Title: Origin of chaperone dependence and assembly complexity in Rubisco's biogenesis

Authors: Jediael Z. Y. Ng,[1,4] Dennis Wiens,[1,4] Luca Schulz,[1,2] Alisa Hergenröder,[1,4] Andreas M. Küffner,[2†] Christin Geil,[1,4] Tobias J. Erb,[2,3] Georg K. A. Hochberg[1,3,4*]

Affiliations:

[1] Evolutionary Biochemistry, Max Planck Institute for Terrestrial Microbiology; Marburg, 35043, Germany

[2] Department of Biochemistry and Synthetic Metabolism, Max Planck Institute for Terrestrial Microbiology; Marburg, 35043, Germany

[3] Center for Synthetic Microbiology (SYNMIKRO), Philipps-Universität Marburg; Marburg, 35043, Germany

[4] Department of Biology, Philipps-Universität Marburg; Marburg, 35043, Germany

† Present address: Max Planck Foundation Research Group, Max Planck Institute for Multidisciplinary Sciences; Göttingen, 37077, Germany

Correspondence: hochberg@staff.uni-marburg.de

Abstract: Molecular chaperones assist with the folding and assembly of protein clients. Consequently, they are essential to diverse cellular functions. In most aerobic photosynthetic organisms such as β-cyanobacteria and all plants, the key CO_2-fixing enzyme Rubisco strictly relies on dedicated chaperones for its assembly. Why these chaperones were recruited is unclear, as they are absent and not required in other prokaryotic autotrophs. By retracing the evolution of Rubisco-chaperone interactions, we find that Rubisco's dependence on chaperones evolved neutrally and not by adaptive optimization, and that this may be reversed without compromising catalytic function. Our work shows how chaperones are broad modifiers of their clients' sequence spaces that can become totally essential through non-adaptive processes.

Main text:

Introduction

Proteins must fold and assemble correctly in a crowded, chaotic cellular environment. To accomplish this, many proteins depend entirely on molecular chaperones, which provide transient assistance to their clients during folding or assembly (*1*). How proteins come to rely on chaperones is not well understood. Chaperone requirements can vary within protein families, suggesting that the need for chaperone assistance during folding or assembly is not always an unavoidable feature of a protein's fold, but can be acquired or lost in the course of history. Evolutionary theory suggests that the acquisition of chaperone dependence (i.e. the inability to fold or assemble without chaperones) may be indirectly beneficial to proteins, through so-called adaptive buffering: Chaperones may allow the fixation of mutations in their clients that are functionally beneficial, but are also destabilizing and intolerable without a chaperone (*2*–*6*). In this theory, improved function of the client comes at the cost of total dependence on a chaperone. An alternative theory is that proteins come to depend on chaperones through the action of a neutral evolutionary ratchet (*7*–*9*): this can happen through mutations that are very detrimental to function on their own, which become neutral in the presence of chaperones (*10*). If such mutations fix through the action of neutral genetic drift, clients can become completely dependent on chaperones without gaining any improvement in their function in return (*11*).

One example of a protein family known to have become dependent on several new chaperones in its history is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Fig. 1A) (*12*–*14*), the enzyme responsible for CO_2 fixation in oxygenic photosynthesis and probably the most abundant protein on Earth (*15*–*17*). The most ancient versions of Form I Rubisco are first folded in the chaperonin cage of the folding chaperone GroEL/ES and then self-assemble without further assistance into either homo-oligomers consisting of eight catalytic large (RbcL), or hetero-oligomers incorporating eight non-catalytic small subunits (RbcS) (*18*–*21*). The latter then became dependent on additional chaperones that help Rubisco's large subunit assemble into complexes after it emerges from the chaperonin cage (Fig. 1A). Some β-cyanobacterial Rubiscos (1B clade) strictly require the action of either of two dedicated chaperones - Raf1 or RbcX - for the assembly of functional Rubisco complexes (*22*–*25*). Another increase in assembly complexity then followed with the addition of assembly chaperones Raf2 and BSD2 on top of Raf1 and RbcX in plants (Fig. 1A) (*26*–*29*).

How and why these chaperones became essential is unclear, but adaptive buffering is a plausible mechanism: Rubisco's history is plagued by a problematic relationship with oxygen, an undesired side substrate that results in the production of a toxic product. Over the course of its history, Rubisco has adapted to rising atmospheric oxygen concentrations by becoming more specific for CO_2 over O_2 (*30*, *31*). Rubisco's evolution is notoriously slow and potentially constrained by trade-offs between its catalytic speed and specificity for CO_2 (*32*–*36*). It thus seems plausible that Rubisco may have become dependent on additional chaperones because they unlocked catalytically beneficial substitutions that would otherwise destabilize the enzyme. This theory, however, has remained untested, because both adaptive buffering and neutral entrenchment would

result in the same indistinguishable phenotype when chaperones are removed in organisms: a complete loss of functional Rubisco assembly.

Here we recapitulate the evolution of Rubisco's dependence on assembly chaperones Raf1 and RbcX using ancestral sequence reconstruction (*9*, *37*). Through the reconstruction of a billion-year-old assembly pathway, we reveal that increased complexity in Rubisco's biogenesis became essential via a neutral evolutionary ratchet. These findings also illuminate the historical trajectory transversed by the biogeochemically important Form I Rubiscos: Our study suggests that modern Rubiscos could be relieved from constraints imposed on them by chaperones, which may ultimately open up new strategies for engineering improved CO_2 fixation in plants.

Resurrection of Rubisco's chaperone network

To investigate how Rubisco-chaperone interactions evolved, we began by inferring the phylogeny of Rubisco's large subunit (LSU) (Fig. 1B, S2A, see methods). We then mapped onto it the occurrence of assembly chaperones within the genomes of extant species. Chaperones Raf1 and RbcX first appeared within β-cyanobacteria. To retrace when assembly chaperones first became essential, we inferred sequences of ancestral Rubiscos that existed before and after the occurrence of chaperones Raf1 and RbcX (Fig. 1B). The phylogeny of each chaperone was also inferred with matching topological constraints as our RbcL tree to obtain synchronized ancestral nodes (Fig. S1). This enabled us to reconstruct the ancient chaperone network that was likely to have co-existed with our Rubisco ancestors within the same ancestral organism as closely as possible (see Methods). This minimizes the potential for artefacts and incompatibility in Rubisco-chaperone pairings, which is a known obstacle to productive assembly (*38*–*41*).

