PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

*Phil. Trans. R. Soc. B* (2012) **367**, 483–492
doi:10.1098/rstb.2011.0145

*Research*

# Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide

**J. N. Young[1], R. E. M. Rickaby[1,*], M. V. Kapralov[2] and D. A. Filatov[2]**

[1]*Department of Earth Sciences, Oxford University, South Parks Road, Oxford OX1 3AN, UK*
[2]*Department of Plant Sciences, Oxford University, South Parks Road, Oxford OX1 3RB, UK*

Rubisco, the most abundant enzyme on the Earth and responsible for all photosynthetic carbon fixation, is often thought of as a highly conserved and sluggish enzyme. Yet, different algal Rubiscos demonstrate a range of kinetic properties hinting at a history of evolution and adaptation. Here, we show that algal Rubisco has indeed evolved adaptively during ancient and distinct geological periods. Using DNA sequences of extant marine algae of the red and Chromista lineage, we define positive selection within the large subunit of Rubisco, encoded by *rbcL*, to occur basal to the radiation of modern marine groups. This signal of positive selection appears to be responding to changing intracellular concentrations of carbon dioxide ($CO_2$) triggered by physiological adaptations to declining atmospheric $CO_2$. Within the ecologically important Haptophyta (including coccolithophores) and Bacillariophyta (diatoms), positive selection occurred consistently during periods of falling Phanerozoic $CO_2$ and suggests emergence of carbon-concentrating mechanisms. During the Proterozoic, a strong signal of positive selection after secondary endosymbiosis occurs at the origin of the Chromista lineage (approx. 1.1 Ga), with further positive selection events until 0.41 Ga, implying a significant and continuous decrease in atmospheric $CO_2$ encompassing the Cryogenian Snowball Earth events. We surmise that positive selection in Rubisco has been caused by declines in atmospheric $CO_2$ and hence acts as a proxy for ancient atmospheric $CO_2$.

**Keywords:** Rubisco; adaptive evolution; algae; palaeoclimate; carbon dioxide

## 1. INTRODUCTION

The living ecosystem and carbon dioxide ($CO_2$) and oxygen ($O_2$) levels in the atmosphere are inexorably linked through tight feedback mechanisms. The advent of oxygenic photosynthesis around 2.4 Ga increased atmospheric and surface ocean $O_2$ levels and reduced atmospheric $CO_2$. However, $O_2$ levels remained low for the next billion years despite declining $CO_2$ [1], partly attributed to anoxygenic photosynthesis preventing oxygenation of the deep ocean [2]. It was not until the Neoproterozoic when oxygenic photosynthesis became dominant, hypothesized to contribute to the second rise of $O_2$ which paved the way for evolution of complex multi-cellular life [1].

Biological innovations certainly exert an influence on the atmosphere, but atmospheric composition also drives biological adaptations. The rate-limiting step of photosynthesis, $CO_2$ fixation, is catalysed by the enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco, EC 4.1.1.39). During oxygenic photosynthesis, Rubisco catalyses the two competitive reactions: $CO_2$ fixation for photosynthesis (carboxylation) and energy-wasting photorespiration using $O_2$ (oxygenation). The ability of a particular Rubisco to discriminate between non-polar, structurally similar substrates of $CO_2$ and $O_2$ is determined by the kinetic properties of Rubisco and the $CO_2$ and $O_2$ concentrations at the catalytic site of the enzyme, denoted as the specificity factor ($\Omega$) [3]:

$$\Omega = \frac{V_c K_o}{V_o K_c} \times \frac{[CO_2]}{[O_2]},$$

where $V_c$ and $V_o$ are maximal velocities of the carboxylase and oxygenase reactions and $K_c$ and $K_o$ are the Michaelis constants for $CO_2$ and $O_2$. Thermodynamic constraints dictate a trade-off between carboxylation velocity ($V_c$) and affinity for $CO_2$ that has led to suggestions that despite being conserved and sluggish, Rubisco is optimized to its physical environment [4]. The variation in Rubisco catalytic properties found within photosynthetic eukaryotes hints that it has undergone adaptations to low $CO_2$ [5–8], but little is known about the detailed timing of evolution of Rubisco and its relationship to the environmental change.

We aim to define the history of adaptation of Rubisco in red and Chromista algae. Our study focused on these oxygenic photosynthesizing algae because they all possess a red chloroplast containing

* Author for correspondence (rosalind.rickaby@earth.ox.ac.uk).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2011.0145 or via http://rstb.royalsocietypublishing.org.

One contribution of 12 to a Theme Issue 'Atmospheric $CO_2$ and the evolution of photosynthetic eukaryotes: from enzymes to ecosystems'.

the same Form 1D Rubisco [9]. Their chloroplast evolutionary history stemmed from a primary endosymbiotic event that resulted in the red algae (Rhodophyta and Cyanidiales), and subsequent secondary endosymbiosis to form the Chromalveolata [10]. Chromalveolata includes the Chromista (Haptophyta, Stramenopiles and Cryptophyta) [11] and the Alveolata. However, as there is debate on the plastid monophyly of the Chromalveolata (for review, see Keeling [12]), we focused only on the Chromista which have a well-supported single secondary endosymbiotic origin of their plastids [13]. Furthermore, many representatives of the Chromista algae, such as diatoms and Haptophyta, dominate the modern ocean and display an extensive fossil history [14,15], making it possible to date the periods of adaptive evolution in Rubisco. Form 1D Rubisco protein is made up of eight large and eight small subunits. The active sites of Rubisco are formed by large subunits which are encoded by the chloroplast gene *rbcL*.

To infer adaptation of Rubisco, we reconstructed the chloroplast evolutionary history of the red and Chromista algae as a template on which we identified positive selection in *rbcL* by applying phylogenetic analysis of maximum likelihood (PAML) [16]. PAML identifies positive selection by comparing the substitution rates of mutations that do and do not affect protein sequence ($d_N$ and $d_S$, respectively) along a phylogeny. The former is likely to affect the survival of the organism, whereas the latter is neutral, or nearly neutral; so $d_N/d_S > 1$, $d_N/d_S = 1$ and $d_N/d_S < 1$ indicate positive, neutral and purifying selection, respectively. Bayesian tree reconstruction and dating were conducted using a number of fossil calibrations, allowing the timing of adaptive events to be constrained between branch nodes providing an evolutionary history to around 1.5 Ga.

## 2. MATERIAL AND METHODS

### (a) *DNA extraction and amplification of* rbcL

DNA was extracted from Haptophyta and diatom cultures maintained at Marine Biological Association, Plymouth, UK using previously published methods [17]. About 95 per cent length of the coding region of the chloroplast *rbcL* gene was PCR amplified using Biomix (Bioline, MA, USA) and primers (electronic supplementary material, table S1), and the following PCR conditions: initial cycle of 95°C for 2 min, 55°C for 30 s, 72°C for 30 s followed by 40 cycles of 92°C for 30 s, 53°C for 30 s and 72°C for 3 min, with final elongation of 72°C for 10 min. DNA sequencing was performed using ABI BigDye v. 3.1 and capillary sequencers 3700 and 3730xl (Applied Biosystems Inc., CA, USA). Sequences of the *rbcL* gene for 17 species were uploaded to NCBI GenBank under accession numbers HQ656822–HQ656838. The new *rbcL* sequences were combined with a further 227 *rbcL* sequences from NCBI GenBank (electronic supplementary material, table S2) and aligned using ClustalX [18], followed by minor manual editing of the alignment and assignment of coding region using ProSeq3 [19].

