Molecular foundations of Precambrian uniformitarianism

Mateusz Kędzior[1], Amanda K. Garcia[1], Meng Li[2], Arnaud Taton[3], Zachary R. Adam[4], Jodi N. Young[3] and Betül Kaçar[1,4*]

[1]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ
[2]School of Oceanography, University of Washington, Seattle, WA
[3]Division of Biological Sciences, University of California San Diego, La Jolla, CA
[4]Department of Planetary Sciences, Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ
*Corresponding author: betul@arizona.edu

Keywords: RuBisCO, carbon fixation, carbon isotope record, molecular paleobiology, paleogenetics, ancestral sequence reconstruction

Abstract

Uniformitarian assumptions underlie the oldest evidence for living organisms on Earth, the distinct isotope fractionation between inorganic and organic carbon. Aside from a handful of compelling deviations, the $^{13}C/^{12}C$ isotopic mean of preserved organic carbon ($\delta^{13}C_{org}$) has remained remarkably unchanged through time. RuBisCO is the principal carboxylase/oxygenase biomolecular component that is thought to primarily account for the generation of these distinct carbon isotopic signals. However, it is difficult to reconcile a mostly unchanging mean $\delta^{13}C_{org}$ with several known factors that can affect the isotope fractionation of RuBisCO, such as atmospheric composition and the amino acid composition of the enzyme itself, which have each changed markedly over Earth history. Here we report the resurrection and genetic incorporation of a Precambrian-age, Form IB RuBisCO in a modern cyanobacterial host. The isotopic composition of biomass relative to CO_2 (ε_p) in ancestral and control strains were much greater when grown under Precambrian CO_2 concentrations compared to modern ambient levels, but displaying values within a nominal envelope of modern-day RuBisCO IB enzyme variants. We infer that these isotopic differences derive indirectly from the decreased fitness of the AncIB strain, which includes diminished growth capacity and total cell RuBisCO activity. We argue that to answer the greatest questions of deep-time paleobiology, ancient biogeochemical signals should be reproduced in the laboratory through the synthesis of the geologic record with experimentally-derived constraints on underlying ancient molecular biology.

Significance Statement

The earliest geochemical indicators of microbes, and the enzymes that powered them, extend back almost 3.8 billion years on our planet. Paleobiologists often attempt to understand these indicators by assuming that the behaviors of modern microbes and enzymes are consistent (uniform) with those of their predecessors. This assumption seems uncomfortably at odds with the great variability of Earth's

environment and its highly adaptive microbes. Here we examine whether a uniformitarian assumption for an enzyme thought to generate these indicators, RuBisCO, can be corroborated by independently studying the history of changes recorded within RuBisCO's genetic sequences. We outline a new approach to paleobiology that informatively links molecule-level evolutionary changes with planet-level geochemical conditions in Earth's deep past.

Introduction

The history of life on Earth may be broadly subdivided into two, mutually exclusive macroevolutionary phases, the Phanerozoic and the Precambrian (1). The Phanerozoic (~542 million years ago to present) may be characterized by physiological and anatomical innovations and their resultant effects on ecosystem expansion, trophic tiering, and sociality (2-4). Hard- and soft-anatomical preservation provides a rich template for reconstructing Phanerozoic adaptive trends, correlating them with geographical and climatological changes (5, 6), and for testing observed diversity trends against possible systematic effects of preservation bias (7-9). By contrast, the Precambrian (the ~4 billion years preceding the Phanerozoic) is primarily characterized by genetic and metabolic biomolecular innovations, traded amongst microscopic organisms of uncertain phylogenetic assignment (10, 11). The Precambrian record of evolutionary change appears to be cryptic and may have been comparatively static. This may be attributable to macroevolutionary dynamics that were distinctly non-Phanerozoic, or it may merely indicate a lack of direct paleontological and geological evidence of the specific timing and extent of intermediate biomolecular adaptive steps (1).

Comparative analyses of extant organisms have traditionally been the most informative means of interpreting the scant direct evidence of Precambrian life, but such analyses inevitably face pitfalls. A reasonable null hypothesis is that evolution is largely a uniformitarian process, such that rates or tempos may change but the underlying processes or modes of evolutionary change (as established through observations of the more comprehensive Phanerozoic record) are likely invariant in deep time (12). Uniformitarian assumptions of ancient biology inferred from extant or Phanerozoic phenotypes are often employed to make sense of the Precambrian record, including body plan function (13, 14), cladistic assignment (15-18), and isotope biosignature traces (19-23). However, more recent studies complicate the use of uniformitarian assumptions, namely indicating that modes of biological variation can actually vary through time (12). A major crux of the problem is that even the simplest modern organisms, as well as the macromolecules that compose them, differ from their ancient predecessors in having been shaped by the cumulative effects of billions of years of Earth-life co-evolution and ecosystem upheaval. For this reason, Precambrian functional or sequence comparisons may be of limited utility at organismal or biomolecular levels of adaptation, undercutting interpretations made possible by uniformitarian assumptions. Novel experimental approaches may help to distinguish inferred paleobiological phenotypes from characteristically modern adaptive overprints.

The interpretation of the Precambrian carbon isotope record, comprising the oldest signatures of life on Earth, may be aided by novel experimental constraints on ancient phenotypes. Interpretations of this record is conventionally subject to uniformitarian assumptions regarding ancient biogeochemistry (19, 24). The $^{13}C/^{12}C$ isotopic mean of preserved organic carbon ($\delta^{13}C_{org} \approx -25‰$) has remained notably static over geologic time (19, 21, 24), and is leveraged as a general signature of ancient biological activity (25-

27). Given its role in the Calvin-Benson-Bassham cycle—likely the predominant mode of carbon fixation for much of Earth history (19)—the majority of contextual information used to assess how carbon isotope biosignatures might have been generated over Earth's deep history comes from studies of the modern enzyme RuBisCO (Ribulose 1,5-Bisphosphate (RuBP) Carboxylase/Oxygenase, EC 4.1.1.39). RuBisCO catalyzes the uptake of inorganic CO_2 from the environment and facilitates CO_2 reduction and incorporation into organic biomass. The $^{13}C/^{12}C$ isotopic fractionation of modern Form I RuBisCO variants in photosynthetic organisms consistently measures ~-25‰ (28-31), approximately the same isotopic difference observed between inorganic and organic carbon in the Precambrian rock record. The carbon isotope discrimination of RuBisCO has therefore been presumed to have remained constant over the history of life. Recent data, however, demonstrate that RuBisCO can produce significantly different carbon isotope signatures within organic matter in response to external factors, such as levels of atmospheric CO_2 and/or O_2 (32-35) or cellular carbon concentrating mechanisms, which affect the catalytic efficiency of RuBisCO (36-38). Internal factors may also affect fractionation, such as single-point mutations that can alter the interaction between RuBisCO and CO_2 (39). Given the sensitivity of RuBisCO to both external and internal variables, it seems unlikely that ancestral forms under ancient environmental conditions generated the same isotope fractionation signal as descendent homologs in modern organisms. An experimental assessment that combines the phylogenetic history of RuBisCO with the study of intra- and extracellular conditions may provide a more insightful basis for comparing extant and Precambrian carbon isotope fractionation patterns.

The Form I clade of the RuBisCO phylogeny (and its macroevolutionary tractability afforded through green plant, algal and cyanobacterial fossil lineages) makes it an exemplary paleomolecular system for assessing uniformitarian assumptions applied to Precambrian biosignatures. Here, we establish an experimental system for the reconstruction of ancestral biomolecules with which to interpret evidence of Precambrian biological activity. Specifically, we report the resurrection and genetic incorporation of a phylogenetically reconstructed, ancient Form IB RuBisCO variant in a modern strain of cyanobacteria, *Synechococcus elongatus* PCC 7942 (thereafter *S. elongatus*) (40, 41). We compared expression and activity levels of RuBisCO variants and the resulting changes in the growth of *S. elongatus* and isotope fractionation under ambient air as well as a CO_2 concentration that reflects Precambrian conditions.

Results

Construction of a *S. elongatus* strain harboring ancestral RuBisCO. To experimentally investigate the generation of carbon isotope biosignatures in deep time, we designed a paleomolecular system to engineer computationally inferred, ancestral RuBisCO enzymes in extant cyanobacteria. We previously reconstructed a comprehensive phylogenetic history of RuBisCO and inferred maximum-likelihood ancestral RuBisCO large-subunit (RbcL) protein sequences (42) (Fig. 1A). For this study, we selected the ancestor of the Form IB RuBisCO clade (cyanobacteria, green algae, and land plants) for laboratory resurrection, designated "AncIB." Chlorophyte and land plant RuBisCO homologs are nested among cyanobacterial sequences within the Form IB clade. The Form IB topology therefore recapitulates a primary plastid endosymbiotic history from cyanobacterial to Chlorophyte ancestors (43-45) and constrains the minimum age of ancestral Form IB to older than the Archaeplastida. As a conservative estimate, AncIB is thus likely older than ~1 Ga (the age of the oldest well-characterized, crown-group red and green algal fossils (16, 46)) and younger than maximum age estimates of cyanobacteria (~3 Ga, as