To determine when Rubisco started to rely on chaperone interactions for assembly, we resurrected 4 ancestors (Fig. 1B; AncE, D, C, A; AncB was insoluble, see Supplementary text). Going from the most ancient to the most recent reconstructions: AncE is the last common ancestor (LCA) of Form IA and Form IB Rubisco and would have existed before chaperones appeared in the genomes of β-cyanobacteria. AncD is the LCA of all Form IB Rubiscos, which would have existed in an ancestral β-cyanobacterium that already contained Raf1 and RbcX. AncC corresponds to the last node that unites cyanobacterial Rubiscos and their plastidal descendants, and therefore existed in an ancestral cyanobacterium immediately prior to the endosymbiosis event that birthed chloroplasts. Lastly, AncA corresponds to the LCA of all green algal and plant Rubiscos.

We next characterized the assembly characteristics of our ancestors by co-expressing them with or without their corresponding assembly chaperones within Escherichia coli (E. coli) and examining for assembled Rubisco complexes with native PAGE and immunoblot analysis (Fig. 1C). We also quantified the amount of functional active sites using the radiolabeled transition state analogue 2-^{14}C-carboxy-D-arabinitol-1,5-bisphosphate (^{14}C-CABP), and by assaying for Rubisco activity directly in lysates (Fig. 1C, bottom row). Our oldest ancestor AncE could assemble into functional complex without assembly chaperones (Fig. 1B), consistent with the fact that it predates the

emergence of Raf1 and RbcX. Additionally, co-production of assembly chaperones Raf1 or RbcX boosted AncE's assembly yields. Thus, the initial recruitment of these chaperones was likely beneficial because they probably would have increased the fidelity of Rubisco's assembly pathway. It is important to note that this initially beneficial effect of adding chaperones on its own does not explain why Rubisco later becomes completely unable to assemble without them.

Our next ancestor, AncD, behaved similarly to its predecessor and could still assemble in absence of chaperones (Fig. 1C). Adding Raf1 also bolstered assembly yields, though RbcX no longer did so. Likewise, we saw the same capacity for self-assembly in extant Form I Rubiscos from Synechococcus sp. JA (SynJA) and Gloeobacter violaceus (*Gv*) that are close descendants of AncD (Fig. S2B) (*42*). In contrast, AncC could no longer assemble on its own and strictly required assistance from Raf1 for its assembly (Fig. 1C). This is also consistent with Rubisco from the cyanobacterium Gloeomargarita lithophora (*Gl*) (Fig. S2B), which descends from AncC and represents the closest cyanobacterial relative of plastidal Rubiscos (*43*). Our latest ancestor AncA remained totally dependent on assembly chaperones. Overall, our data show that assembly chaperones were not essential for assembly immediately after they first appeared, and that Rubisco became dependent on assembly chaperones within cyanobacteria sometime between AncD and AncC. This dependence, and a dominant role for Raf1, was then inherited by plastidal Rubiscos.

We next asked if the evolution of chaperone dependence coincided with improved enzyme function. To do so, we purified our ancestors and measured carboxylation kinetics with ribulose-1,5-bisphophate (RuBP) and specificity for CO_2 over O_2 ($S_{C/O}$) (Table S1). From prokaryotic Form 1A/B (AncE) to the plastidal type Form 1B Rubisco (AncA), the Michaelis constant for CO_2 (K_m^{CO2}) decreased 5.7-fold (Fig. 2B). This is coupled with somewhat fluctuating maximal carboxylation rates (k_{cat}^c) between 1.3 s^{-1} to 2.5 s^{-1} (Fig. 2A), resulting in a ~7.5-fold increase in overall carboxylation efficiencies over this interval (Fig. 2C). Likewise, a 1.6-fold improvement in specificity occurs between AncE to AncA (Fig. 2D). Intriguingly, the most significant improvements occurred between AncD and AncC, coinciding with the same interval over which chaperone dependence evolved. The temporal overlap between improved enzyme function and dependence thus hints that adaptive buffering by chaperones could have occurred.

These results were verified against two kinds of uncertainty. First, we resurrected sequences of less likely ancestral Rubiscos from our constrained phylogeny (Fig. S2C), in which the maximum a posteriori amino acid was replaced with the next most plausible amino acid whenever the second state had a posterior probability of at least 0.2. In the second, we tested ancestors inferred using an unconstrained ML RbcL phylogeny (Fig. S2D-E), which would require many horizontal gene transfer (HGT) events to explain the branching order of RbcL within cyanobacteria (Fig. S2A). Our inferences remained robust in both cases: Rubisco is already dependent on chaperones at the node that unites plastid and cyanobacterial Rubiscos. Therefore, we proceeded with ancestors inferred from the constrained phylogeny for the rest of this work.

Genetic basis of dependence on chaperones Raf1/RbcX

Next, we sought to assess if the evolution of chaperone dependence is causally linked to improved Rubisco function. To do this, we first needed to determine which changes in which proteins were responsible for the dependence on chaperones and improved catalysis. We therefore constructed hybrid ancestors containing swapped large and small subunits of AncD and AncC ($AncD_LC_S$, $AncC_LD_S$). Both hybrid ancestors assembled functionally when chaperones were provided (Fig. 2E, S3E-F). However, only the hybrid ancestor comprised of AncD's LSU and AncC's SSU ($AncD_LC_S$) retained the ability to assemble without chaperones, suggesting that historical substitutions in the LSU were responsible for chaperone dependence. We also conducted a similar swap of Rubisco-chaperone-ancestor pairs to exclude changes in Raf1 and RbcX as the cause of dependence. Our observations remained the same even when mis-matched chaperone ancestors were utilized, proving that only changes in Rubisco, and not assembly chaperones, were responsible. In addition, improvements to carboxylation kinetics and $S_{C/O}$ were encoded by substitutions in the LSU (Fig. 2F-I). These results demonstrate that the genetic determinants for both improved function and chaperone dependence lie within the LSU.