### (b) *Phylogenetic and molecular clock analysis*

The phylogeny shown in figure 1 was reconstructed using four plastid protein-coding genes used by Yoon *et al.* [22] with a few changes: removal or replacement of species owing to short or poor quality *rbcL* sequences and additional sequences of heterokonts, Haptophyta and coralline red algae. The Haptophyta tree (electronic supplementary material, figure S1) was reconstructed only using *rbcL* sequences. The diatom tree (electronic supplementary material, figure S2) was constructed with the 18S *rRNA* and *rbcL* genes. Bayesian tree reconstruction and dating were conducted using BEAST software [23]. The calibration constraints used for each tree are summarized in electronic supplementary material, table S3 [13–15,24–26]. Phylogeny reconstruction assumed the general time-reversible substitution model [27] with gamma-distributed rate heterogeneity. For protein-coding genes, we assumed separate rates for the three codon positions. We conducted runs with chain lengths of $5 \times 10^7$ to $10^8$ steps assuming the Yule model of speciation process and both clock and uncorrelated lognormal-relaxed clock across the tree [28,29] with similar results. Given the standard deviation of the uncorrelated lognormal-relaxed clock (ucld.stdev parameter) was consistently below 1, the data appeared to fit the molecular clock model quite well, hence the latter was used for dating of individual nodes. The convergence of parameter estimates was checked using Tracer [30]. The data were saved every $10^3$ steps and the first $5 \times 10^3$ trees were ignored as a burn-in. TreeAnnotator v. 1.5.4 was used to summarize the post-burn-in trees, and the maximum credibility tree along with 95% probability density of ages was visualized in FigTree v. 1.3.1 [31].

### (c) *Phylogenetic analysis of maximum likelihood*

Detection of positive selection in the *rbcL* gene was conducted using nested maximum-likelihood models A and A1 allowing for variation in the ratio of non-synonymous to synonymous substitution rates ($d_N/d_S$) across codons and across branches implemented in PAML v. 4 [16]. The analysis was conducted assuming the phylogenies in figure 1 and electronic supplementary material, figures S1 and S2. The null model A1 allows $d_N/d_S$ ratios to vary both among sites and among branches; it allows $0 < d_N/d_S < 1$ and $d_N/d_S = 1$ for both foreground and background branches, and also two additional classes of codons with fixed $d_N/d_S = 1$ on foreground branches while restricted as $0 < d_N/d_S < 1$ and $d_N/d_S = 1$ on background branches. The alternative A model allows $0 < d_N/d_S < 1$ and $d_N/d_S = 1$ for both foreground and background branches, and also two additional classes of codons under positive selection with $d_N/d_S > 1$ on foreground branches while restricted as $0 < d_N/d_S < 1$ and $d_N/d_S = 1$ on background branches. One branch at a time was labelled as a foreground branch with allowed positive selection, whereas all other branches were labelled as background branches with no positive selection allowed. Likelihood-ratio tests (LRTs) were repeated with a different branch labelled as a foreground branch until all branches had been tested [32,33]. The significance of the A/A1 LRTs was calculated assuming

that twice the difference in the log of maximum likelihood between the two models was distributed as a chi-square distribution with one degree of freedom. The Bonferroni correction procedure implemented in SISA-Binomial [34] was used to correct the statistical significance for multiple comparisons.

These tests were applied to three datasets, red and Chromista algae, Haptophyta and diatoms, consisting of 39, 28 and 30 species, respectively (figure 1 and electronic supplementary material, figures S1 and S2).

Phylogenies shown in figure 1 and electronic supplementary material, figures S1 and S2 were also used to detect positive selection within a group using the site test M1a/M2a [35]. Four other red and Chromista algal groups were also tested using M1a/M2a LRTs consisting of 49 Rhodophyta, 45 Phaeophyceae, 28 Cryptophyta, 33 Chrysophyceae and Synurophyceae *rbcL* sequences (electronic supplementary material, figure S3 and table S2). Phylogenies of these groups were reconstructed using Bayesian method (MrBayes v. 3.1, GTR model with gamma-distributed rate variation across sites and a proportion of invariable sites with at least $10^6$ runs) [33]. In M1a/M2a LRTs, the null model M1a (nearly neutral) that allows $0 \leq d_N/d_S \leq 1$ was compared with the M2a model (same as the M1a model plus an extra class under positive selection with $d_N/d_S \geq 1$). The significance of the M1a/M2a LRTs was calculated assuming that twice the difference in the log of maximum likelihood between the two models was distributed as a chi-square distribution with two degrees of freedom.

The Bayes empirical Bayes method [36] implemented in PAML was used to calculate the posterior probabilities that particular sites fall into classes with different $d_N/d_S$ and to identify sites with a high probability of being under positive selection (having $d_N/d_S > 1$). Rubisco sites under positive selection were numbered against the spinach (*Spinacia oleracea*) sequence for comparison with other studies. Residue location in the tertiary structure was visualized using the structural data file of spinach Rubisco (1RBO) available on RCSB Protein Data Bank with DeepView-Swiss PdbViewer v. 3.7 [37].

### (d) *Geological history of atmospheric carbon dioxide and oxygen*

Ranges of atmospheric $CO_2$ during the Proterozoic were taken from the literature [38–41]. Phanerozoic $CO_2$ was reconstructed from modelled (GEOCARBSULF, incorporating the full variability of basalt/granite ratios [42]) and proxy ($CO_2$ estimates from $\delta^{13}C$ values from phytoplankton, liverworts, boron and stomatal indices as compiled by Royer *et al.* [43] with revised $\delta^{13}C$ palaeosol data [44], binned into 10 Ma and 1 s.d.) data. Proterozoic $O_2$ levels were taken from Canfield [1]. Phanerozoic $O_2$ was derived from models [45–47] and proxies ([48], showing 1 s.d.).

