acid substitutions and applications to mitochondrial protein evolution. *Mol Biol Evol*. 15:1600–1611.
- Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acids sites under positive selection. *Mol Biol Evol*. 22:1107–1118.
- Yeoh HH, Badger MR, Watson L. 1980. Variations in $K_m(\text{CO}_2)$ of ribulose-1,5-bisphosphate carboxylase among grasses. *Plant Physiol*. 66:1110–1112.
- Yeoh HH, Badger MR, Watson L. 1981. Variations in kinetic properties of ribulose-1,5-bisphosphate carboxylase among plants. *Plant Physiol*. 67:1151–1155.
- Yu GX, Park BH, Chandramohan P, Geist A, Samatova NF. 2005. An evolution-based analysis scheme to identify CO_2/O_2 specificity-determining factors for ribulose 1,5-bisphosphate carboxylase/oxygenase. *Protein Eng Des Sel*. 18:589–596.
- Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol*. 22:2472–2479.
- Zhu XG, Portis AR, Long SP. 2004. Would transformation of C_3 crops plants with foreign Rubisco increase productivity? A computational analysis extrapolating from kinetic properties to canopy photosynthesis. *Plant Cell Environ*. 27:155–165.

David Irwin, Associate Editor

Accepted August 6, 2008