To pinpoint causal changes in the LSU, we scrutinized all 20 historical substitutions that occurred between AncD to AncC. These were grouped into three spatial clusters of highly conserved substitutions (cluster 1, 2, 3, comprising 3, 3, and 5 substitutions respectively), as well as a fourth cluster (cluster 0 with 9 substitutions) containing all remaining substitutions (Fig. 3A, S4A). We reverted the amino acids of these clusters in AncC back to their ancestral states in AncD (Fig. 3B). Clusters 0 (rv0) or 1 (rv1) variants remained incapable of assembly without chaperones. Reversion of cluster 3 (rv3) weakly restored assembly in the absence of chaperones, requiring combinations with other clusters ($rv1^+3$, $rv1^+2^+3$) to amplify this effect (Fig. 3C). Remarkably, reverting cluster 2 (rv2) substitutions alone was sufficient to achieve robust rescue of assembly without chaperones (Fig. 3B-C). We further tested whether reversion of the same three substitutions in cluster 2 would work in an extant enzyme. Rubisco from *G. lithophora* (*GL*), which branches after AncC and is sister to all plastid Rubiscos on our tree (*43*), contains all derived cluster 2 substitutions and is dependent on assembly chaperones (Fig. 3E). Reverting only cluster 2 substitutions in *Gl* Rubisco was likewise able to abolish its dependence on chaperones Raf1 and RbcX. Therefore, cluster 2 substitutions, which lie within the central solvent pore of Rubisco (Fig. 3D), remain as the key determinant for chaperone dependence even in extant descendants.

Entrenchment of Raf1/RbcX interaction was catalytically neutral

We now had the means to examine if Rubisco's reliance on chaperones came hand-in-hand with improved catalytic function, as predicted by adaptive buffering. To do this, we compared the enzymatic parameters of AncC (the better, chaperone-dependent enzyme) and AncD (the inferior, but chaperone-independent ancestor) to AncC rv2. Remarkably, the chaperone-independent reversion was as good a catalyst as AncC: we observed no significant changes in $S_{C/O}$ and carboxylation kinetics of AncC rv2 compared to AncC (Fig. 3F-I). To verify that this is not an artefact of our ancestral sequences, we performed the same experiment on the chaperone-independent rv2

205 variant of the extant *G. lithophora* enzyme (*Gl* rv2). Again, $S_{C/O}$ was unaffected in the
206 rv2 variant of the extant *G. lithophora* enzyme (*Gl* rv2), though k_{cat}^{c} and K_m^{CO2} increased
207 slightly (Fig. 3F-I). We also did not observe any significant differences in the Michaelis
208 constant for RuBP (K_m^{RuBP}) or apparent rate of inhibitor release (k_{obs}) of reverted
209 variants (Fig. S4F-G). We further tested whether our reversions made the enzyme less
210 stable. To do this, we exposed our variant to elevated temperatures and measured its
211 remaining activity. We found no difference between the reverted variant and its wild-
212 type counterpart (Fig. S4D-E). Furthermore, we also did not find any significant
213 differences in Rubisco assembly yields of the rv2 variant if chaperones were co-
214 produced, indicating that chaperone-dependent variants do not gain added benefits to
215 assembly fidelity from the presence of assembly chaperones (Fig. 3C).

216 These data imply that entrenchment and catalytic improvements are genetically
217 separable from each other. We next wondered whether we could determine which
218 occurred first: catalytic improvement or chaperone dependence. On our constrained
219 topology there are two nodes between AncD and AncC, which we term AncD1 and
220 AncD2. These proteins already contain cluster 1 and 3 substitutions, but not cluster 2
221 substitutions, and are not yet dependent on chaperones (Fig. 3A, S3G, S4B-C). If
222 clusters 1 and 3 were responsible for improved catalysis, entrenchment would have
223 come after these improvements. We verified this by experimentally resurrecting AncD2,
224 as well as AncD with clusters 1 and 3 substituted to their states in AncC. Both proteins
225 exhibited elevated $S_{C/O}$ values relative to AncD (Table S1), implying that catalytic
226 improvement at least partially preceded the entrenchment of chaperones. Together,
227 these data show that entrenchment of the assembly chaperones Raf1 and RbcX was
228 neutral on the level of the catalytic parameters of Rubisco. These inferences are robust
229 to the topological uncertainty in our tree: causal substitutions and their impact on
230 catalysis remain unchanged even when examined on the unconstrained maximum
231 likelihood topology (Fig. S2D ,S3A-D).

232

233 ***An epistatic rachet keeps assembly chaperones essential***

234 Within green algae and plants, Raf1 remains essential for Rubisco's assembly and
235 enhances photosynthetic growth when overexpressed (*29*, *44*, *45*). However, if only
236 three catalytically neutral reversions are enough to cure Rubisco of this dependence,
237 why are chaperones still necessary in the plastid lineage to this day (*23*, *27*, *29*, *38*)? To
238 answer this, we turned to the youngest of our Rubisco ancestors, AncA. This enzyme
239 represents the LCA of all green algal and plant Rubiscos and is dependent on Raf1
240 (Fig. 1C). AncA, as well as most other plastid Rubiscos, maintains conservation of the
241 same derived rv2 states as AncC (Fig. S5A). We therefore asked if reverting these
242 substitutions would restore AncA's ability to assemble without Raf1. AncA rv2, however,
243 remained incapable of assembly without chaperones (Fig. 4B, S5E-F). This implies
244 further changes must have accumulated which deepened dependence on assembly
245 chaperones.

246 To identify what triggered deeper dependence, we reasoned that the elongated βA-βB
247 loop of plant SSUs, which extends into the central solvent channel of Rubisco, may
248 coordinate with cluster 2 entrenching elements, which may influence the effect of
249 reversions at these sites (Fig. S5B). However, hybrid Rubiscos containing AncC's SSU,

which lacks this elongated lool, remain insoluble without chaperones, excluding a role for the SSU in causing deeper dependence (Fig. S5C). Another plausible contributor could be the folding chaperone: Whereas the simpler prokaryotic GroEL chaperonin suffices for prokaryotic AncC (Fig. S5D; lanes 1 and 2), AncA requires the more structurally complex plastidal chaperonins CPN60α/β (Fig. S5D; lanes 3 and 4). Utilizing GroEL instead of CPN60 also failed to restore chaperone independent assembly of AncA rv2 (Fig. S5F-G, lanes 3). Therefore, we searched once more amongst the 39 historical substitutions that emerged between AncC and AncA, reverting subsets of these in an AncA rv2 background (Fig. S5E, S6A). Three clusters near cluster 2 residues (Fig. 4D) were found to confer weak rescue from chaperones as determined by assembly band intensities (Fig. S5E-F), but rescue was minimal and could be eliminated when co-expression was carried out at slightly elevated temperatures (30°C instead of 28°C) (Fig. S6B-C). In contrast, combining these subsets (AncA rvC; reversion of clusters $2^{+}4^{+}5^{+}6$, 11 total substitutions) resulted in robust self-assembly when eukaryotic folding chaperonins were utilized (Fig. S5F). The effect of these further reversions was contingent on also reverting the original cluster 2 sites, since AncA variants without rv2 reversions still required chaperones (Fig. S5F, lanes 6 and 9). These data show that further substitutions entrenched the need for assembly chaperones, and that this was reliant on the original dependence-causing substitutions (also see Supplementary text).