## 3. RESULTS AND DISCUSSION

### (a) *Positive selection of algal Rubisco occurred during distinct geological periods*

The analysis of individual branches revealed that positive selection was restricted to branches basal to the radiation of the red and Chromista algae between 1.56 and 0.41 Ga (average ages of the nodes; for the range at each node see figure 1). Positive selection in *rbcL* was prior to the divergence of large algal taxonomic groups, but not within the groups. To further explore this pattern, we focused on two geochemically and ecologically important groups with relatively good fossil records: the Haptophyta (including coccolithophores) and Bacillariophyta (diatoms). Once again, positive selection was restricted to a few, deep branches. The Haptophyta (electronic supplementary material, figure S1) showed a strong signal of positive selection along the branch leading to, and basal to, the radiation of the Prymnesiophyceae (which includes the coccolithophores *Emiliania huxleyi*, *Coccolithus pelagicus* and colony-forming *Phaeocystis* spp.) between 375 and 285 Ma. Once this group diverged, there is little evidence of positive selection apart from the branch leading to the Phaeocystaceae. In the diatoms (electronic supplementary material, figure S2), adaptation also occurred in deep branches within radial centrics (including *Aulacoseira* and *Coscinodiscus* spp.) at 142–97 Ma and pennates (including *Fragilariopsis, Phaeodactylum* and *Pinnularia* spp.) at 73–56 Ma and in bipolar centrics (including *Thalassiosira*, *Chaetocera* and *Skeletonema* spp.) at 114–92 Ma.

Positive selection hints at changes of amino acids that would influence the catalytic activity of Rubisco. Indeed, positive selection does appear to explain the variation in known Rubisco specificity factors (figure 1*b*). However, specificity factors are only a ratio and are unable to reveal the true sensitivity of kinetic changes in Rubisco. There are few measurements of $K_c$ and $V_c$ in algae but the $K_c$ for both diatoms and Haptophyta are higher than Rhodophyta, which is higher again when compared with the Cyanidiales (figure 1*b*).

Using a Bayesian approach [36], we identified positively selected amino acid positions in *rbcL* in the basal branches where selection occurred. Only a few domains have been identified to influence activity or kinetics [9] and our sites under positive selection do not fall within these domains. However, there is some commonality to the residues under positive selection between land plants [49] and algae (figure 2), suggesting that these additional residues outside known important protein domains may also significantly affect Rubisco performance.

### (b) *Positive selection of Rubisco in algae contrasts to that of land plants*

Our results of positive selection only in the deep branches of the algal phylogeny contrasts with the ubiquitous positive selection found throughout the land plant groups [49]. To test whether this result may be owing to differences in the analytical approaches between studies, we applied the methodology previously used for plants [49] to an expanded set of red and Chromista algal groups, including Rhodophyta, Phaeophyta, Cryptophyta, Chrysophyceae and Synurophyceae, along with Haptophyta and diatoms with phylogenies constructed using *rbcL* sequences (electronic supplementary material, figure S3). This analysis confirmed the absence of positive selection

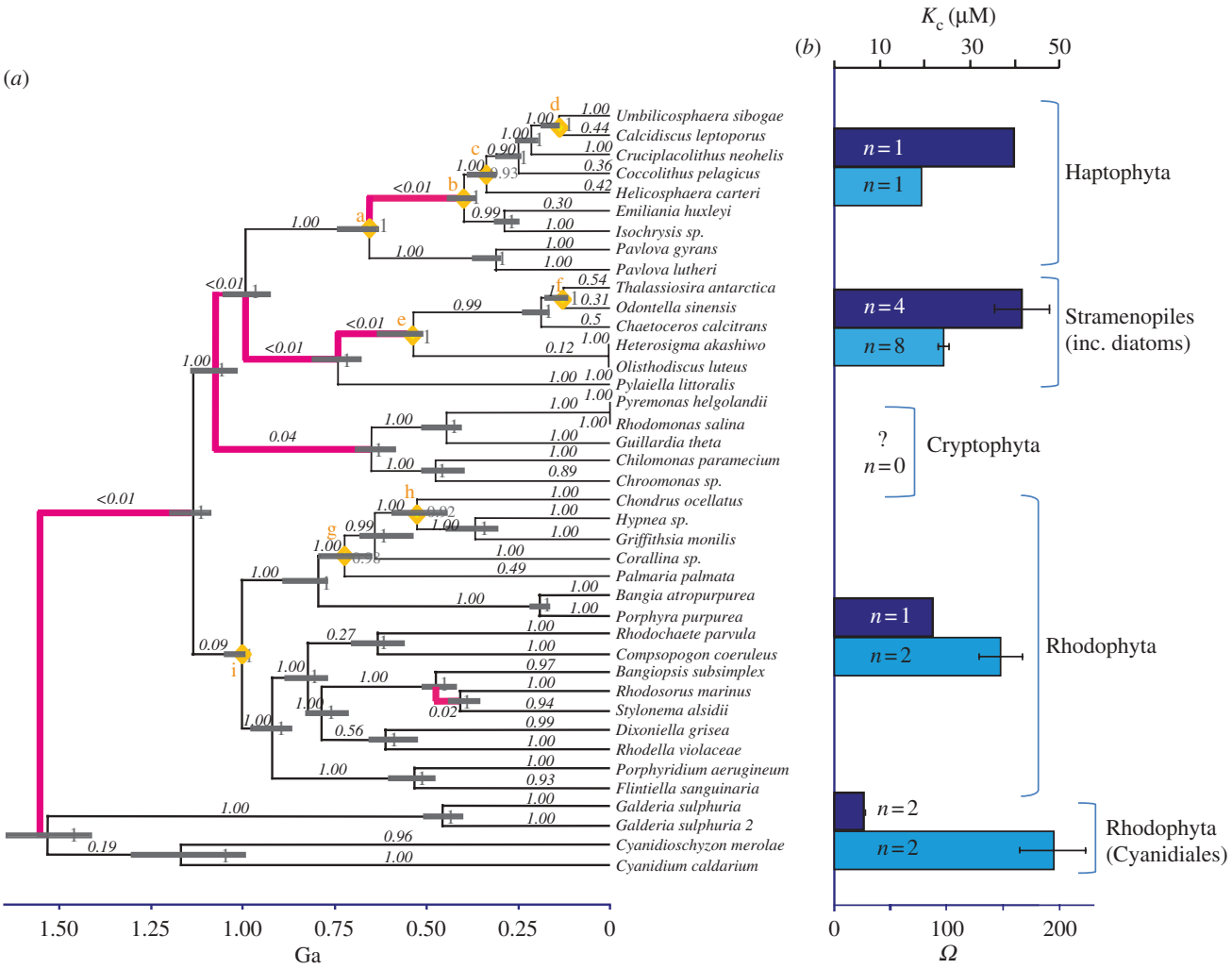


Figure 1. (a) *RbcL* phylogeny of red and Chromista algae with branches under positive selection (magenta) and those with no evidence for positive selection in Rubisco (black). Black italic numbers along branches are *p*-values for the probability of likelihood-ratio test values being as high as is seen by chance (see §2). Grey numbers are posterior probability values for individual nodes. Date estimates of nodes with grey bars denoting 95% probability density and yellow diamonds with corresponding letter are fossil calibration dates (electronic supplementary material, table S3). (*b*) Available measurements of Rubisco specificity factors ($\Omega$; light blue bars) [4,6,20] and $K_c$ (dark blue bars) [6,21] with standard error bars and number of species measured (*n*) are shown.

within algal taxonomic groups, with none of the expanded algal groups displaying a signal of positive selection using this test (table 1).