129 constrained by oxidized sediments potentially indicating the early presence of oxygenic photosynthesis
130 (47, 48)).
131
132 The ancestral AncIB and *S. elongatus* native RbcL proteins differ at 37 sites and share 92% amino acid
133 identity (Fig. 2). This site variation is evenly spread across the length of the protein, except for a highly
134 conserved region between approximately site 170 to 285 (site numbering here and hereafter based on
135 aligned WT *S. elongatus* RbcL; Fig. 2A) that constitutes a portion of the catalytic C-terminal domain and is
136 proximal to the L-L interface and active site. Critical residues for carboxylase activity, including the Lys-
137 198 site that binds CO_2, are conserved in AncIB. Homology modeling of AncIB using the *S. elongatus* RbcL
138 template (PDB: 1RBL (49)) indicates high structural conservation without predicted disruption to
139 secondary structure (Fig. 2B). The nucleotide sequence for the reconstructed AncIB RbcL protein was
140 codon-optimized for *S. elongatus* and cloned within a copy of the *rbc* operon into pSyn02 (50) for insertion
141 into the *S. elongatus* chromosomal neutral site 2 (NS2). The native *rbc* operon was subsequently deleted
142 to create the AncIB strain. In addition, we generated the control strain Syn01 harboring WT *rbcL* at NS2
143 (see Materials and Methods for full genome strategy). Thus, the AncIB and Syn01 strains carry a single
144 ectopic copy of the *rbc* operon at NS2 and only differ in the coding sequence of the large RuBisCO subunit
145 (Table 1).
146
147 **Ancestral RuBisCO complements photoautotrophic growth of extant *S. elongatus*.** We cultured wild-
148 type (WT) and engineered (AncIB and Syn01) strains of *S. elongatus* in both ambient air and 2% CO_2 to
149 evaluate the physiological impact of ancestral RuBisCO under estimated Precambrian CO_2 concentrations
150 (51) (Fig. 3A). The AncIB strain was capable of photoautotrophic growth in both ambient air and 2% CO_2.
151 Maximum growth rates for all strains were relatively comparable under each atmospheric condition
152 (averaging doubling times of ~15 to 20 hours), and generally increased under 2% CO_2 relative to air ($p <$
153 0.001; Table 2). The AncIB strain exhibited a significantly diminished maximum cell density ($OD_{750} \approx 5$)
154 relative to both WT and Syn02 strains ($OD_{750} \approx 8$; $p < 0.001$). No significant difference between WT and
155 Syn01 growth rate or maximum cell density was observed under ambient air or 2% CO_2. In addition, a
156 moderate increase in the midpoint time for the AncIB strain was observed under air, indicating lag in
157 growth ($p < 0.01$). These growth parameters taken together suggest decreased fitness of the AncIB strain
158 relative to WT and Syn01.
159
160 AncIB RuBisCO protein produced a more modest impact on the oxygen evolution of *S. elongatus* compared
161 to that on growth parameters. Cell suspensions were briefly incubated in the dark and subsequently
162 exposed to saturated light in an electrode chamber to detect evolution of molecular oxygen (normalized
163 to chlorophyll *a* concentration (52)). For cells sampled from cultures grown in air, there was no statistical
164 difference detected between any of the WT, Syn01, or AncIB strains (Fig. 3B). We did find a modest but
165 significant decrease in photosynthetic activity for AncIB relative to WT for cells cultured in 2% CO_2,
166 generating ~200 and ~320 nmol $O_2 \cdot h^{-1} \cdot \mu g^{-1}$ chlorophyll *a*, respectively. However, a slower O_2 evolution
167 rate was also observed for Syn01 at 2% CO_2.
168
169 **Ancestral RuBisCO is overexpressed and less catalytically active relative to WT.** We assessed the impact
170 of ancestral RuBisCO on gene expression at both the transcript and protein levels. *S. elongatus rbcL*
171 transcript was measured by quantitative reverse-transcription PCR (RT-qPCR) and normalized to that of
172 the *secA* reference gene (53). For strains cultured in ambient air, we found that the AncIB strain produced

173 a ~29-fold increase in *rbcL* transcript relative to WT or the control strain Syn01 ($p < 0.001$; Fig. 4A). The
174 magnitude of AncIB *rbcL* overexpression was lower in 2% CO_2, with a ~4 and ~2-fold increase observed
175 relative to WT and the control strain, respectively ($p < 0.001$).
176
177 RbcL protein was quantified for all *S. elongatus* strains by immunodetection using rabbit anti-RbcL
178 antibody. We found that the amount of RbcL protein was also increased in the AncIB strain by ~3-fold (p
179 < 0.05) and ~5-fold ($p < 0.001$) relative to WT or Syn01 in air and 2% CO_2, respectively (Fig. 4B). No
180 difference in RbcL quantity was detected between WT and Syn01 strains at either atmospheric condition.
181 Finally, we confirmed assembly of the hexadecameric L_8S_8 RuBisCO complex in the AncIB strain by native
182 PAGE and detection by anti-RbcL antibody (Fig. S1).
183
184 The total carboxylase activity of *S. elongatus* harboring ancestral RuBisCO was measured from crude cell
185 lysates. Activity was assessed by an *in vitro* spectrophotometric coupled-enzyme assay that measures
186 NADH oxidation and is reported as the RuBP consumption rate normalized to total soluble protein content
187 (54). For two sets of assays using either 2.5 mM or 5 mM HCO_3^-, AncIB lysate generated less than half the
188 carboxylase activity of WT lysate ($p < 0.05$) (Fig. 4C).
189
190 ***S. elongatus* harboring ancestral RuBisCO produces greater carbon isotopic fractionation than wild-**
191 **type.** We measured the carbon isotope discrimination of *S. elongatus* strains cultured in ambient air and
192 2% CO_2 to evaluate how the fractionation behavior of ancestral RuBisCO might influence the
193 interpretation of ancient isotopic biosignatures preserved in the geologic record. The $^{13}C/^{12}C$ carbon
194 isotope composition of *S. elongatus* was measured for biomass ($\delta^{13}C_{biomass}$) as well as dissolved inorganic
195 carbon (DIC; $\delta^{13}C_{DIC}$) in the growth medium (Table S1). We calculated the carbon isotope fractionation
196 associated with photosynthetic CO_2 fixation (ε_p) following Freeman and Hayes (33) after estimating $\delta^{13}C_{CO2}$
197 from measured $\delta^{13}C_{DIC}$ (55, 56) (see Materials and Methods).
198
199 Overall, ε_p values were greater for *S. elongatus* strains cultured in 2% CO_2 compared to ambient air. At 2%
200 CO_2, ε_p ranged between 21‰ and 26‰ compared to only 8‰ and 14‰ in air. We found that *S. elongatus*
201 engineered with AncIB RbcL had an ε_p ~5‰ greater than both WT and Syn01 when cultured in air ($p <$
202 0.001) and ~2‰ to 4‰ greater than WT and Syn01 when cultured at 2% CO_2 ($p < 0.01$) (Fig. 1C), though
203 these differences in ε_p appear driven by the inorganic carbon pool composition rather than biomass (Table
204 S1). A substantially smaller increase in ε_p (~1‰; $p < 0.001$) was observed for Syn01 relative to WT under
205 ambient air. Conversely, under 2% CO_2, Syn01 ε_p was decreased by ~2‰ ($p < 0.001$) relative to WT.
206

207 Discussion

208
209 Form IB ancestral RuBisCO, when engineered into an extant strain of *S. elongatus*, decreased both the
210 organismal growth capacity and the total cell RuBisCO activity. The genetic engineering strategy for
211 insertion of the AncIB *rbcL* sequence in the cyanobacterial genome cannot solely account for these
212 physiological differences since differences between the WT and Syn01 control strains were insignificant
213 for most measured properties, or not comparable in magnitude to differences between the WT and AncIB
214 strains. Rather, the observed differences for the AncIB strain indicate that the resultant phenotype is likely
215 attributable to the functionality of the ancestral enzyme itself. The unique phenotype of the AncIB strain
216 could be a direct result of the ancestral RbcL subunit or due to impediments to the assembly and activation

of a hexadecamer RuBisCO complex containing both the ancestral RbcL and modern RbcS subunits. Another possibility is hampered integration of ancestral RbcL given a modern suite of associated proteins required for RuBisCO folding and assembly (57). Further, while overexpression of the ancestral RbcL occurred at the level of transcription and translation, the ancestral strain appears to have comparable levels of assembled hexadecamer RuBisCO, suggesting lower rates of RuBisCO assembly (or faster degradation). Even lower rates of measured total carboxylase activity suggest that the AncIB has decreased efficacy, which could be directly representing ancestral RuBisCO kinetics as well as the challenges associated with hybrid enzyme activation and activity.

Overexpression of the amount of ancestral RuBisCO shown by RT-qPCR and immunodetection assays is a common physiological response to decreased enzymatic efficacy throughout the cell (e.g., (58, 59)). However, expression compensation is insufficient to fully restore the extant WT phenotype, as indicated by the reduced fitness (i.e., decreased maximum cell density, oxygen evolution, and total carboxylase activity) of the ancestral strain harboring the ancestral RuBisCO compared to WT and Syn01.

There are few *in vitro* measurements of the kinetic isotope effect of Form IB RuBisCO in modern-day organisms, but those available range from ~22‰ to 28‰ for cyanobacteria and C_3 plants, respectively (28-31). The ~26‰ ε_p of AncIB strain biomass grown under 2% CO_2 suggests that the ancestral RuBisCO also fractionates within this range. It has been theorized that RuBisCO kinetics have adapted in response to CO_2 availability, either due to increased environmental CO_2 (36) or the emergence of CCMs (e.g., C_4 photosynthesis in plants (60)). Considering the positive relationship between enzymatic fractionation and RuBisCO's specificity to CO_2 (36), reconstructing ancient RuBisCO kinetic isotope effect could provide insights into the co-evolution of atmospheric concentrations of CO_2 and O_2 and carbon fixation strategies during the Precambrian, in particular the emergence of carbon concentrating mechanisms (CCMs) (34, 61-63). This is relevant as precise estimates of the magnitude of atmospheric CO_2 elevation during the Precambrian relative to the present, as well as the emergence and effectiveness of Precambrian CCMs, are unknown.