The assembly-competent AncA rvC exhibited compromised catalytic rates compared to AncA (Fig. 4E-H). We therefore identified 3 additional reversions in the inter-dimer interface of the LSU (cluster 11; Fig. 4D) that together with our prior reversions, resulted in a variant (AncA rvCC) with catalytic properties indistinguishable from AncA (Fig. 4E-H), but which could assemble independent of chaperones (Fig. 4C; lanes 5 and 6). Unlike wild-type AncA and our other reversion constructs of it (Fig. 4C lanes 3 and 4, Fig. S5G lanes 5 and 7), AncA rvCC could also fold using GroEL instead of the more complex CPN60 from plants (Fig. 4C; lanes 7 and 8). This demonstrates that it is possible to rid a plastidal Rubisco of its derived folding and assembly chaperone requirements. This prompted us to ask whether we could also rid Rubiscos from *Arabidopsis thaliana* of their need for these chaperones using the same reversions. This experiment, however, failed (Fig. S6D-E), implying that even more entrenching states have accumulated along the lineage to plants since AncA, potentially due to their additional assembly requirements like reliance on BSD2 (*27*).

Together, these results demonstrate that assembly chaperones remain essential because additional entrenching states accumulated along the lineage to plants, and because these states have been ratcheted in through other changes in the protein that makes their reversion functionally unfavourable. At the level of Rubisco's catalytic function, these additional entrenching substitutions confer no functional advantage. We can at present only speculate about the structural mechanism of Rubisco's dependence on assembly chaperones. Sites involved in causing the dependence on chaperones cluster at the surface facing the inside of the pore formed by Rubisco complexes, far from known chaperone binding-sites on Rubisco. A plausible mechanism would be that dependence-causing substitutions either make this surface aggregation-prone or unstable due to the loss of specific inter-residue contacts that stabilize the native fold (Fig. S4A, S7). This would pose a problem only during assembly, because this surface

is buried and shielded inside Rubisco complexes once assembly is complete. Assembly chaperones, which facilitate the formation of RbcL octamers during assembly, may therefore allow Rubisco to avoid aggregation by minimizing the time it spends with this surface exposed to the cellular environment.

Discussion

In this work we discover how Rubisco became dependent on its assembly chaperones Raf 1 and RbcX. Addition of these chaperones allowed Rubisco to accumulate substitutions that are neutral when chaperones are present, but highly deleterious when they are not. Though the initial gain of these chaperones was likely adaptive because it boosted assembly yields, the changes that made them essential later were likely non-adaptive, consistent with theoretical models of the evolution multi-layered biological systems (Fig. S8) (*46*, *47*).

We note that while this entrenchment was neutral at the level of Rubisco's catalytic function, a chaperone dependence could be useful, for example, as a means of regulating Rubisco's abundance in vivo. Such adaptative use of a chaperone dependence itself could also evolve long after the initial dependence arose, as an evolutionary spandrel (*48*). But even if the dependence itself is somehow useful, this cannot explain why Rubisco accumulated substitutions later in in its history that deepened the genetic dependence (i.e. made it more difficult to revert) in an enzyme that could already not assemble at all without chaperones anymore. At present, our only explanation for this process is that the new chaperones unlocked a large and previously inaccessible part of Rubisco's sequence space, into which it has since been drifting. This is reminiscent of the SSU's role earlier in Rubisco's history (*20*, *49*). Unlike the SSU, however, Rubisco's assembly chaperones appear not to have unlocked substitutions that improved Rubisco's enzymatic function. This is clearly at odds with the theory that chaperones are generally adaptive buffering agents (*2*, *5*) and raises the possibility that other protein-chaperone interactions are also the result of similar neutral ratchets.

Rather than adaptive evolutionary buffers, chaperones may be better understood as broad modifiers of their clients' accessible sequence spaces, which sometimes serve as adaptive buffers and sometimes become essential neutrally. The potential for neutral entrenchment may add a selfish dimension to chaperone evolution: Chaperones can ensure their preservation by becoming essential for other proteins with important functions, even if this doesn't improve their clients' functions. The more promiscuous the chaperone, the more certain it is that at least one important protein forgets how to reach its functional conformation without the chaperone. Conversely the more important a potential chaperone client's function is to an organism, the more effectively purifying selection will preserve any chaperone that becomes essential to that client.

As we show, we can still undo such purposeless increases in cellular complexity by reversing the genetic changes that made these increases essential in history. It may thus be possible to rid modern genomes of any useless cellular bureaucracy that evolution has allowed to accumulate.