### (c) *Carbon dioxide as a driver for Rubisco adaptation*

There is evidence that Rubisco adaptation in land plants was driven by a number of environmental factors in addition to atmospheric composition, including aridity and high temperatures [50]. However, in the marine realm, we propose that it is the change in atmospheric $CO_2$ equilibrated with surface waters which acts as the ultimate driver of positive selection in Rubisco. As Rubisco is packaged within the chloroplast, the adaptive response of Rubisco must be driven by changing intracellular conditions. Intracellular $CO_2$ conditions are under physiological control, and as external $CO_2$ becomes increasingly limited algae are known to induce carbon-concentrating mechanisms (CCMs) to boost the internal supply of $CO_2$ to Rubisco [5]. However, this can also lead to an elevation in intracellular $O_2$ of levels up to 6–6.5 times external concentrations [51], owing to reduced leakiness of both $CO_2$ and $O_2$ [52,53]. While both intracellular $CO_2$ and $O_2$ levels can change by induction of a CCM, the trigger for the employment of CCMs is driven by changes in external $CO_2$. CCMs are regulated by external $CO_2$, and on geological scales of eukaryotic evolution, changes of atmospheric $CO_2$ are orders of magnitude greater than changes in $O_2$. Additionally, the Form 1D Rubisco present in the studied algal species appears to have reduced oxygenase potential and therefore a reduced tendency to be inhibited by $O_2$ [6]. Finally, optimization of Rubisco efficiency is driven by the trade-off between $V_c$ and $CO_2$ affinity, and therefore is in tune with intracellular $CO_2$ concentrations [54].

In land plants, it has been established that positive selection in *rbcL* emerges coincident with the development of a $C_4$ CCM which elevates $CO_2$ to almost saturation at the site of Rubisco [55,56]. This relaxes pressure for Rubisco to have a high affinity for $CO_2$ and therefore allows an increase in $V_c$, which results in increased photosynthetic efficiency as the plant requires less nitrogen to achieve a given $CO_2$ fixation capacity

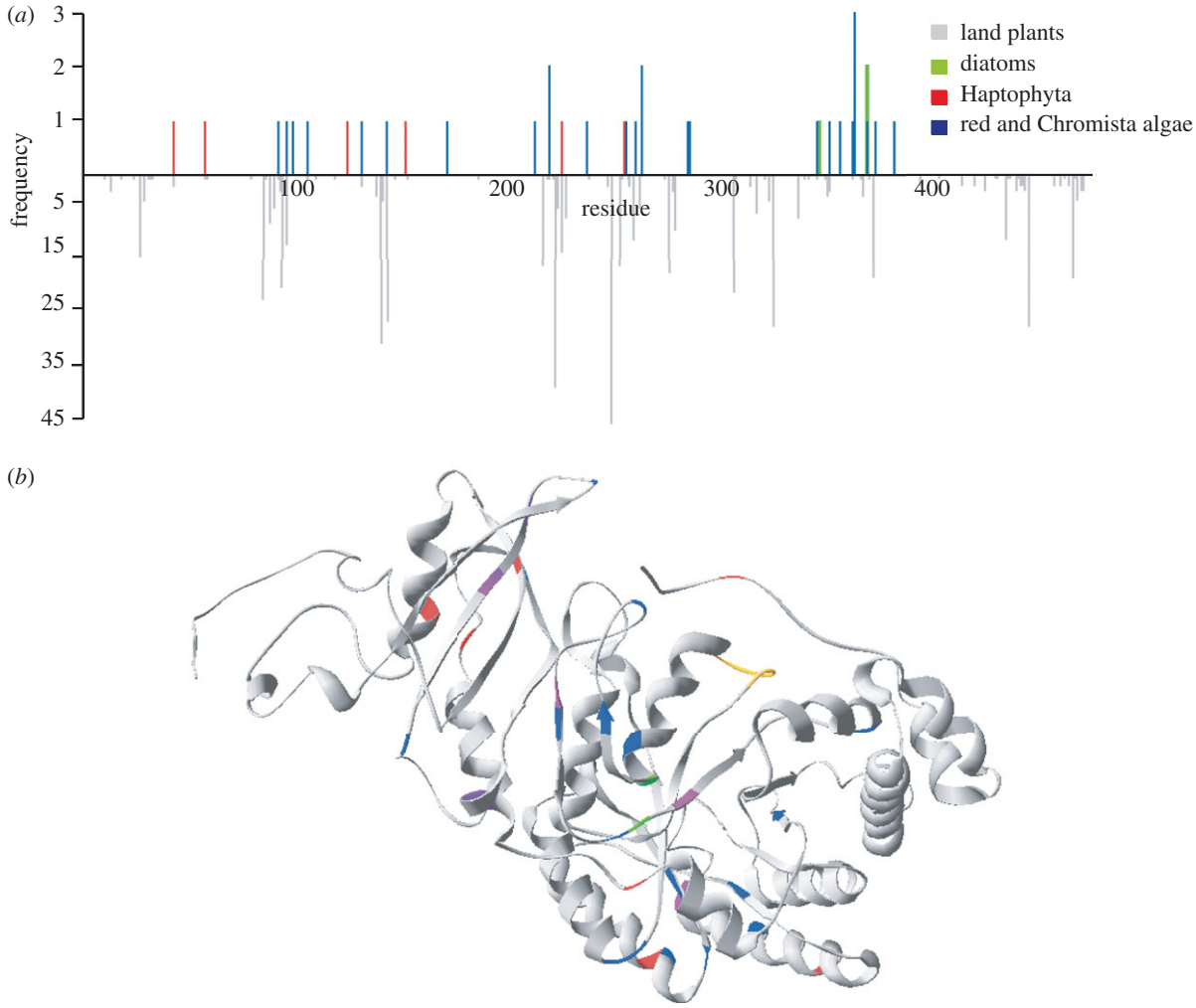


Figure 2. Location of codons identified under positive selection aligned to *Spinacia oleracea* amino acid sequence, with posterior probabilities greater than 95% detected with Bayes empirical Bayes methods [36]. (*a*) Frequency shows the number of times when a particular site was detected as being under positive selection in trees; figure 1 (blue), Haptophyta (electronic supplementary material, figure S1, red) and diatoms (electronic supplementary material, figure S2, green). This is compared with results found in land plants [49] shown in grey. (*b*) Three-dimensional tertiary structure of *S. oleracea* with algal positive selection sites highlighted (same colours as (*a*)). Loop 6 is denoted in yellow.

[55,56]. By analogy, therefore, positive selection in Rubiscos of the Haptophyta and diatoms is likely to have occurred also in response to emergence of CCMs as both Haptophyta and diatoms are thought to possess them [57]. We can infer that, like in the $C_4$ land plants, positive selection indicates a lowering of $CO_2$ affinity and an increase in $V_c$ in these algae after induction of CCMs. Indeed, modern day Haptophyta and diatoms do display lower specificities ($\Omega$) and $CO_2$ affinity (i.e. higher $K_c$) than their red algal counterparts (their ancestral endosymbionts; figure 1*b*). So, although atmospheric $CO_2$ acts as the ultimate driver of Rubisco change, the mechanistic driver is the physiological innovation in boosting intracellular carbon (the CCMs) in response to declining $CO_2$.