We did observe statistically significant differences in ε_p of the ancestral strain compared to WT *S. elongatus* under both ambient air and 2% CO_2 atmospheric conditions. However, upon inspection it appears that the diminished activity of the AncIB strain is influencing the composition of the DIC pool (both $\delta^{13}C$ and concentration) in our cultures, and it is in fact the differences in DIC composition driving the calculated differences in ε_p. Though strains were harvested at similar cell densities, small differences in cell concentrations at high densities can strongly influence carbonate chemistry of the media (64). The lower CO_2 availability in the air treatment is more sensitive to cellular influence, resulting in a larger difference in ε_p compared to the 2% CO_2 treatment. These differences in DIC are unlikely to be due to experimental setup (e.g., variations in CO_2 bubbling), as biological replicates showed similar values. Therefore, the differences in fractionation reported here are likely implicated indirectly with the less efficient AncIB ancestral enzyme. Further comparative biomolecular characterization of AncIB and WT *S. elongatus* RbcL forms is needed to determine the degree to which enzymatic inefficiencies are contributing directly to the AncIB strain phenotype.

The observed carbon isotopic fractionation values corroborate a uniformitarian assumption for applying the maximal range of extant organism-level isotope fractionation values to interpret deep time isotopic

261 biosignatures. There are, however, several potentially important contextual caveats. The most obvious is
262 that the isotopic fractionation values of all strains (WT, Syn01, and AncIB) are increased under simulated
263 Precambrian conditions with elevated CO_2. The Form IB ancestor represents predecessors that are at least
264 1 billion years old, but it is also genetically and functionally still likely to be very different from the putative
265 'root' or common ancestor of all RuBisCO variants that emerged much earlier. Reconstruction of older
266 ancestors may further expand this maximal envelope of RuBisCO-generated carbon fractionation, or it
267 may indicate that the extant maximal envelope is pervasive (and perhaps characteristic) across all
268 functional variants of RuBisCO.
269
270 Another important caveat lies in the observation that, whereas all strains produce increased isotopic
271 fractionation under elevated CO_2, the comparative difference between ancestral AncIB and WT RbcL
272 fractionation is relatively muted under 2% CO_2 relative to ambient air. One possibility is that elevated CO_2
273 brings the intrinsic fractionation properties of RuBisCO into relief (35, 65-67), at least compared to
274 fractionation effects deriving from the overlying organismal physiology. By contrast, in present-day
275 conditions, RuBisCO-mediated fractionation processes may be more significantly overprinted by physical
276 factors that can affect RuBisCO catalytic efficiency, including cellular diffusion of O_2/CO_2 or other factors
277 such as the presence of carbon concentration mechanisms.
278
279 There are many fundamental attributes of extant and ancestral metabolism for which the systemic effects
280 on biosignature production have yet to be characterized. Disentangling these effects is critical for
281 interpretation of the oldest biogeochemical record. A host cyanobacterium *S. elongatus* engineered with
282 a Form IB RbcL ancestor confirms that organism- and enzyme-level effects on biosignature production are
283 not always synonymous but differ in nuanced ways. These differences are contingent upon changes to
284 internal (cellular, physiological) and external (environmental) conditions that have demonstrably varied
285 over Earth's long history. Cyanobacteria, with well-characterized genetic and morphological features (40,
286 41, 61-63) and a tractable paleobiological history (18, 68), are ideal hosts for investigating a range of early
287 Precambrian metabolic processes (68-70).
288
289 Discernible trends (or steadfast consistencies) in metabolic outputs over macroevolutionary timescales
290 can lead to foundational uniformitarian approaches to deep time molecular paleobiology. The available
291 rock record becomes vanishingly sparse with greater age, but it is arguably well-sampled across key global-
292 scale and biotically relevant isotopic systems at least through the early Archaean. Greater geologic
293 sampling will therefore likely generate diminishing returns for shedding new light on deep time
294 paleobiological trends. Innovative approaches that can chart a comprehensive envelope of biomolecular
295 variability over time are a promising new means of reconciling coarse geochemical data with the nuance
296 and complexity of ancient biological activity.
297
298 The engineering of ancient-modern hybrid organisms and their characterization can be used to
299 complement the existing array of fossil remains, biogeochemical signatures, and modern organismal and
300 molecular proxies to assess and contextualize plausible ranges of Precambrian carbon isotope
301 biosignature production. Hybrid organisms may be particularly useful to disentangle the regulatory,
302 physiological, and inter- and intramolecular factors that have impacted isotope fractionation, none of
303 which are individually expressed in the geologic record. These factors must be systematically accounted

for when interpreting bulk fractionation signals, even if only to elucidate the evolutionary molecular underpinnings of uniformitarian phenomena over geologic time.

Conclusions

After engineering a cyanobacterium with an ancient RuBisCO large protein subunit and cultivating it under conditions that mimic those prevailing through much of the Precambrian, we found the resultant carbon isotope fractionation to be within the range of organisms utilizing modern Form IB RuBisCO. The underlying biomolecular and organismal adjustments made by the cell to accommodate the ancestral gene were tracked, and we conclude that the small fractionation differences observed are likely attributable indirectly to decreased fitness of the AncIB strain, which influenced the inorganic carbonate chemistry of the media. The consistency of isotopic signatures generated by this strain indicates that uniformitarian assumptions based on the range of phenotypes of modern RuBisCO variants may apply for Precambrian environmental conditions, but that further study is warranted to discern organism- and enzyme-level trends in carbon isotope fractionation that may extend deeper into the early Precambrian.

Materials and Methods

Inference of ancestral AncIB RbcL protein sequence. A RuBisCO RbcL phylogeny was reconstructed as previously described (42). Briefly, RbcL orthologs were identified from the NCBI protein database by BLAST (sequence dataset and the tree can be found at https://github.com/kacarlab/rubisco). Phylogenetic analysis was performed by Phylobot (71), a web portal that integrates alignment, phylogenetic reconstruction by RAxML (72), and ancestral sequence inference by PAML (73). A maximum likelihood phylogeny was built using a MSAProbs alignment (74) and the best-fit PROTCATWAG model (75, 76), determined by the Akaike information criterion (77). Ancestral states were reconstructed at each amino acid site for all phylogenetic nodes, and gap characters were inferred according to Fitch's parsimony (78).

Cyanobacterial growth and maintenance. *S. elongatus* PCC 7942 strains were cultured in BG-11 medium (79) as liquid cultures or on agar plates (1.5% (w/v) agar and 1 mM $Na_2S_2O_3{\cdot}5H_2O$). Liquid cultures were grown at 30°C, continuous shaking at 120 rpm, sparged with ambient air or 2% CO_2, and with 115 µmol photon·m^{-2}·s^{-1} (with the exception of cultures used to prepare samples for O_2 evolution, which were grown with 80 µmol photon·m^{-2}·s^{-1}). The 2% CO_2 gas mix was controlled by an environment chamber (Percival, Cat. No. I36LLVLC8) with a CO_2 tank input. For recombinant strains, liquid and solid media were supplemented with appropriate antibiotics: 2 µg·ml^{-1} Spectinomycin (Sp) plus 2 µg·ml^{-1} Streptomycin (Sm), 5 µg·ml^{-1} Kanamycin (Km). Cyanobacterial growth was measured at an optical density of 750 nm (OD_{750}) and growth parameters were estimated using the Growthcurver package for R (80) (Growthcurver analysis script can be found at https://github.com/kacarlab/rubisco). Cultures were sampled at the middle exponential growth phase, i.e., at an OD_{750} of ~2.5 (AncIB) or ~4.5 (WT and Syn01) for all subsequent experiments.

Genetic engineering of cyanobacteria. Recombinant strains of *S. elongatus* were constructed by natural transformation using standard protocols (81) with minor modifications (50). The plasmids and strains used in this study are listed in Table 1. The construction of *S. elongatus* Syn01, carrying a single ectopic copy of the *rbc* operon at NS2, as well as the plasmids pSyn01 and pSyn02 used to construct strain Syn01, were described previously (50). The construction of strain AncIB was performed similarly to Syn01. Briefly,