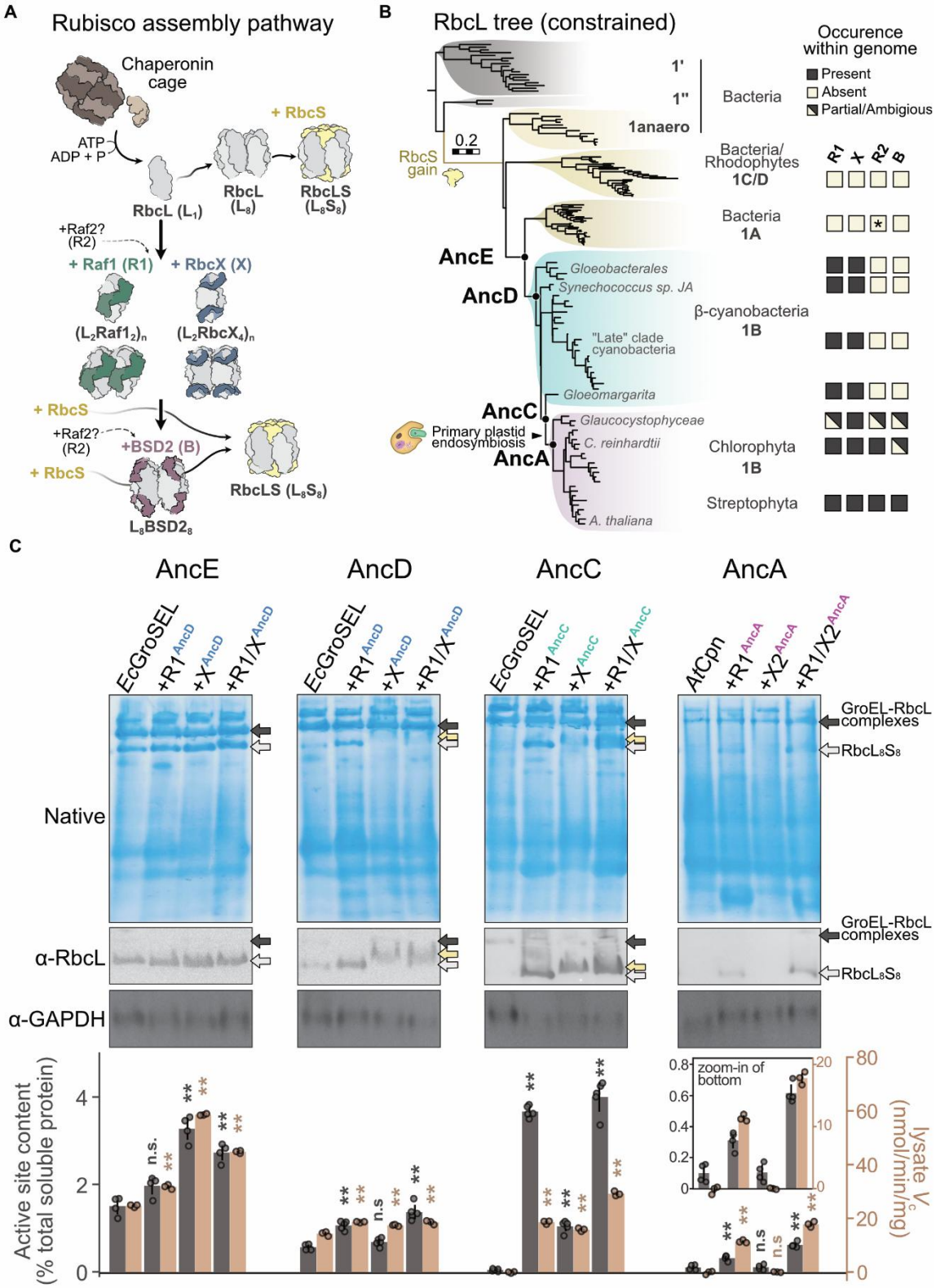

Figure 1. Evolution of Rubisco and assembly chaperones.

(A) Biogenesis pathway in Form I Rubiscos. Early branching Form I Rubiscos may co-assemble directly with the small subunit (RbcS). Reliance on assembly chaperones

Raf1 (R1) and RbcX (X) before RbcS addition first arose in β-cyanobacteria. Requirement of additional chaperones factors Raf2 (R2) and BSD2 (B) emerged later within green plastids. **(B)** Constrained Maximum Likelihood phylogeny of RbcL based on known species relationships. Ancestral Rubiscos at indicated nodes were resurrected and characterized. Clades according to their biogenesis requirements and final assemblies are coloured as per (A), with their designated Form 1 type and host organisms indicated on the right. Right column panel: The presence or absence of chaperones within the genomes of extant organisms within corresponding clades shown to the left in the tree. A putative Raf2-like homolog of a similar fold is found in α-Cyanobacteria as indicated by the asterisk. Partial or ambiguous classifications were assigned due to poor genome coverage or hits with low sequence identities. **(C)** Native PAGE and western blot analysis of ancestral Rubisco-chaperone co-expression combinations. White arrows indicate L_8S_8 assembly bands, black arrows indicate chaperonin-bound LSU complexes, and yellow arrows for plausible chaperone-bound LSU complexes. Bottom row left axis/ grey bars: The amount of Rubisco active sites shown as percentages relative to total soluble protein obtained from lysate extracts of co-expressed combinations in the columns above. Error bars represent SDs. Significance as tested by one-way ANOVA (ANalysis Of VAriance) with post-hoc Tukey HSD (Honestly Significant Difference) test. *n.s.*, nonsignificant; $^{**}P < 0.01$. Data are from at least three technical replicates as indicated by coloured circles. Bottom row right axis: Rubisco activities measured in soluble lysate extracts from expression conditions indicated above at fixed CO_2 (~200 μM) concentrations.

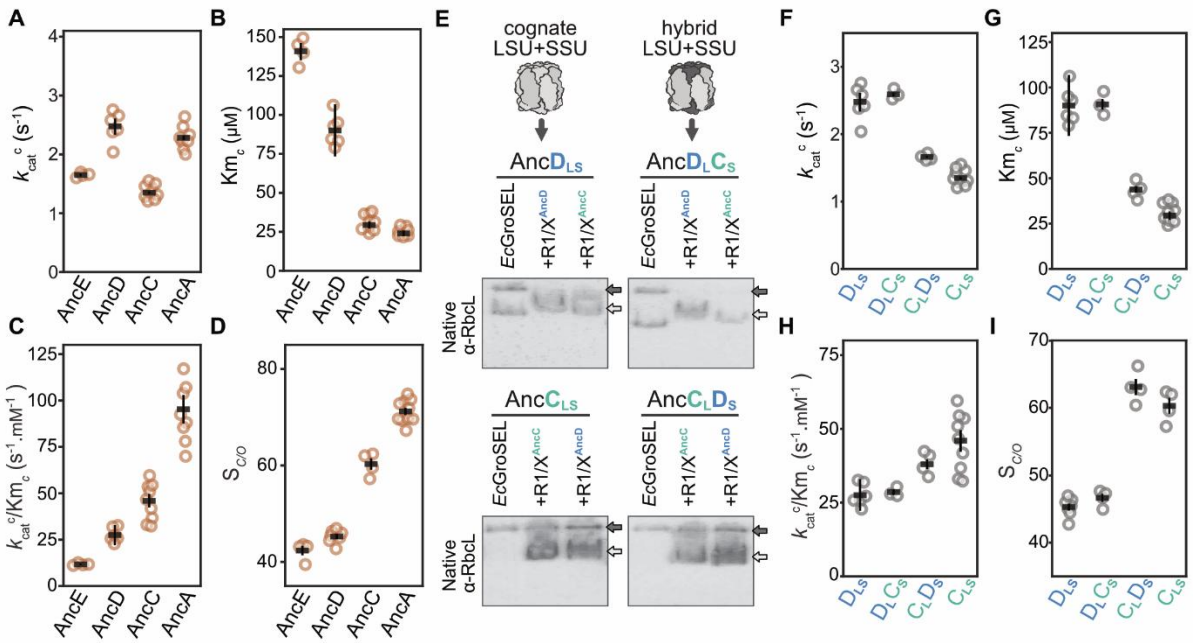

Figure 2. Changes in the LSU determine catalytic properties and assembly requirements.