The term CCM encompasses a wide variety of poorly understood mechanisms that concentrate carbon to different degrees. While the pattern of positive selection in *rbcL* hints at a link between the presence of CCMs and positive selection within the Haptophyta and diatoms, there is insufficient knowledge to explain whether a simple relationship exists between gradational improvements in carbon concentration and positive selection in *rbcL*.

It does appear that the presence of positive selection certainly indicates elevated internal carbon. We have already outlined that positive selection in Haptophtya and diatoms suggest an increase in intracellular carbon concentrations above the Rhodophyta but many members of the Rhodophyta are also thought to possess CCMs [6], though their relative abilities to concentrate carbon are poorly documented [6]. Like the presence of positive selection that separates the Rhodophyta and Chromista (Haptophyta and diatoms), positive selection also occurs at the divergence of Rhodophyta and Cyanidiales, and again correlates with an increase in $K_c$ (figure 1*b*). This suggests that positive selection of *rbcL* is linked to a gradual reduced affinity of Rubisco for $CO_2$, most likely driven by development of a better functioning CCM.

On the other hand, absence of positive selection does not necessarily indicate a lack of CCMs. No positive selection is detected leading to the Phaeophyta (the branch leading to *Pylaiella littoralis* in figure 1).

Table 1. PAML results for models M0 and M1a/M2a.

| | | M0 | | | | | | | | | M1a/M2a | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| group | tree | $n_F$ | $n_G$ | $n_S$ | $n_N$ | BL | $\kappa$ | $d_S$ | $d_N$ | $d_N/d_S$ | $\chi^2$ | $p$-value |
| Rhodophyta | electronic supplementary material, figure S4*a* | 31 | 37 | 50 | 1200 | 21.748 | 2.022 | 37.483 | 0.749 | 0.020 | 0 | 1 |
| Phaeophyceae | electronic supplementary material, figure S4*b* | 8 | 29 | 45 | 1362 | 2.732 | 3.170 | 3.905 | 0.179 | 0.046 | 0.01 | 0.99 |
| Chrysophyceae and Synuophyceae | electronic supplementary material, figure S4*c* | 4 | 5 | 33 | 930 | 8.950 | 0.968 | 14.821 | 0.449 | 0.030 | 0 | 1 |
| Cryptophyta | electronic supplementary material, figure S4*d* | 4 | 8 | 28 | 972 | 9.227 | 1.778 | 0.312 | 13.086 | 0.024 | 0 | 1 |
| Haptophyta | electronic supplementary material, figure S1 | 10 | 21 | 32 | 1377 | 8.501 | 1.377 | 8.723 | 0.437 | 0.053 | 0 | 1 |
| diatoms | electronic supplementary material, figure S2 | 18 | 23 | 30 | 900 | 5.617 | 1.380 | 8.178 | 0.444 | 0.054 | 0 | 1 |
| red and brown algae | figure 1 | 28 | 39 | 40 | 1206 | 19.826 | 1.108 | 29.800 | 1.084 | 0.036 | 0 | 1 |

Using M0 model in PAML: $n_F$, number of families; $n_G$, number of genera; $n_S$, number of sequences; $n_N$, number of nucleotides; BL, branch length; $\kappa$, transition/transversion rate; $d_S$, synonymous substitutions; $d_N$, non-synonymous substitutions; $d_N/d_S$, average $d_N/d_S$ across tree (using 1 class for $d_N/d_S$). Comparing M1a (nearly neutral) with M2a (positive selection) using likelihood-ratio test (LRT), where $\chi^2 = 2 \times$ (LRTM2a – LRTM1a) with degrees of freedom = 2 to calculate probability ($p$-value) for the LRT value being as high as is seen by chance.

While there is no information on the carbon physiology of *P. littoralis*, Phaeophyta studied to date do appear to concentrate carbon, but to lower levels than their diatom counterparts. Furthermore, some orders are known to lack pyrenoids (Dicytotales, Sphacelariales, Laminariales and Fucales) although they do appear to be able to use $HCO_3^-$ [5]. Even though we find no positive selection in groups which are established as lacking CCMs (table 1 and electronic supplementary material, figure S3; the Rhodophyta genera *Batrachospermum*, *Caloglossa*, *Membranoptera*, *Nitophyllum*, *Phycodrys* and *Ptilota* [58–60], and the Chrysophyceae and Synurophyceae [60,61]), the test used for this particular analysis is very conservative. This test averages positive selection across the group and was unable to detect positive selection, even within the Haptophyta and diatoms, in which positive selection is established when individual branches were tested.

### (d) *Rubisco adaptation correlates with declining carbon dioxide*

The timing of events of positive selection in Rubisco further corroborates its relationship with atmospheric $CO_2$. A comparison of the occurrence of positive selection within Haptophyta and diatoms during the Phanerozoic against the most recent compilation of proxy-derived reconstructions and geochemical models of $CO_2$ and $O_2$ show that positive selection always corresponds with falling $CO_2$ (inset, figure 3). This correlation of Rubisco adaptation events with declining $CO_2$ demonstrates the potential for use of adaptation of Rubisco as an indicator of past atmospheric $CO_2$.

Our pattern of adaptation prior to radiation, which characterizes the history of algal Rubisco, is suggestive that adaptive change at the physiological and molecular level was associated with the enhancement of photosynthetic performance, and that this increased efficiency could be the foundation for the successful subsequent proliferation of the different algal groups in novel ecological niches. Adaptation in the Haptophyta between 375 and 285 Ga occurs over the lowest limit of $CO_2$ of the Carboniferous and coincides with estimation of the emergence of calcification and switch to full autotrophy in this group [24]. Diatoms have carbon acquisition physiology that would suggest a CCM [62,63] and it has even been controversially suggested that some possess a $C_4$ mechanism [57–60]. The three separate adaptation events within diatoms occur during falling $CO_2$ of the Mesozoic and Cenozoic, which coincide with their dominance and diversification, in particular the Thalassiosirales, between 100 Ma and 30 Ma [61].