349 pSyn03, which carries the AncIB nucleotide sequence within the entire *rbc* operon (including flanking
350 sequences and homologous regions for recombination at neutral site 2 (NS2) of the *S. elongatus*
351 chromosome), was transformed in *S. elongatus*. Transformation of WT *S. elongatus* with pSyn03
352 generated strain Syn03 carrying a second copy of the *rbc* operon at NS2. Strain Syn03 was subsequently
353 transformed with pSyn01 to replace the native *rbc* operon with a spectinomycin/streptomycin resistance
354 gene as described previously (50), producing strain AncIB. Transformants of Syn03 and AncIB were
355 screened for complete segregation by colony PCR using primers F06, R06, F07, and R07 (Fig. S2; Table S2)
356 and the strain sequences at the deletion and insertion sites were further verified by Sanger sequencing
357 using the primers R07, F08, R08, F15, and F16 (Table S2). To construct plasmid pSyn03, pSyn02 excluding
358 the *rbcL* coding sequence was PCR-amplified and linearized using primers F13/R13 and assembled with
359 the AncIB *RbcL* coding sequence codon-optimized for *S. elongatus* and synthesized by Twist Bioscience.
360 Both DNA fragments were assembled using the GeneArt™ Seamless Cloning and Assembly Kit (Invitrogen,
361 Cat. No. A13288).
362
363 **Analysis of *rbcL* expression by RT-qPCR.** Cells were pelleted by centrifugation and resuspended in TE
364 buffer (10 mM Tris, pH 8.0, 1 mM EDTA). Total RNA was extracted using the RNeasy® Protect Bacteria Mini
365 Kit (QIAGEN, Cat. No. 74524). DNase I-treated RNA was then used in reverse transcription (RT) performed
366 with the SuperScript™ IV First-Strand Synthesis System (Invitrogen, Cat. No. 18091050). F09/R09,
367 F11/R11, F14/R14 pairs of qPCR primers (Table S2) were designed with Primer3Plus
368 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). The quality of cDNA and primer
369 specificity was assessed by PCR using cDNA templates (RT positive reactions) and RT negative controls.
370 qPCR was performed by the real-time thermal cycler qTOWER^3 G (Analytik Jena AG) using qPCRsoft
371 software. The relative expression of native and AncIB *rbcL* was calculated as the average fold change
372 normalized to the *secA* reference gene (53) using the delta-delta Ct method. The experiment was carried
373 out using three biological replicates and three technical replicates.
374
375 **Immunodetection of RbcL protein.** Cells were pelleted by centrifugation and resuspended in 95°C TE
376 buffer supplemented with 1% (w/v) SDS and incubated at 95°C for 10 min. The mixture was sonicated and
377 centrifuged to remove cell debris. Total protein concentration in the crude cell lysates was measured using
378 the Pierce™ BCA Protein Assay Kit (Thermo Scientific, Cat. No. 23225). Lysates containing 5 µg of total
379 protein in Laemmli sample buffer were loaded onto a 6% (v/v) polyacrylamide stacking gel. Proteins were
380 electrophoresed in a 12% polyacrylamide resolving gel and blotted onto a nitrocellulose membrane.
381 Detection of RbcL and total protein load was performed as previously described (50). The densitometric
382 analysis of RbcL signal intensity, normalized to total protein load, was performed with Quantity One®
383 software (Bio-Rad) for three to six biological replicates.
384
385 **Confirmation of RuBisCO assembly.** Assembly of the RuBisCO large and small subunits into a
386 hexadecameric complex in each strain was evaluated by native gel electrophoresis and immunodetection,
387 as previously described (50). Immunodetection of the RuBisCO complex was performed for three
388 biological replicates with the same primary and secondary antibodies that were used to detect RbcL, as
389 described above.
390
391 **Catalytic activity of RuBisCO.** The activity of RuBisCO in cyanobacterial lysates was measured using a
392 spectrophotometric coupled-enzyme assay that links this activity with the rate of NADH oxidation (82).
393 Cell lysis and the activity assay were carried out as previously described (50) with either 2.5 mM or 5 mM
394 $NaHCO_3$. After 20 min at 25 °C for activation of Rubisco, the reaction was initialized with the addition of
395 ribulose 1,5-bisphosphate (RuBP) (0.5 mM) and the absorbance at 340 nm was recorded using a Synergy

396 H1 plate reader (BioTek). RuBisCO activity was reported as the RuBP consumption rate normalized to total
397 soluble protein content. The assay was performed for three biological replicates.
398
399 **Photosynthetic oxygen evolution rate.** *S. elongatus* strain photosynthetic activity was assayed using a
400 Clark-type oxygen electrode chamber to measure the level of molecular oxygen produced in
401 cyanobacterial cultures. Cells were pelleted and resuspended in fresh BG-11 to an OD_{750} of ~1 following
402 De Porcelinis (83). Concentration of chlorophyll *a* (for normalization) was measured following the protocol
403 by Zavrel et al. (84). The remaining suspension was incubated in the dark for 20 min with gentle agitation.
404 Samples from each suspension were analyzed in an oxygen electrode chamber under saturated light, using
405 the Oxygraph+ System (Hansatech Instruments) equipped with the OxyTrace+ software. Oxygen evolution
406 rate was monitored for 10 min and expressed as nanomoles of molecular oxygen evolved per hour per
407 microgram of chlorophyll *a*. The assay was performed for three biological replicates.
408
409 **Carbon isotope fractionation in bulk cyanobacterial biomass.** Cells were pelleted by centrifugation and
410 washed in 10 mL of 10 mM NaCl (OD_{750} for Syn-1 ~ 4.5, OD_{750} for AncIB ~2.5). Pellets were then dried at
411 50°C. In parallel, the supernatant from centrifuged culture samples was sterilized through 0.2 µm filtration
412 for DIC isotopic analysis of growth media. Sterilized media was transferred to Exetainer vials leaving no
413 headspace and stored at 4°C until analysis. Isotopic analysis was performed for three biological replicates.
414 The carbon isotope composition of bulk biomass ($\delta^{13}C_{biomass}$) and DIC ($\delta^{13}C_{DIC}$) was determined at the UC
415 Davis Stable Isotope Facility. $\delta^{13}C_{biomass}$ was analyzed using a PDZ Europa ANCA-GSL elemental analyzer
416 interfaced to a PDZ Europa 20-20 isotope ratio mass spectrometer (Sercon Ltd.). DIC samples were
417 analyzed by gas evolution and composition was measured by a GasBench II system interfaced to a Delta
418 V Plus IRMS (Thermo Scientific). The carbon isotopic composition values were reported relative to the
419 Vienna PeeDee Belemnite standard (V-PDB):
420

$$\delta^{13}C_{sample} = \left(\frac{{}^{13}C / {}^{12}C_{sample}}{{}^{13}C / {}^{12}C_{V-PDB}} - 1 \right) \times 1000$$

421
422
423 The isotopic composition of dissolved molecular CO_2 ($\delta^{13}C_{CO2}$) was estimated from $\delta^{13}C_{DIC}$ following Rau et
424 al. (55) and Mook et al. (56):
425

$$\delta^{13}C_{CO_2} = \delta^{13}C_{DIC} + 23.644 - \frac{9701.5}{T_K}$$

426
427
428 The carbon isotope fractionation associated with photosynthetic CO_2 fixation (ε_p) was calculated relative
429 to $\delta^{13}C_{CO2}$ in the post-culture medium according to Freeman and Hayes (1992):
430

$$\varepsilon_p = \frac{\delta^{13}C_{CO_2} - \delta^{13}C_{biomass}}{1 + \delta^{13}C_{biomass} / 1000}$$

431
432
433 **Statistical analyses.** Results for experimental analyses were presented as the mean and the sample
434 standard deviation (1σ) values of at least three biological replicates. For comparisons of two groups,
435 statistical significance was analyzed by an unpaired, two-tailed *t*-test assuming equal variance. For
436 comparisons of three or more groups, significance was analyzed by one-way ANOVA and a post-hoc Tukey
437 HSD test.

Acknowledgements

We sincerely thank Emily Peñaherrera, Jenan Kharbush, Ryan Ward, Sky Dominguez, and the University of California-Davis Stable Isotope Facility for assistance. This work was supported by the National Aeronautics and Space Administration Early Career Faculty (ECF) Award No. 80NSSC19K1617 (BK), the National Science Foundation Emerging Frontiers Program Award No. 1724090 (BK), National Aeronautics and Space Administration Postdoctoral Fellowship (AKG), Simons Foundation Early Career Award No. 561645 (JNY and ML), and National Institute of General Medical Sciences of the National Institutes of Health award number R01GM118815 (AT, to James W. Golden at the University of California-San Diego).

References

1. N. J. Butterfield, Macroevolution and Macroecology through Deep Time. *Palaeontology* **50**, 41-55 (2007). https://doi.org/10.1111/j.1475-4983.2006.00613.x
2. B. Hannisdal, S. E. Peters, Phanerozoic Earth System Evolution and Marine Biodiversity. *Science* **334**, 1121-1124 (2011). https://doi.org/10.1126/science.1210695
3. R. K. Bambach, "Classes and Adaptive Variety: The Ecology of Diversification in Marine Faunas Through the Phanerozoic" in *Phanerozoic Diversity Patterns,* J. W. Valentine, Ed. (Princeton University Press, 1986), pp. 191-254.
4. D. J. Bottjer, W. I. Ausich, Phanerozoic development of tiering in soft substrata suspension-feeding communities. *Paleobiology* **12**, 400-420 (2016). https://doi.org/10.1017/s0094837300003134
5. J. J. Sepkoski, Biodiversity: Past, Present, and Future. *J Paleontol* **71**, 533-539 (2015). https://doi.org/10.1017/s0022336000040026
6. G. J. Vermeij, Adaptation, Versatility, and Evolution. *Syst Zool* **22** (1973). https://doi.org/10.2307/2412953
7. A. J. W. Hendy, "Taphonomic Overprints on Phanerozoic Trends in Biodiversity: Lithification and Other Secular Megabiases" in *Taphonomy,* P. A. Allison, D. J. Bottjer, Eds. (Springer, Dordrecht, 2010), pp. 19-77.
8. S. E. Peters, Geologic constraints on the macroevolutionary history of marine animals. *Proc Natl Acad Sci USA* **102**, 12326-12331 (2005). https://doi.org/10.1073/pnas.0502616102
9. S. E. Peters, M. Foote, Biodiversity in the Phanerozoic: a reinterpretation. *Paleobiology* **27**, 583-601 (2016). https://doi.org/10.1666/0094-8373(2001)027<0583:Bitpar>2.0.Co;2
10. A. H. Knoll, Paleobiological Perspectives on Early Microbial Evolution. *Cold Spring Harbor Perspectives in Biology* **7**, a018093 (2015). https://doi.org/10.1101/cshperspect.a018093
11. J. W. Schopf, Disparate rates, differing fates: tempo and mode of evolution changed from the Precambrian to the Phanerozoic. *Proc Natl Acad Sci USA* **91**, 6735-6742 (1994). https://doi.org/10.1073/pnas.91.15.6735
12. D. H. Erwin, Evolutionary uniformitarianism. *Dev Biol* **357**, 27-34 (2011). https://doi.org/10.1016/j.ydbio.2011.01.020
13. Z. R. Adam, M. L. Skidmore, D. W. Mogk, N. J. Butterfield, A Laurentian record of the earliest fossil eukaryotes. *Geology* **45**, 387-390 (2017). https://doi.org/10.1130/g38749.1
14. Z. Yin, M. Zhu, D. J. Bottjer, F. Zhao, P. Tafforeau, Meroblastic cleavage identifies some Ediacaran Doushantuo (China) embryo-like fossils as metazoans. *Geology* **44**, 735-738 (2016). https://doi.org/10.1130/g38262.1