(A-D) Kinetic properties of ancestral Rubiscos. **(A)** Carboxylation rates, **(B)** Michaelis constant for CO_2, **(C)** carboxylation efficiency, and **(D)** substrate specificity. **(E)** Native blot analysis of wild type and hybrid ancestral Rubisco assembly with matching or non-matching chaperone pairs. Neither the SSU nor chaperone identities influence the requirement for assembly chaperones. **(F-I)** Comparisons of kinetic parameters exhibited by hybrid AncD/C Rubiscos against wild-type enzymes. LSU identities are observed to dictate kinetic characteristics of hybrid enzymes. For all kinetic data, catalytic parameters extracted from individual Michaelis-Menten fits of technical replicates are shown as hollow circles, with globally fitted parameters denoted as a horizontal bar. Error bars represent SEs of the global fit. Values of global fits and replicate numbers are further indicated in table. S1 and kinetic curve fits shown in fig. S17.

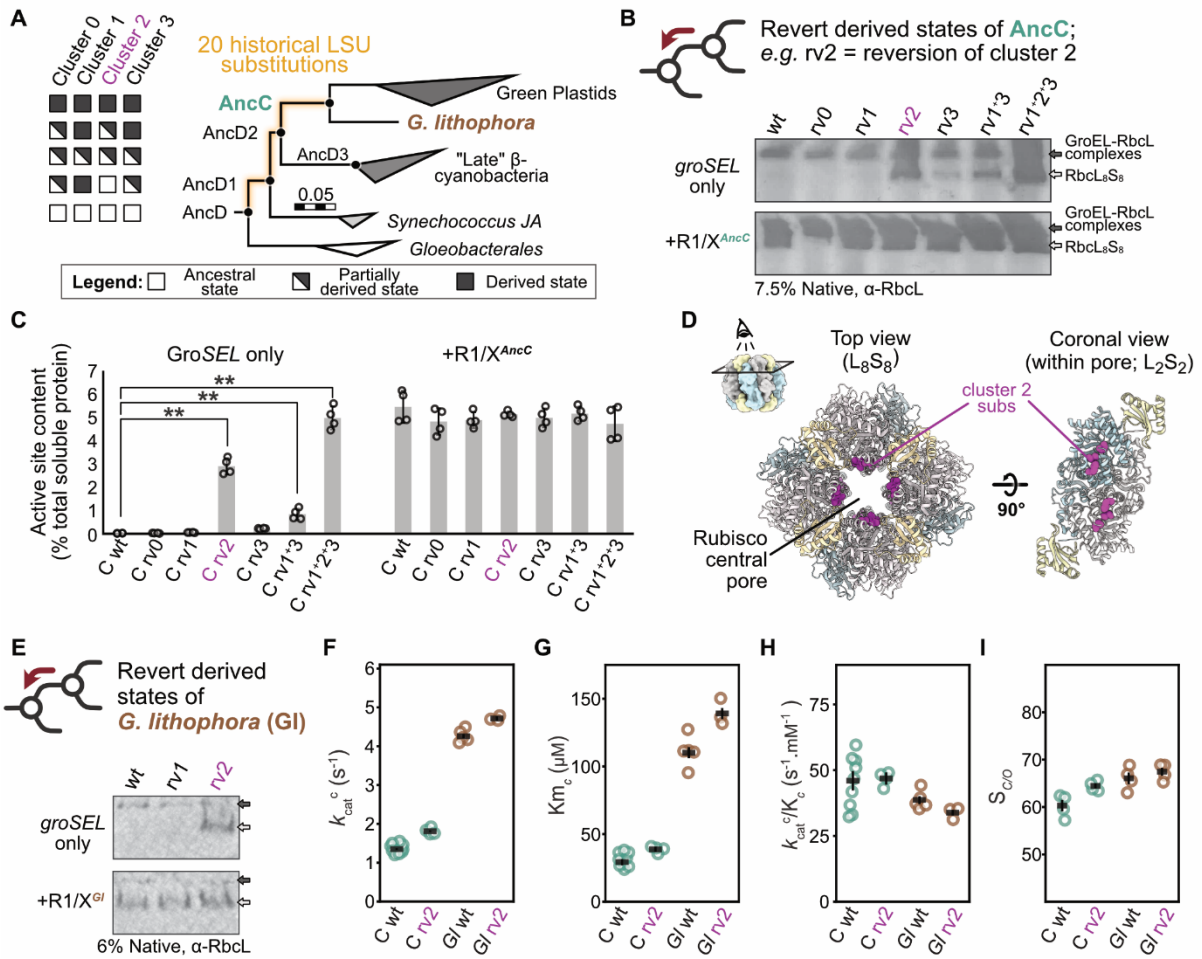

Figure 3. Dependence on chaperones is decoupled from improved catalysis.

(A) Reduced phylogeny of RbcL between AncD and AncC. The presence or absence of derived substitutions in each cluster is shown on the left, with each row corresponding to the identity of ancestral nodes to the right. **(B)** Native blot analysis of AncC reversion variants carrying ancestral amino acid states at indicated clusters, co-expressed with (bottom row) and without (top row) chaperones. rv: reversion mutant. Black arrows indicate probable chaperonin bound RbcL complexes, White arrows indicate assembled Rubiscos. **(C)** The effect of cluster reversions on assembly yields as in (B), quantified by active site content measurements. Only cluster 2 reversion variant rescues chaperone independent assembly. Significance as tested by one-way ANOVA with post-hoc Tukey HSD test. $^{**}P < 0.01$. **(D)** Location of cluster 2 (α4-β5) substitutions in the central solvent channel of Rubisco. On the right, a single L_2S_2 dimer is visualized from the coronal plane within the pore. **(E)** Native blot analysis of wild type and reverted mutant Rubisco from the extant cyanobacterium *G. lithophora*. Reversion of cluster 2 but not cluster 1 states rescue self-assembly. **(F-I)** Reversion of cluster 2 substitutions in AncC (wt and rv2 in green) and *G. lithophora* (wt and rv2 in brown) Rubisco do not perturb catalysis. Kinetic parameters extracted from individual fits of technical replicates are shown as hollow circles, with global parameters denoted by a horizontal bar. Error bars represent SEs of the global fit. Exact values of global fits and replicate numbers are indicated in table S1 and kinetic fits in fig. S17.