### (e) *Extending the carbon dioxide record into the Proterozoic*

We can apply the Phanerozoic relationship between declining $CO_2$ and positive selection to investigate possible $CO_2$ change using the deeper branches in the tree of the whole red/Chromista algal lineage (figure 3). It is challenging to constrain timing of adaptation in the Proterozoic owing to long branches and the lack of a fossil record. Nonetheless, we find positive selection on the earliest divergence of the Cyanidiales (1.56–1.14 Ga). There is a cluster of signals of positive selection in a number of branches immediately after the event of secondary endosymbiosis, the origin of the Chromista

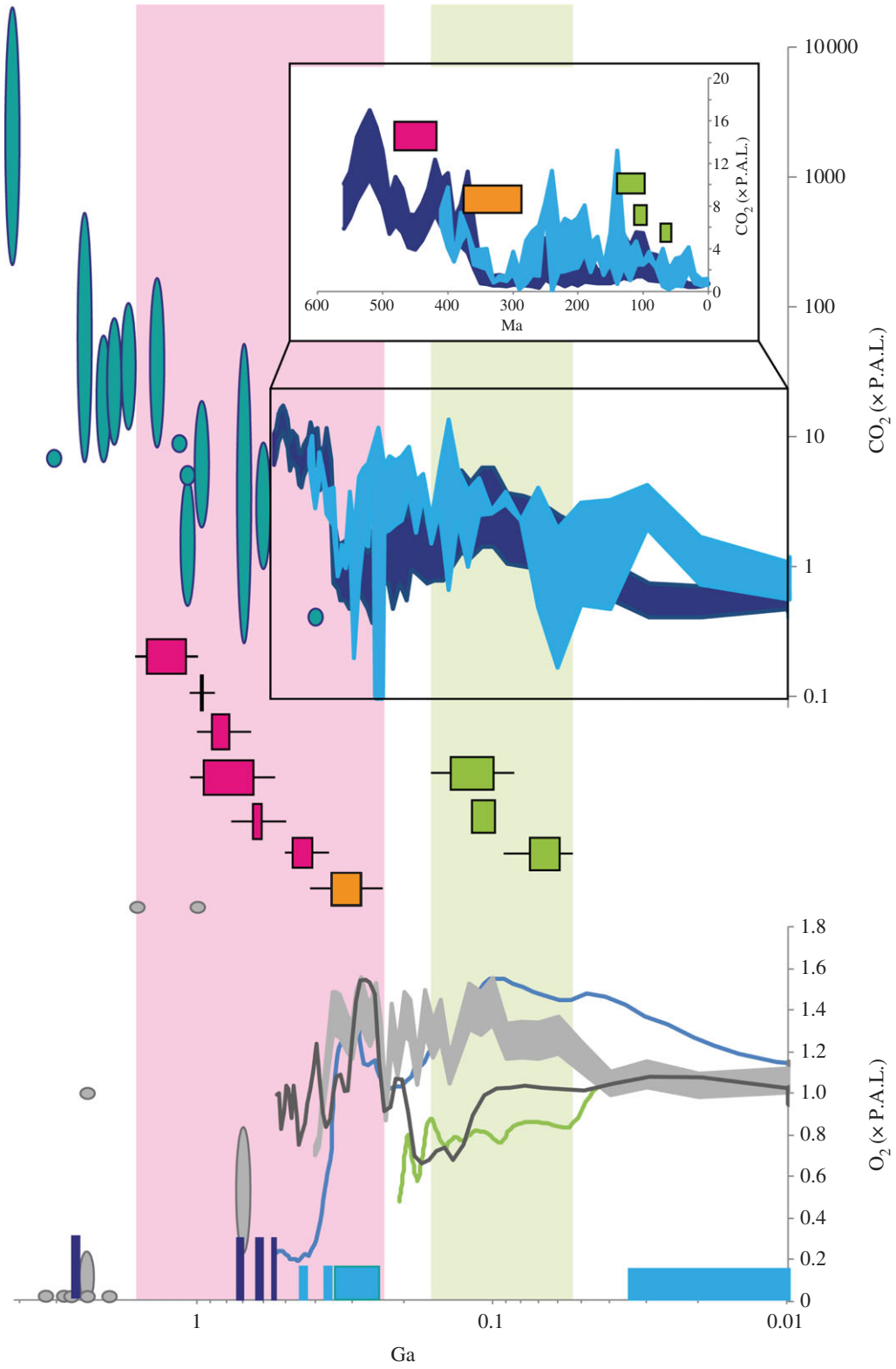


Figure 3. Timing of adaptation of Rubisco in response to changing atmospheric $CO_2$ and $O_2$. Timing of positive selection in figure 1 (magenta) and within Haptophyta (electronic supplementary material, figure S1, orange) and diatoms (electronic supplementary material, figure S2, green), where boxed area denotes time between age of nodes and line encompasses 95% probability density of ages. This is compared with proxy-derived Proterozoic ranges of atmospheric $CO_2$ (turquoise ovals [38–41]) and $O_2$ (grey ovals [1]). A continuous record of Phanerozoic $CO_2$ was reconstructed from proxies [43,44] binned into 10 Ma with 1 s.d. (light blue) and overlaid over modelled data (dark blue [42]). Phanerozoic $O_2$ was reconstructed from proxies showing 1 s.d. (shaded grey) and modelled data (dark grey line [45], blue line [47] and green line [46]). Dark blue blocks denote possible snowball Earths, and light blue blocks denote periods of glaciations. Inset shows Phanerozoic data on a linear scale, where coloured bars represent the average age of nodes.

algae, approximately 1.1 Ga. The environmental trigger for this symbiotic relationship between host and chloroplast is unknown but this relationship confirms that Rubisco evolves in response to changing intracellular conditions, plus there are certain advantages to an endosymbiotic habitat in a low $CO_2$ environment.

The continued positive selection, on two branches subsequent to those immediately after the secondary endosymbiosis, points to sustained declining $CO_2$ (0.93–0.52 Ga). We present a continuous record suggesting decreasing $CO_2$ encompassing periods when there are currently few or no constraints. Because advances in evolutionary innovation generally occurred at times of major environmental or geochemical change, the trend of declining $CO_2$ during the Proterozoic suggests this billion years of time may not have been so 'boring'. We suggest instead a backdrop of environmental change associated with the expansion of the eukaryotes during the very late Mesoproterozoic and Neoproterozoic [22]. Further, such a decrease in $CO_2$ likely played a role in triggering the subsequent Snowball Earth events and at least contributed to dictating Proterozoic climates [64].

## 4. CONCLUSION

The wide adoption of photosynthesis by life, resulting in reduced atmospheric $CO_2$–$O_2$ ratios, created increased evolutionary pressure on Rubisco which has evolved its kinetic properties in response to declining atmospheric $CO_2$ [8]. Determining the timing of adaptation of the $CO_2$–$O_2$ sensitive enzyme, Rubisco, presents a novel approach in understanding the biological response to changing atmosphere, and the periods of emergence of CCMs. It supports the current estimation of the periods of decreasing $CO_2$ beyond the Phanerozoic and provides important additional constraints to the current scarcity of proxies available during these time periods, uniquely delivering a continuous record extending 1.5 Ga. The future challenge is to further document the physiological innovations responsible for driving the Rubisco change, and to discover other genetic signatures which can open a window on past major environmental change, such as the oxygenation of the planet.

We thank the Royal Society for hosting and funding, and David Beerling for the invitation to the meeting at the Kavli Royal Society International Centre from which this paper and the Theme Issue arose. This research was supported by an ERC Starting grant (SP2-GA-2008-200915) and the Leverhulme Trust to R.E.M.R. and NERC grant (NE/H007741/1) to D.A.F. J.N.Y. acknowledges financial support through the Clarendon Scholarship, Oxford. R.E.M.R. conceived the project and directed this work. J.N.Y. performed the bulk of the sequencing, and data analysis. M.V.K. helped with analysis of positive selection, D.A.F. conducted phylogenetic analyses and dating of the nodes. J.N.Y. led the writing and all authors contributed to discussions and the writing. *RbcL* sequences generated for this research have been deposited with the NCBI GenBank under accession codes HQ656822–HQ656838. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.E.M.R. (rosalind.rickaby@earth.ox.ac.uk).