15. N. J. Butterfield, *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386-404 (2000). https://doi.org/10.1666/0094-8373(2000)026<0386:Bpngns>2.0.Co;2
16. Q. Tang, K. Pang, X. Yuan, S. Xiao, A one-billion-year-old multicellular chlorophyte. *Nat Ecol Evol* **4**, 543-549 (2020). https://doi.org/10.1038/s41559-020-1122-9
17. S. Golubic, L. Seong-Joo, Early cyanobacterial fossil record: preservation, palaeoenvironments and identification. *Eur J Phycol* **34**, 339-348 (1999). https://doi.org/10.1080/09670269910001736402
18. J. W. Schopf, The paleobiological record of photosynthesis. *Photosynth Res* **107**, 87-101 (2011). https://doi.org/10.1007/s11120-010-9577-1
19. M. Schidlowski, Carbon isotopes as biogeochemical recorders of life over 3.8 Ga of Earth history: evolution of a concept. *Precambrian Res* **106**, 117-134 (2001). https://doi.org/Doi 10.1016/S0301-9268(00)00128-5
20. B. L. Beard, Iron Isotope Biosignatures. *Science* **285**, 1889-1892 (1999). https://doi.org/10.1126/science.285.5435.1889
21. A. K. Garcia, C. M. Cavanaugh, B. Kacar, The curious consistency of carbon biosignatures over billions of years of Earth-life coevolution. *ISME J* (2021). https://doi.org/10.1038/s41396-021-00971-5
22. A. D. Anbar, O. Rouxel, Metal Stable Isotopes in Paleoceanography. *Annu Rev Earth Pl Sc* **35**, 717-746 (2007). https://doi.org/10.1146/annurev.earth.34.031405.125029
23. J. W. Schopf *et al.*, Sulfur-cycling fossil bacteria from the 1.8-Ga Duck Creek Formation provide promising evidence of evolution's null hypothesis. *Proc Natl Acad Sci USA* **112**, 2087-2092 (2015). https://doi.org/10.1073/pnas.1419241112
24. J. Krissansen-Totton, R. Buick, D. C. Catling, A statistical analysis of the carbon isotope record from the Archean to Phanerozoic and implications for the rise of oxygen. *Am J Sci* **315**, 275-316 (2015). https://doi.org/10.2475/04.2015.01
25. M. Schidlowski, A 3,800-Million-Year Isotopic Record of Life from Carbon in Sedimentary-Rocks. *Nature* **333**, 313-318 (1988). https://doi.org/10.1038/333313a0
26. J. W. Schopf, K. Kitajima, M. J. Spicuzza, A. B. Kudryavtsev, J. W. Valley, SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. *Proc Natl Acad Sci USA* **115**, 53-58 (2018). https://doi.org/10.1073/pnas.1718063115
27. E. A. Bell, P. Boehnke, T. M. Harrison, W. L. Mao, Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc Natl Acad Sci USA* **112**, 14518-14521 (2015). https://doi.org/10.1073/pnas.1517557112
28. K. M. Scott *et al.*, Kinetic isotope effect and biochemical characterization of form IA RubisCO from the marine cyanobacterium *Prochlorococcus marinus* MIT9313. *Limnol Oceanogr* **52**, 2199-2204 (2007).
29. R. D. Guy, M. L. Fogel, J. A. Berry, Photosynthetic fractionation of the stable isotopes of oxygen and carbon. *Plant Physiol* **101**, 37-47 (1993). https://doi.org/10.1104/pp.101.1.37
30. S. von Caemmerer, Y. Tazoe, J. R. Evans, S. M. Whitney, Exploiting transplastomically modified Rubisco to rapidly measure natural diversity in its carbon isotope discrimination using tuneable diode laser spectroscopy. *J Exp Bot* **65**, 3759-3767 (2014). https://doi.org/10.1093/jxb/eru036
31. C. A. Roeske, M. H. O'Leary, Carbon isotope effects on the enzyme-catalysed carboxylation of ribulose bisphosphate. *Biochemistry* **23**, 6275-6284 (1984).
32. M. Eichner, S. Thoms, S. A. Kranz, B. Rost, Cellular inorganic carbon fluxes in *Trichodesmium*: a combined approach using measurements and modelling. *J Exp Bot* **66**, 749-759 (2015). https://doi.org/10.1093/jxb/eru427

33. K. H. Freeman, J. M. Hayes, Fractionation of carbon isotopes by phytoplankton and estimates of ancient CO_2 levels. *Global Biogeochem Cycles* **6**, 185-198 (1992). https://doi.org/10.1029/92gb00190
34. S. J. Hurley, B. A. Wing, C. E. Jasper, N. C. Hill, J. C. Cameron, Carbon isotope evidence for the global physiology of Proterozoic cyanobacteria. *Sci Adv* **7** (2021). https://doi.org/10.1126/sciadv.abc8998
35. E. B. Wilkes, R. B. Y. Lee, H. L. O. McClelland, R. E. M. Rickaby, A. Pearson, Carbon isotope ratios of coccolith–associated polysaccharides of *Emiliania huxleyi* as a function of growth rate and CO_2 concentration. *Org Geochem* **119**, 1-10 (2018). https://doi.org/10.1016/j.orggeochem.2018.02.006
36. G. G. Tcherkez, G. D. Farquhar, T. J. Andrews, Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci USA* **103**, 7246-7251 (2006). https://doi.org/10.1073/pnas.0600605103
37. R. A. Studer, P. A. Christin, M. A. Williams, C. A. Orengo, Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci USA* **111**, 2223-2228 (2014). https://doi.org/10.1073/pnas.1310811111
38. A. I. Flamholz *et al.*, Revisiting trade-offs between Rubisco kinetic parameters. *Biochemistry* **58**, 3365-3376 (2019). https://doi.org/10.1021/acs.biochem.9b00237
39. D. B. McNevin *et al.*, Differences in carbon isotope discrimination of three variants of D-ribulose-1,5-bisphosphate carboxylase/oxygenase reflect differences in their catalytic mechanisms. *J Biol Chem* **282**, 36068-36076 (2007). https://doi.org/10.1074/jbc.M706274200
40. B. M. Berla *et al.*, Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front Microbiol* **4** (2013). https://doi.org/10.3389/fmicb.2013.00246
41. Y. Chen, C. K. Holtman, A. Taton, S. S. Golden, "Functional Analysis of the Synechococcus elongatus PCC 7942 Genome" in *Functional Genomics and Evolution of Photosynthetic Systems,* R. Burnap, W. Vermaas, Eds. (Springer, Dordrecht, 2012), pp. 119-137.
42. B. Kacar, V. Hanson-Smith, Z. R. Adam, N. Boekelheide, Constraining the timing of the Great Oxidation Event within the Rubisco phylogenetic tree. *Geobiology* **15**, 628-640 (2017). https://doi.org/10.1111/gbi.12243
43. H. S. Yoon, J. D. Hackett, C. Ciniglia, G. Pinto, D. Bhattacharya, A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Mol Biol Evol* **21**, 809-818 (2004). https://doi.org/10.1093/molbev/msh075
44. S. Maruyama, E. Kim, "Evolution of Photosynthetic Eukaryotes; Current Opinion, Perplexity, and a New Perspective" in *Symbiosis: Cellular, Molecular, Medical and Evolutionary Aspects,* M. Kloc, Ed. (Springer, Cham, 2020), pp. 337-351.
45. P. M. Shih *et al.*, Biochemical characterization of predicted Precambrian RuBisCO. *Nat Commun* **7**, 10382 (2016). https://doi.org/10.1038/ncomms10382
46. T. M. Gibson *et al.*, Precise age of *Bangiomorpha pubescens* dates the origin of eukaryotic photosynthesis. *Geology* **46**, 135-138 (2017). https://doi.org/10.1130/g39829.1
47. S. A. Crowe *et al.*, Atmospheric oxygenation three billion years ago. *Nature* **501**, 535-538 (2013). https://doi.org/10.1038/nature12426
48. A. D. Anbar *et al.*, A whiff of oxygen before the great oxidation event? *Science* **317**, 1903-1906 (2007). https://doi.org/10.1126/science.1140325
49. J. Newman, C. I. Brändén, T. A. Jones, Structure determination and refinement of ribulose 1,5-bisphosphate carboxylase/oxygenase from *Synechococcus* PCC6301. *Acta Crystallogr D* **49**, 548-560 (1993). https://doi.org/10.1107/s090744499300530x