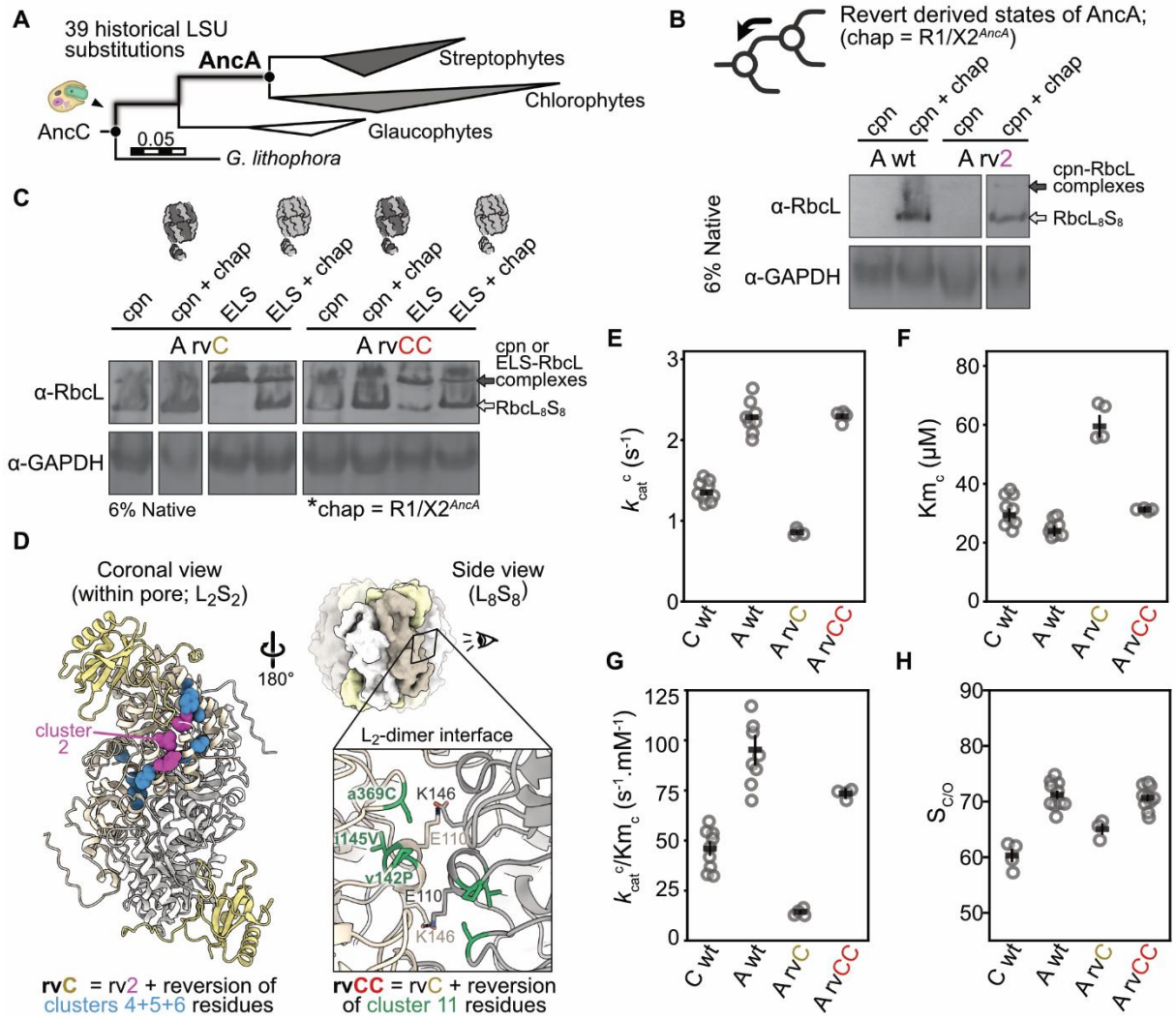

Figure 4. Persistence of chaperone interactions in the green plastid lineage.

(A) Reduced phylogeny of RbcL between AncC and AncA **(B)** Native blot analysis of AncA wild type and reverted variants as indicated either in the absence or presence of assembly chaperones, **(C)** or with prokaryotic (ELS) or eukaryotic (cpn) chaperonin co-production. cpn: CPN60. chap: Raf1 and RbcX. **(D)** Schematic representation of the location of chaperone and complex plastid type chaperonin entrenching substitutions modelled on AncA's Rubisco as predicted by Alphafold 3 (*50*). Clusters 4 and 6 surround initial cluster 2 entrenching substitutions, while cluster 5 substitutions occur in proximity to the active site of Rubisco, as additionally depicted in Fig. S6A. Right inset: Cluster 11 substitutions are found at the inter LSU-dimer interface supporting residues key to inter-LSU dimer interactions. **(F-I)** Kinetic parameters exhibited by assembly competent AncA Rubisco variants. Kinetic parameters extracted from individual fits are shown as hollow circles, with global parameters depicted as black horizontal bars. Errors bars denote SEs of the global fit. Replicate numbers are indicated in table. S1 and fitted kinetic curves in fig. S17.

References and Notes

1. F. U. Hartl, M. Hayer-Hartl, Converging concepts of protein folding in vitro and in vivo. *Nature Structural & Molecular Biology* **16**, 574–581 (2009).

2. S. L. Rutherford, S. Lindquist, Hsp90 as a capacitor for morphological evolution. *Nature* **396**, 336–342 (1998).

3. T. A. Williams, M. A. Fares, The Effect of Chaperonin Buffering on Protein Evolution. *Genome Biology and Evolution* **2**, 609–619 (2010).

4. M. A. Fares, M. X. Ruiz-González, A. Moya, S. F. Elena, E. Barrio, GroEL buffers against deleterious mutations. *Nature* **417**, 398–398 (2002).

5. N. Tokuriki, D. S. Tawfik, Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).

6. J. Aguilar-Rodríguez, B. Sabater-Muñoz, R. Montagud-Martínez, V. Berlanga, D. Alvarez-Ponce, A. Wagner, M. A. Fares, The Molecular Chaperone DnaK Is a Source of Mutational Robustness. *Genome Biology and Evolution* **8**, 2979–2991 (2016).

7. A. Stoltzfus, On the Possibility of Constructive Neutral Evolution. *Journal of Molecular Evolution* **49**, 169–181 (1999).