## REFERENCES

1 Canfield, D. E. 2005 The early history of atmospheric oxygen: homage to Robert M. Garrels. *Annu. Rev. Earth Planet Sci.* **33**, 1–36. (doi:10.1146/annurev.earth.33.092203.122711)
2 Johnston, D. T., Wolfe-Simon, F., Pearson, A. & Knoll, A. H. 2009 Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc. Natl Acad. Sci. USA* **106**, 16 925–16 929. (doi:10.1073/pnas.0909248106)
3 Laing, W. A., Ogren, W. L. & Hageman, R. H. 1974 Regulation of soybean net photosynthetic $CO_2$ fixation by the interaction of $CO_2$, $O_2$, and ribulose 1,5-diphosphate carboxylase. *Plant Physiol.* **54**, 678–685. (doi:10.1104/pp.110.159855)
4 Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. 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. (doi:10.1073/pnas.0600605103)
5 Badger, M. R. & Andrews, T. J. 1987 Co-evolution of Rubisco and $CO_2$ concentrating mechanisms. In *Progress in photosynthesis research*, vol. III (ed. J. Biggins), pp. 601–609. Dordrecht, The Netherlands: Martinus Nijhoff Publishers.
6 Badger, M. R. T., Andrews, J., Whitney, S. M., Ludwig, M., Yellowlees, D. C., Leggat, W. & Price, G. D. 1998 The diversity and coevolution of Rubisco, plastids, pyrenoids, and chloroplast-based $CO_2$-concentrating mechanisms in algae. *Can. J. Bot.* **76**, 1052–1071. (doi:10.1139/b98-074)
7 Badger, M. R., Hanson, D. & Price, G. D. 2002 Evolution and diversity of $CO_2$ concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* **29**, 161–173. (doi:10.1071/PP01213)
8 Tortell, P. D. 2000 Evolutionary and ecological perspectives on carbon acquisition in phytoplankton. *Limnol. Oceanogr.* **45**, 744. (doi:10.4319/lo.2000.45.3.0744)
9 Spreitzer, R. J. & Salvucci, M. E. 2002 Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* **53**, 449–475. (doi:10.1146/annurev.arplant.53.100301.135233)
10 Cavalier-Smith, T. 1999 Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. *J. Eukaryot. Microbiol.* **46**, 347–366. (doi:10.1111/j.1550-7408.1999.tb04614.x)
11 Cavalier-Smith, T. 1981 Eukaryote kingdoms: seven or nine? *BioSystems* **14**, 461–481. (doi:10.1016/0303-2647(81)90050-2)
12 Keeling, P. J. 2009 Chromalveolates and the evolution of plastids by secondary endosymbiosis. *J. Eukaryot. Microbiol.* **56**, 1–8. (doi:10.1111/j.1550-7408.2008.00371.x)
13 Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. 2002 The single, ancient origin of chromist plastids. *Proc. Natl Acad. Sci. USA* **99**, 15 507–15 512. (doi:10.1073/pnas.242379899)
14 Sims, P. A., Mann, D. G. & Medlin, L. K. 2006 Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* **45**, 361–402. (doi:10.2216/05-22.1)
15 Medlin, L. K., Sáez, A. G. & Young, J. R. 2008 A molecular clock for coccolithophores and implications for selectivity of phytoplankton extinctions across the K/T boundary. *Mar. Micropaleontol.* **67**, 69–86. (doi:10.1016/j.marmicro.2007.08.007)
16 Yang, Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
17 Richlen, M. & Barber, P. H. 2005 A technique for the rapid extraction of microalgal DNA from single live and preserved cells. *Mol. Ecol. Notes* **4**, 688–691. (doi:10.1111/j.1471-8286.2005.01032.x)
18 Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. 1997 The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882. (doi:10.1093/nar/25.24.4876)