50. A. K. Garcia *et al.*, System-level effects of CO_2 and RuBisCO concentration on carbon isotope fractionation. bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.04.20.440233 (accessed 28 May 2021).
51. D. C. Catling, K. J. Zahnle, The Archean atmosphere. *Sci Adv* **6**, eaax1420 (2020). https://doi.org/10.1126/sciadv.aax1420
52. T. Zavřel, M. A. Sinetova, D. Búzová, P. Literáková, J. Červený, Characterization of a model cyanobacterium *Synechocystis* sp. PCC 6803 autotrophic growth in a flat-panel photobioreactor. *Eng Life Sci* **15**, 122-132 (2015). https://doi.org/10.1002/elsc.201300165
53. E. Szekeres, C. Sicora, N. Dragos, B. Druga, Selection of proper reference genes for the cyanobacterium Synechococcus PCC 7002 using real-time quantitative PCR. *FEMS Microbiol Lett* **359**, 102-109 (2014). https://doi.org/10.1111/1574-6968.12574
54. D. S. Kubien, S. M. Whitney, P. V. Moore, L. K. Jesson, The biochemistry of Rubisco in Flaveria. *J Exp Bot* **59**, 1767-1777 (2008). https://doi.org/10.1093/jxb/erm283
55. G. H. Rau, U. Riebesell, D. Wolf-Gladrow, A model of photosynthetic ^{13}C fractionation by marine phytoplankton based on diffusive molecular CO_2 uptake. *Mar Ecol Prog Ser* **133**, 275-285 (1996). https://doi.org/10.3354/meps133275
56. W. G. Mook, J. C. Bommerson, W. H. Staverman, Carbon isotope fractionation between dissolved bicarbonate and gaseous carbon dioxide. *Earth Planet Sc Lett* **22**, 169-176 (1974). https://doi.org/10.1016/0012-821x(74)90078-8
57. L. H. Gunn, E. Martin Avila, R. Birch, S. M. Whitney, The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth. *Proc Natl Acad Sci USA* **117**, 25890-25896 (2020). https://doi.org/10.1073/pnas.2011641117
58. B. Kacar, X. Ge, S. Sanyal, E. A. Gaucher, Experimental Evolution of Escherichia coli Harboring an Ancient Translation Protein. *J Mol Evol* **84**, 69-84 (2017). https://doi.org/10.1007/s00239-017-9781-0
59. M. Goldsmith, D. S. Tawfik, Potential role of phenotypic mutations in the evolution of protein expression and stability. *Proc Natl Acad Sci USA* **106**, 6197-6202 (2009). https://doi.org/10.1073/pnas.0809506106
60. P.-A. Christin *et al.*, Evolutionary Switch and Genetic Convergence on rbcL following the Evolution of C4 Photosynthesis. *Mol Biol Evol* **25**, 2361-2368 (2008). https://doi.org/10.1093/molbev/msn178
61. B. M. Long, B. Förster, S. B. Pulsford, G. D. Price, M. R. Badger, Rubisco proton production can drive the elevation of CO_2 within condensates and carboxysomes. *Proc Natl Acad Sci USA* **118** (2021). https://doi.org/10.1073/pnas.2014406118
62. R. Burnap, M. Hagemann, A. Kaplan, Regulation of CO_2 Concentrating Mechanism in Cyanobacteria. *Life* **5**, 348-371 (2015). https://doi.org/10.3390/life5010348
63. Jeffrey C. Cameron, Steven C. Wilson, Susan L. Bernstein, Cheryl A. Kerfeld, Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell* **155**, 1131-1140 (2013). https://doi.org/10.1016/j.cell.2013.10.044
64. D. Shi, Y. Xu, F. M. M. Morel, Effects of the pH/pCO_2 control method on medium chemistry and phytoplankton growth. *Biogeosciences* **6**, 1199-1207 (2009). https://doi.org/10.5194/bg-6-1199-2009
65. J. M. Hayes, Factors controlling 13C contents of sedimentary organic compounds: Principles and evidence. *Mar Geol* **113**, 111-125 (1993). https://doi.org/10.1016/0025-3227(93)90153-m
66. R. R. Bidigare *et al.*, Consistent fractionation of ^{13}C in nature and in the laboratory: growth-rate effects in some haptophyte algae. *Global Biogeochem Cycles* **11**, 279-292 (1997). https://doi.org/10.1029/96gb03939

624 67. B. A. Schubert, A. H. Jahren, The effect of atmospheric CO_2 concentration on carbon isotope
625 fractionation in C_3 land plants. *Geochim Cosmochim Acta* **96**, 29-43 (2012).
626 https://doi.org/10.1016/j.gca.2012.08.003
627 68. B. Kacar, L. Guy, E. Smith, J. Baross, Resurrecting ancestral genes in bacteria to interpret ancient
628 biosignatures. *Philos Trans R Soc A* **375** (2017). https://doi.org/10.1098/rsta.2016.0352
629 69. A. K. Garcia, H. McShea, B. Kolaczkowski, B. Kacar, Reconstructing the evolutionary history of
630 nitrogenases: Evidence for ancestral molybdenum-cofactor utilization. *Geobiology* **18**, 394-411
631 (2020). https://doi.org/10.1111/gbi.12381
632 70. A. K. Garcia, B. Kacar, How to resurrect ancestral proteins as proxies for ancient
633 biogeochemistry. *Free Radic Biol Med* **140**, 260-269 (2019).
634 https://doi.org/10.1016/j.freeradbiomed.2019.03.033
635 71. V. Hanson-Smith, A. Johnson, PhyloBot: A Web Portal for Automated Phylogenetics, Ancestral
636 Sequence Reconstruction, and Exploration of Mutational Trajectories. *PLoS Comput Biol* **12**,
637 e1004976 (2016). https://doi.org/10.1371/journal.pcbi.1004976
638 72. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
639 phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
640 https://doi.org/10.1093/bioinformatics/btu033
641 73. Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591
642 (2007). https://doi.org/10.1093/molbev/msm088
643 74. Y. Liu, B. Schmidt, D. L. Maskell, MSAProbs: multiple sequence alignment based on pair hidden
644 Markov models and partition function posterior probabilities. *Bioinformatics* **26**, 1958-1964
645 (2010). https://doi.org/10.1093/bioinformatics/btq338
646 75. N. Lartillot, H. Philippe, A Bayesian mixture model for across-site heterogeneities in the amino-
647 acid replacement process. *Mol Biol Evol* **21**, 1095-1109 (2004).
648 https://doi.org/10.1093/molbev/msh112
649 76. S. Whelan, N. Goldman, A general empirical model of protein evolution derived from multiple
650 protein families using a maximum-likelihood approach. *Mol Biol Evol* **18**, 691-699 (2001).
651 https://doi.org/10.1093/oxfordjournals.molbev.a003851
652 77. F. Abascal, R. Zardoya, D. Posada, ProtTest: selection of best-fit models of protein evolution.
653 *Bioinformatics* **21**, 2104-2105 (2005). https://doi.org/10.1093/bioinformatics/bti263
654 78. W. M. Fitch, Toward Defining the Course of Evolution: Minimum Change for a Specific Tree
655 Topology. *Syst Zool* **20** (1971). https://doi.org/10.2307/2412116
656 79. R. Rippka, J. Deruelles, J. B. Waterbury, M. Herdman, R. Y. Stanier, Generic Assignments, Strain
657 Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology* **111**, 1-61 (1979).
658 https://doi.org/https://doi.org/10.1099/00221287-111-1-1
659 80. K. Sprouffske, A. Wagner, Growthcurver: an R package for obtaining interpretable metrics from
660 microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016). https://doi.org/10.1186/s12859-
661 016-1016-7
662 81. E. M. Clerico, J. L. Ditty, S. S. Golden, Specialized techniques for site-directed mutagenesis in
663 cyanobacteria. *Methods Mol Biol* **362**, 155-171 (2007). https://doi.org/10.1007/978-1-59745-
664 257-1_11
665 82. D. S. Kubien, C. M. Brown, H. J. Kane, Quantifying the amount and activity of Rubisco in leaves.
666 *Methods Mol Biol* **684**, 349-362 (2011). https://doi.org/10.1007/978-1-60761-925-3_27
667 83. A. J. De Porcellinis *et al.*, Overexpression of bifunctional fructose-1,6-
668 bisphosphatase/sedoheptulose-1,7-bisphosphatase leads to enhanced photosynthesis and
669 global reprogramming of carbon metabolism in *Synechococcus* sp. PCC 7002. *Metabolic*
670 *Engineering* **47**, 170-183 (2018). https://doi.org/10.1016/j.ymben.2018.03.001

84. T. Zavřel, M. Sinetova, J. Červený, Measurement of Chlorophyll *a* and Carotenoids Concentration in Cyanobacteria. *Bio-Protocol* **5** (2015). https://doi.org/10.21769/BioProtoc.1467