19 Filatov, D. A. 2009 Processing population genetic analysis of multigenic datasets with ProSeq3 software. *Bioinformatics* **25**, 3183–3189. (doi:10.1093/bioinformatics/btp572)
20 Haslam, R. P., Keys, A. J., Andralojc, P. J., Madgwick, P. J., Andersson, I., Grimsrud, A., Eilertsen, H. C. & Parry, M. A. J. 2005 Specificity of diatom Rubisco. In *Plant responses to air pollution and global change* (eds K. Omasa, I. Nouchi & L. J. De Kok), pp. 157–164. Tokyo, Japan: Springer.
21 Webster, R. J. 2009 The effects of light and $CO_2$ on photosynthesis in *Emiliania Huxleyi*. PhD thesis, University of Essex, UK.
22 Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. 2004 A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* **21**, 809–818. (doi:10.1093/molbev/msh075)
23 Drummond, A. & Rambaut, A. 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214. (doi:10.1186/1471-2148-7-21)
24 Liu, H., Aris-Brosou, S., Probert, I. & de Vargas, C. 2010 A time line of the environmental genetics of the haptophytes. *Mol. Biol. Evol.* **27**, 161–176. (doi:10.1093/molbev/msp222)
25 Saunders, G. W. & Hommersand, M. H. 2004 Assessing red algal supraordinal diversity and taxonomy in the context of contemporary systematic data. *Am. J. Bot.* **91**, 1494–1507. (doi:10.3732/ajb.91.10.1494)
26 Rampen, S. W., Schouten, S., Hopmans, E. C., Abbas, B., Noordeloos, A. A. M., van Bleijswijk, J. D. L., Geenevasen, J. A. J. & Sinninghe Damsté, J. S. 2009 Diatoms as a source for 4-desmethyl-23,24-dimethyl steroids in sediments and petroleum. *Geochim. Cosmochim. Acta* **73**, 377–387. (doi:10.1016/j.gca.2008.10.024)
27 Rodríguez, F., Oliver, J. L., Marín, A. & Medina, J. R. 1990 The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**, 485–501. (doi:10.1186/1471-2148-7-21)
28 Yule, G. U. 1925 A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Phil. Trans. R. Soc. Lond. B* **213**, 21–87. (doi:10.1098/rstb.1925.0002)
29 Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. 2006 Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88. (doi:10.1371/journal.pbio.0040088)
30 Rambaut, A. & Drummond, A. J. 2009 Tracer v. 1.5. See http://beast.bio.ed.ac.uk/Tracer (accessed 04 June 2010).
31 Rambaut, A. 2009 FigTree. See http://tree.bio.ed.ac.uk/software/figtree/ (accessed 04 June 2010).
32 Zhang, J., Nielsen, R. & Yang, Z. 2005 Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **22**, 2472–2479. (doi:10.1093/molbev/msi237)
33 Ronquist, F. & Huelsenbeck, J. P. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)
34 Uitenbroek, D. G. 1997 SISA-Binomial. See http://www.quantitativeskills.com/sisa/distributions/binomial.htm (accessed 20 June 2010).
35 Wong, W. S. W., Yang, Z., Goldman, N. & Nielsen, R. 2004 Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**, 1041–1051. (doi:10.1534/genetics.104.031153)
36 Yang, Z., Wong, W. S. & Nielsen, R. 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118. (doi:10.1093/molbev/msi097)
37 Guex, N. & Peitsch, M. C. 1997 SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723. (doi:10.1002/elps.1150181505)
38 Kasting, J. F. 1993 Earth's early atmosphere. *Science* **5097**, 920–926. (doi:10.1126/science.11536547)
39 Sheldon, N. D. 2006 Precambrian paleosols and atmospheric $CO_2$ levels. *Precambrian Res.* **147**, 148–155. (doi:10.1016/j.precamres.2006.02.004)
40 Kaufman, A. J. & Xiao, S. 2003 High $CO_2$ levels in the Proterozoic atmosphere estimated from analyses of individual microfossils. *Nature* **425**, 279–282. (doi:10.1038/nature01902)
41 Rosing, M. T., Bird, D. K., Sleep, N. H. & Bjerrum, C. J. 2010 No climate paradox under the faint early Sun. *Nature* **464**, 744–747. (doi:10.1038/nature08955)
42 Berner, R. A. 2008 Addendum to 'Inclusion of the weathering of volcanic rocks in the GEOCARBSULF model': (R. A. Berner, 2006, **306,** pp. 295–302). *Am. J. Sci.* **308**, 100–103. (doi:10.2475/01.2008.04)
43 Royer, D. L., Berner, R. A. & Park, J. 2007 Climate sensitivity constrained by $CO_2$ concentrations over the past 420 million years. *Nature* **446**, 530–532. (doi:10.1038/nature05699)
44 Breecker, D. O., Sharp, Z. D. & McFadden, L. D. 2010 Atmospheric $CO_2$ concentrations during ancient greenhouse climates were similar to those predicted for A.D. 2100. *Proc. Natl Acad. Sci. USA* **107**, 576–580. (doi:10.1073/pnas.0902323106)
45 Berner, R. A. 2009 Phanerozoic atmospheric oxygen: new results using the GEOCARBSULF model. *Am. J. Sci.* **309**, 603–606. (doi:10.2475/07.2009.03)
46 Falkowski, P. G., Katz, M. E., Milligan, A. J., Fennel, K., Cramer, B. S., Aubry, M. P., Berner, R. A., Novacek, M. J. & Zapol, W. M. 2005 The rise of oxygen over the past 205 million years and the evolution of large placental mammals. *Science* **309**, 2202–2204. (doi:10.1126/science.1116047)
47 Bergman, N. M., Lenton, T. M. & Watson, A. J. 2004 COPSE: a new model of biogeochemical cycling over Phanerozoic time. *Am. J. Sci.* **304**, 397–437. (doi:10.2475/ajs.304.5.397)
48 Glasspool, I. J. & Scott, A. C. 2010 Phanerozoic concentrations of atmospheric oxygen reconstructed from sedimentary charcoal. *Nat. Geosci.* **3**, 627–630. (doi:10.1038/ngeo923)
49 Kapralov, M. & Filatov, D. 2007 Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* **7**, 73. (doi:10.1186/1471-2148-7-73)
50 Galmes, J., Flexas, J., Keys, A. J., Cifre, J., Mitchell, R. A. C., Madgwick, P. J., Haslam, R. P., Medrano, H. P. & Parry, M. A. J. 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. (doi:10.1111/j.1365-3040.2005.01300.x)
51 Raven, J. & Larkum, A. 2007 Are there ecological implications for the proposed energetic restrictions on photosynthetic oxygen evolution at high oxygen concentrations? *Photosynth. Res.* **94**, 31–42. (doi:10.1007/s11120-007-9211-z)
52 Hopkinson, B. M., Dupont, C. L., Allen, A. E. & Morel, F. M. M. 2011 Efficiency of the $CO_2$-concentrating mechanism of diatoms. *Proc. Natl Acad. Sci. USA* **108**, 3830–3837. (doi:10.1073/pnas.1018062108)
53 Badger, M. R., Bassett, M. & Comins, H. N. 1985 A model for $HCO_3^-$ accumulation and photosynthesis in the cyanobacterium *Synechococcus* sp. Theoretical predictions and experimental observations. *Plant Physiol.* **77**, 465–471. (doi:10.1104/pp.77.2.465)

54 Savir, Y., Noor, E., Milo, R. & Tlusty, T. 2010 Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc. Natl Acad. Sci. USA* **107**, 3475–3480. (doi:10.1073/pnas.0911663107)
55 Christin, P., Salamin, N., Muasya, A. M., Roalson, E. H., Russier, F. & Besnard, G. 2008 Evolutionary switch and genetic convergence on *rbcL* following the evolution of C4 photosynthesis. *Mol. Biol. Evol.* **25**, 2361–2368. (doi:10.1093/molbev/msn178)
56 Kapralov, M. V., Kubien, D. S., Andersson, I. & Filatov, D. A. 2011 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.* **28**, 1491–1503. (doi:10.1093/molbev/msq335)
57 Reinfelder, J. R. 2011 Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Annu. Rev. Mar. Sci.* **3**, 291–315. (doi:10.1146/annurev-marine-120709-142720)
58 Reinfelder, J. R., Kraepiel, A. M. & Morel, F. M. 2000 Unicellular C4 photosynthesis in a marine diatom. *Nature* **407**, 996–999. (doi:10.1038/35039612)
59 Roberts, K., Granum, E., Leegood, R. C. & Raven, J. A. 2007 Carbon acquisition by diatoms. *Photosynth. Res.* **93**, 79–88. (doi:10.1007/s11120-007-9172-2)
60 Raven, J. 2010 Inorganic carbon acquisition by eukaryotic algae: four current questions. *Photosynth. Res.* **106**, 123–134. (doi:10.1007/s11120-010-9563-7)
61 Armbrust, E. V. 2009 The life of diatoms in the world's oceans. *Nature* **459**, 185–192. (doi:10.1038/nature08057)
62 Burkhardt, S., Amoroso, G., Riebesell, U. & Sültemeyer, D. 2001 $CO_2$ and $HCO_3^-$ uptake in marine diatoms acclimated to different $CO_2$ concentrations. *Limnol. Oceanogr.* **46**, 1378–1391. (doi:10.4319/lo.2001.46.6.1378)
63 Rost, B., Riebesell, U., Burkhardt, S. & Sültemeyer, D. 2003 Carbon acquisition of bloom-forming marine phytoplankton. *Limnol. Oceanogr.* **48**, 55–67. (doi:10.4319/lo.2003.48.1.0055)
64 Hoffman, P. F., Kaufman, A. J., Halverson, G. P. & Schrag, D. P. 1998 A Neoproterozoic snowball earth. *Science* **281**, 1342–1346. (doi:10.1126/science.281.5381.1342)