Figures and Figure Legends

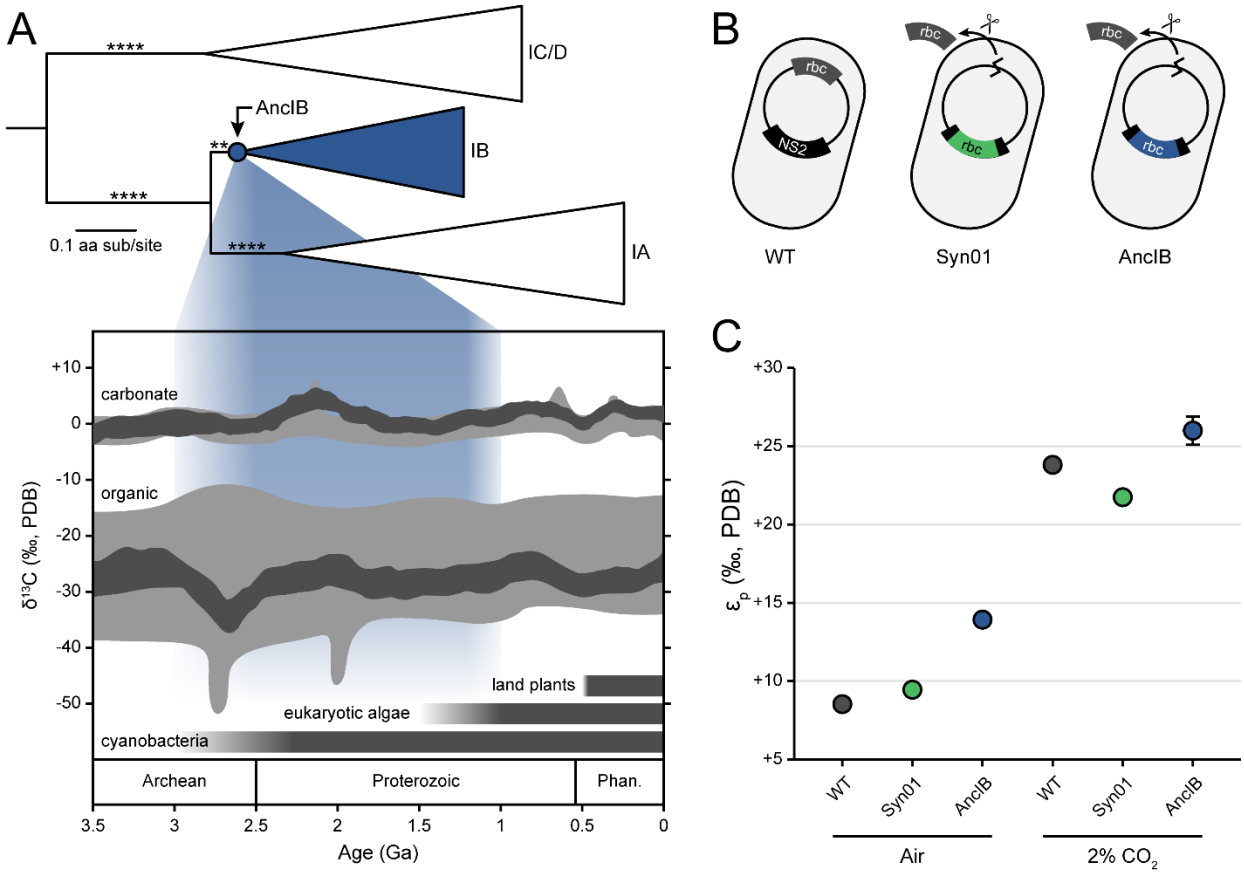


Figure 1. Reconstruction of ancestral RuBisCO biogeochemical signatures. (*A*) Maximum likelihood Form I RuBisCO RbcL phylogeny (derived from full RuBisCO phylogeny described in Kacar et al. (42)). Ancestral AncIB node and descendent Form IB clade highlighted in blue. Approximate likelihood ratio (aLR) branch support indicated by asterisks (**: >10, ****: >1000). Carbon isotope record figure adapted from Garcia et al. (21), with data from Schidlowski et al. (25) (grey) and Krissansen-Totton et al. (24) (dark grey). Approximate age range of AncIB indicated by blue field (see text for discussion). (*B*) Genetic engineering of *S. elongatus* strains. Strain Syn01 was constructed by inserting a second copy of the *rbc* operon in the chromosomal neutral site 2 (NS2). Strain AncIB was constructed by inserting the genetic sequence encoding for the ancestral AncIB *rbcL* within the NS2 *rbc* operon. The native *rbc* operon was removed in both strains Syn01 and AncIB. (*C*) Photosynthetic carbon isotope fractionation (ε_p) of *S. elongatus* strains in this study, cultured in ambient air or 2% CO_2. n = 3 for each data point and error bars indicate 1σ (error bars smaller than some datapoints).

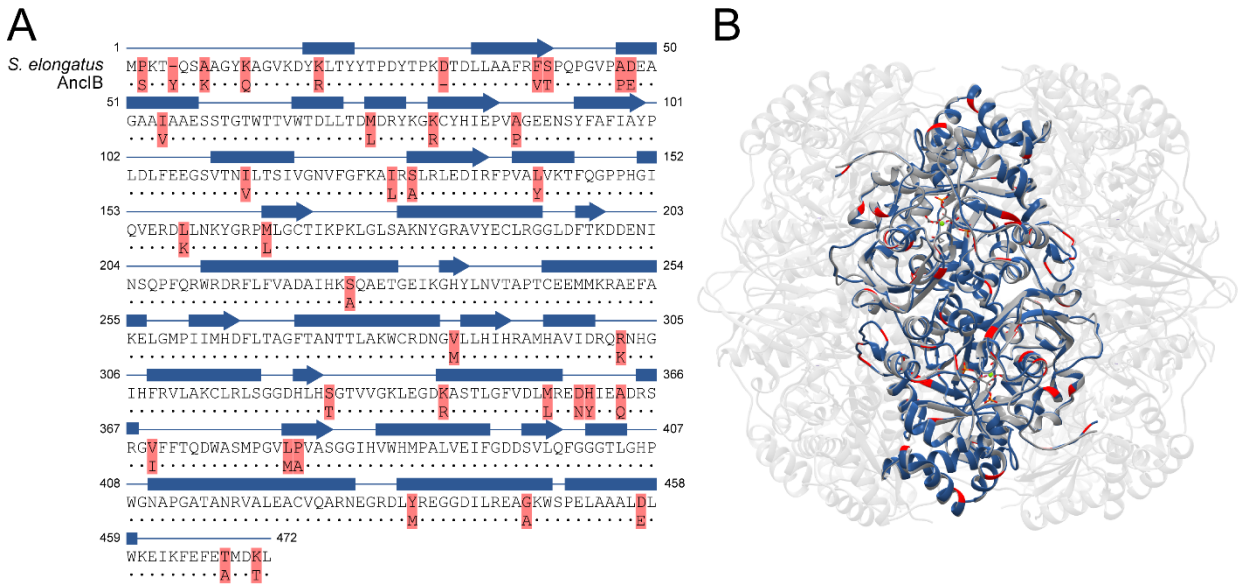
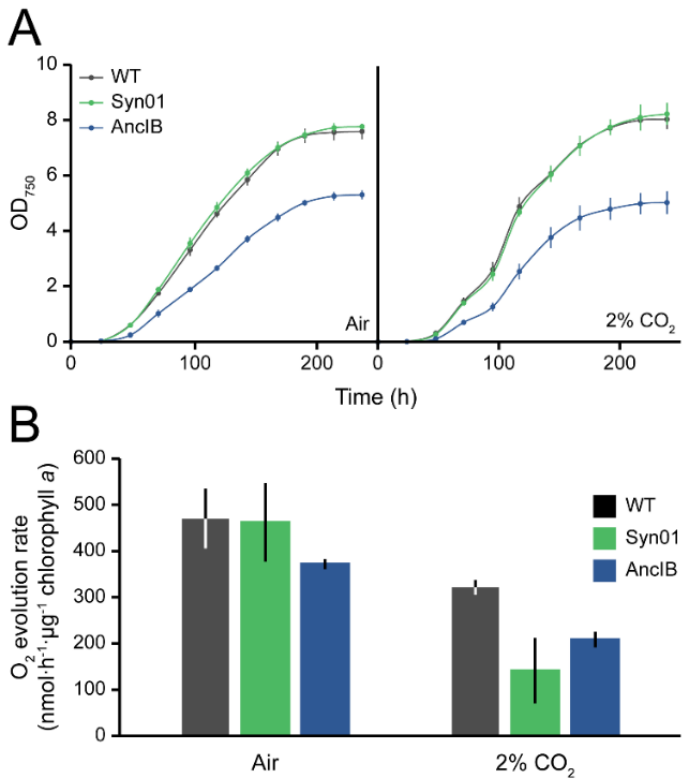


Figure 2. Structure and sequence features of ancestral RuBisCO. (*A*) Amino acid sequence alignment between ancestral AncIB and extant *S. elongatus* RbcL. Ancestral site variation relative to the *S. elongatus* template is highlighted in red. (*B*) Modeled structure of the ancestral AncIB L_2 dimer (blue), aligned to the active conformation of the extant *S. elongatus* L_8S_8 hexadecamer (grey; PDB: 1RBL (49)). Highlighted residues in (*A*) are also highlighted in (*B*). Site numbering from extant *S. elongatus*. Conserved residues are indicated by dots and secondary structure is indicated above the sequences (blue rectangle: α-helix; blue arrow: β-sheet).

701



702

703 **Figure 3.** (*A*) Growth curves and (*B*) photosynthetic oxygen evolution of *S. elongatus* strains cultured in
704 ambient air or 2% CO_2. (*A, B*) n = 3 for each data point or bar and error bars indicate 1σ.

705

706

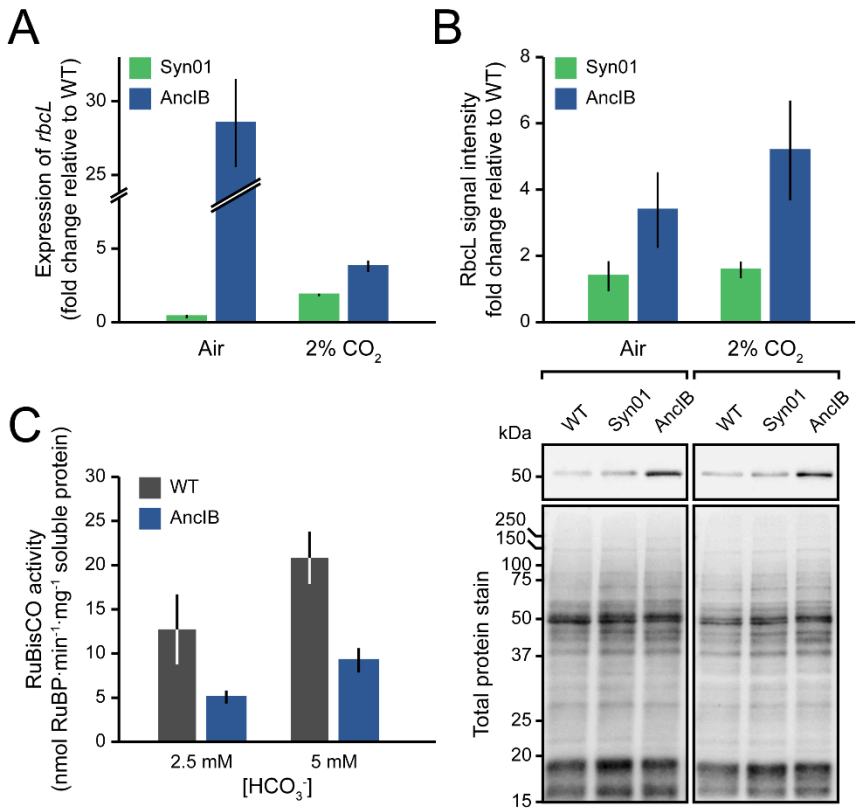


Figure 4. Expression and total cell activity of RuBisCO in *S. elongatus* strains. (*A*) Expression of *rbcL* in Syn01 and AncIB detected by RT-qPCR (*secA* reference gene), relative to WT. (*B*) Immunodetection of RbcL protein. *Top*, RbcL signal intensities normalized to those for total soluble protein load. *Bottom*, Western blot showing RbcL protein detected by anti-RuBisCO antibody and total protein stain from crude cell lysates (*C*) Total cell RuBisCO activity, measured with 2.5 mM and 5 mM HCO_3^- concentrations. (*A-C*) n = 3 for each bar (except 2% CO_2 RbcL signal intensity data, n = 6) and error bars indicate 1σ.

Tables

Table 1. Strains and plasmids used in this study.

Strain or plasmid	Description/Genotype	Antibiotic resistance	Source/Reference
WT	Wild-type strain of *S. elongatus* PCC 7942	-	Susan S. Golden (UC San Diego)
Syn01	*S. elongatus* strain Syn02 with the native *rbc* operon removed: Syn02 and Δ(*rbcL-rbcS-purK)*::*aadA*.	Km, Sp+Sm	(50)
Syn02	*S. elongatus* PCC 7942 carrying a second copy of the *rbc* operon and flanking sequences at NS2: NS2::*aphI-rbcL-rbcS-purK*-Synpcc7942_1429-Synpcc7942_1430	Km	(50)
Syn03	*S. elongatus* PCC 7942 carrying a modified copy of the *rbc* operon (with the ancestral *rbcL*) and flanking sequences at NS2: NS2::*aphI-ancIB-rbcS-purK*-Synpcc7942_1429-Synpcc7942_1430	Km, Sp+Sm	This study
AncIB	*S. elongatus* strain Syn03 carrying the ancestral AncIB *rbcL* gene in the *rbc* operon copy at NS2 and having the native *rbc* operon replaced with a Sp/Sm resistance gene: Δ(*rbcL-rbcS-purK)*::*aadA*.	Km	This study
pSyn01	Plasmid to replace *S. elongatus'* native *rbc* operon (CP000100: 1479070-1482595) with a Sp/Sm resistance gene: Δ(*rbcL-rbcS-purK)*::*aadA*.	Sp+Sm	This study
pSyn02	Plasmid for recombination at NS2 of *S. elongatus* chromosome carrying the *rbc* operon including *rbcL*, *rbcS*, *purK,* and flanking sequences from *S. elongatus* PCC 7942 (CP000100: 1479071-1484283)	Km	(50)
pSyn03	pSyn02 in which the coding sequence of RbcL was replaced with the coding sequence of AncIB	Km	This study

723 **Table 2.** Growth parameters of *S. elongatus* strains. Values represent the mean of three replicates ± 1σ.
724 Asterisks indicate significance relative to WT for the same atmospheric condition, determined by one-
725 way ANOVA and post-hoc Tukey HSD tests (**: $p < 0.01$; ***: $p < 0.001$).
726

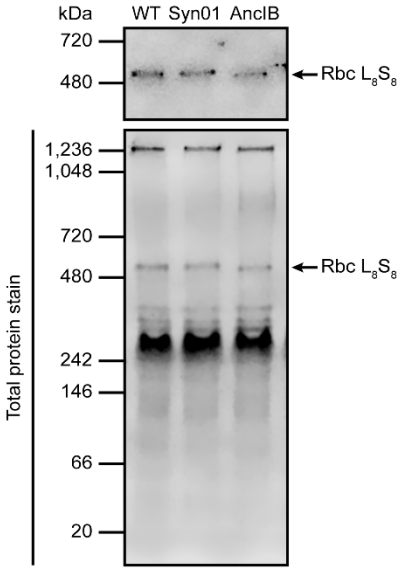
Strain	Atmosphere	Doubling time (h)	Midpoint time (h)	Maximum cell density (OD_{750})
WT	Air	18.9 ± 1.0	106.4 ± 2.8	7.7 ± 0.3
Syn01		18.6 ± 0.3	103.8 ± 1.9	7.7 ± 0.1
AncIB		20.5 ± 0.4	118.4 ± 2.1**	5.4 ± 0.2***
WT	2% CO_2	16.9 ± 1.5	110.8 ± 6.1	8.0 ± 0.3
Syn01		17.3 ± 1.0	113.9 ± 2.3	8.1 ± 0.5
AncIB		15.8 ± 0.5	117.8 ± 1.7	5.0 ± 0.4***

727
728
729

730 # Supplemental Information

731

732

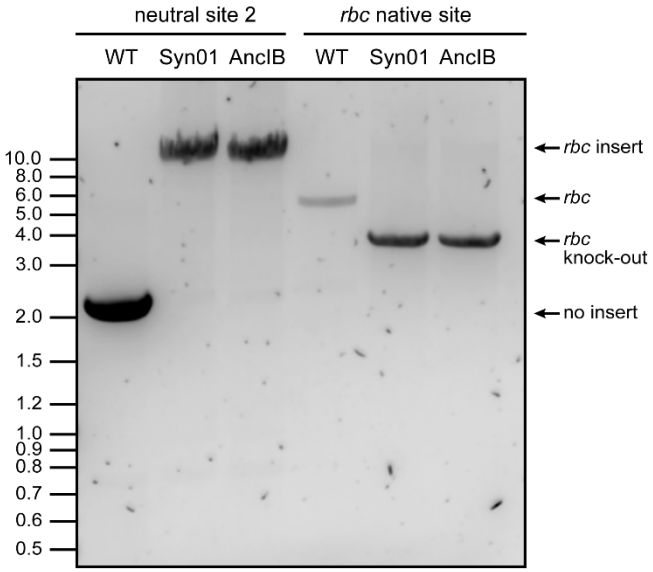


733

734 **Figure S1.** Immunodetection of assembled RuBisCO. Western blot showing the assembly of RbcL (WT and
735 AncIB) and RbcS into the L_8S_8 hexadecameric complex (520 kDa), detected by anti-RbcL antibody.

736

737



738

739 **Figure S2.** Genotyping of *S. elongatus* strains. Primers F06/R06 (Table S2) were used to confirm the *rbc*
740 operon insertion at the neutral site 2 (NS2), either with WT *rbcL* (for Syn01) or AncIB *rbcL* (*rbc* insert:
741 8,617 bp; no insert: 1,970 bp). Primers F07/R07 were used to confirm the presence or absence of the rbc
742 operon at the native site (*rbc* present at its native site: 5,140 bp; *rbc* knocked out and replaced with the
743 aadA gene: 3,342 bp).

744

745

746 **Table S1.** Isotopic composition of biomass and DIC in growth medium of *S. elongatus* cultures.

Strain	Atmosphere	$\delta^{13}C_{biomass}$ (‰)	$\delta^{13}C_{DIC}$ (‰)	$\delta^{13}C_{CO2}$ (‰)	[DIC] (mM)
WT	Air	-19.63 ± 0.07	-2.92 ± 0.02	-11.28 ± 0.02	5.14 ± 0.07
	2% CO_2	-26.41 ± 0.15	5.77 ± 0.30	-2.59 ± 0.30	7.35 ± 0.24
Syn01	Air	-20.20 ± 0.19	-2.63 ± 0.10	-11.00 ± 0.10	5.23 ± 0.02
	2% CO_2	-26.84 ± 0.10	3.29 ± 0.25	-5.07 ± 0.25	7.53 ± 0.34
AncIB	Air	-20.60 ± 0.08	1.26 ± 0.10	-7.10 ± 0.10	3.44 ± 0.05
	2% CO_2	-31.2 ± 0.34	3.15 ± 0.62	-5.21 ± 0.62	5.70 ± 0.70

747
748
749 **Table S2.** Primers used in this study.

Primer[a]	**Sequence (5'-3')**	**Description**
F06	GACAATCCTGTTCTCCGGCA	Genotyping *S. elongatus* strains to screen for the *rbc* operon insert (either with native or AncIB *rbcL*) at NS2 (PCR product size: with *rbc* insert – 8,617 bp, without *rbc* insert – 1,970 bp).
R06	ATCAACGCCGTACCCGTATC	
F07	GGAGTCAATTCTGCAAGAGC	Genotyping *S. elongatus* strains to confirm the presence or absence of the *rbc* operon at the native site (PCR product size: with the *rbc* operon – 5,140 bp, without the *rbc* operon – 3,342 bp) R07 was also used to sequence the native *rbc* operon deletion site.
R07	TCAAGCTCGGTCTACTGC	
F08	GAATGCTCCGCTGGACTTGC	Sequencing the *rbc* operon insertion site at NS2. F08 was also used to sequence the native *rbc* operon deletion site.
R08	TGTACTCGATTTGTGCAGCG	
F09	ACCACCTTGGCAAAATGGTG	qPCR analysis of the expression of *rbcL* (ID: Synpcc7942_1426) that encodes the RuBisCO large subunit.
R09	TTTGTCGCCTTCCAGTTTGC	
F11	ATTACCTGCGCGACAACATG	qPCR analysis of the expression of *secA* (reference gene, ID: Synpcc7942_0289) that encodes the preprotein translocase subunit SecA.
R11	TGCCCGCATGTATTTTTCGC	
F13	GGAGCCTCTGACTATCGCTGGGGGAG	Linearization of pSyn02 without the *rbcL* coding sequence.
R13	GTCGTCTCTCCCTAGAGATATG	
F14	AAACCGGCGAAATCAAAGGC	qPCR analysis of the expression of AncIB *rbcL* that encodes the AncIB RuBisCO large subunit.
R14	CGTGCATGATGATGGGCATG	
F15	TGTTTGAAGAAGGCAGCGTG	Sequencing the internal region of AncIB *rbcL* inserted at NS2.
F16	CCAAATGGTGCCGCGATAAC	

750 [a]F - forward, R – reverse