8. S. A. Muñoz-Gómez, G. Bilolikar, J. G. Wideman, K. Geiler-Samerotte, Constructive Neutral Evolution 20 Years Later. *Journal of Molecular Evolution* **89**, 172–182 (2021).

9. G. K. A. Hochberg, Y. Liu, E. G. Marklund, B. P. H. Metzger, A. Laganowsky, J. W. Thornton, A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503–508 (2020).

10. L. Schulz, F. L. Sendker, G. K. A. Hochberg, Non-adaptive complexity and biochemical function. *Current Opinion in Structural Biology* **73**, 102339 (2022).

11. T. D. P. Brunet, W. F. Doolittle, J. P. Bielawski, The role of purifying selection in the origin and maintenance of complex function. *Studies in History and Philosophy of Science Part A* **87**, 125–135 (2021).

12. M. Hayer-Hartl, F. U. Hartl, Chaperone Machineries of Rubisco – The Most Abundant Enzyme. *Trends in Biochemical Sciences* **45**, 748–763 (2020).

13. A. Vitlin Gruber, L. Feiz, Rubisco Assembly in the Chloroplast. *Frontiers in Molecular Biosciences* **Volume 5-2018** (2018).

14. P. Durão, H. Aigner, P. Nagy, O. Mueller-Cajar, F. U. Hartl, M. Hayer-Hartl, Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nature Chemical Biology* **11**, 148–155 (2015).

15. N. Prywes, N. R. Phillips, O. T. Tuck, L. E. Valentin-Alvarado, D. F. Savage, Rubisco Function, Evolution, and Engineering, *Annual Review of Biochemistry*. **92** (2023)pp. 385–410.

16. I. Andersson, A. Backlund, Structure and function of Rubisco. *Plant Physiology and Biochemistry* **46**, 275–291 (2008).

17. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* **115**, 6506–6511 (2018).

18. P. Goloubinoff, A. A. Gatenby, G. H. Lorimer, GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose bisphosphate carboxylase oligomers in Escherichia coli. *Nature* **337**, 44–47 (1989).

19. D. M. Banda, J. H. Pereira, A. K. Liu, D. J. Orr, M. Hammel, C. He, M. A. J. Parry, E. Carmo-Silva, P. D. Adams, J. F. Banfield, P. M. Shih, Novel bacterial clade reveals origin of form I Rubisco. *Nature Plants* **6**, 1158–1166 (2020).

20. L. Schulz, Z. Guo, J. Zarzycki, W. Steinchen, J. M. Schuller, T. Heimerl, S. Prinz, O. Mueller-Cajar, T. J. Erb, G. K. A. Hochberg, Evolution of increased complexity and specificity at the dawn of form I Rubiscos. *Science* **378**, 155–160 (2022).

21. A. K. Liu, B. Kaeser, L. Chen, J. West-Roberts, L. J. Taylor-Kearney, A. Lavy, D. Günzing, W.-J. Li, M. Hammel, E. Nogales, J. F. Banfield, P. M. Shih, Deep-branching evolutionary intermediates reveal structural origins of form I rubisco. *Current Biology* **33**, 5316-5325.e3 (2023).

22. C. Liu, A. L. Young, A. Starling-Windhof, A. Bracher, S. Saschenbrecker, B. V. Rao, K. V. Rao, O. Berninghausen, T. Mielke, F. U. Hartl, R. Beckmann, M. Hayer-Hartl, Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* **463**, 197–202 (2010).

23. L. Feiz, R. Williams-Carrier, K. Wostrikoff, S. Belcher, A. Barkan, D. B. Stern, Ribulose-1,5-Bis-Phosphate Carboxylase/Oxygenase Accumulation Factor1 Is Required for Holoenzyme Assembly in Maize. *The Plant Cell* **24**, 3435–3446 (2012).

24. F. Huang, W.-W. Kong, Y. Sun, T. Chen, G. F. Dykes, Y.-L. Jiang, L.-N. Liu, Rubisco accumulation factor 1 (Raf1) plays essential roles in mediating Rubisco assembly and carboxysome biogenesis. *Proceedings of the National Academy of Sciences* **117**, 17418–17428 (2020).

25. D. Emlyn-Jones, F. J. Woodger, G. D. Price, S. M. Whitney, RbcX Can Function as a Rubisco Chaperonin, But is Non-Essential in Synechococcus PCC7942. *Plant and Cell Physiology* **47**, 1630–1640 (2006).

26. L. Feiz, R. Williams-Carrier, S. Belcher, M. Montano, A. Barkan, D. B. Stern, A protein with an inactive pterin-4a-carbinolamine dehydratase domain is required for Rubisco biogenesis in plants. *The Plant Journal* **80**, 862–869 (2014).

27. H. Aigner, R. H. Wilson, A. Bracher, L. Calisse, J. Y. Bhat, F. U. Hartl, M. Hayer-Hartl, Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2. *Science* **358**, 1272–1278 (2017).

28. B. Conlan, R. Birch, C. Kelso, S. Holland, A. P. De Souza, S. P. Long, J. L. Beck, S. M. Whitney, BSD2 is a Rubisco-specific assembly chaperone, forms intermediary hetero-oligomeric complexes, and is nonlimiting to growth in tobacco. *Plant, Cell & Environment* **42**, 1287–1301 (2019).

29. Z. G. Oh, T. A. Robison, D. H. Loh, W. S. L. Ang, J. Z. Y. Ng, F.-W. Li, L. H. Gunn, Unique biogenesis and kinetics of hornwort Rubiscos revealed by synthetic biology systems. *Molecular Plant* **17**, 1833–1849 (2024).

30. F. R. Tabita, S. Satagopan, T. E. Hanson, N. E. Kreel, S. S. Scott, Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of Experimental Botany* **59**, 1515–1524 (2008).

31. T. J. Erb, J. Zarzycki, A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Current Opinion in Biotechnology* **49**, 100–107 (2018).

32. J. W. Bouvier, D. M. Emms, S. Kelly, Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants. *Proceedings of the National Academy of Sciences* **121**, e2321050121 (2024).

33. J. W. Bouvier, D. M. Emms, T. Rhodes, J. S. Bolton, A. Brasnett, A. Eddershaw, J. R. Nielsen, A. Unitt, S. M. Whitney, S. Kelly, Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off. *Molecular Biology and Evolution* **38**, 2880–2896 (2021).

34. A. I. Flamholz, N. Prywes, U. Moran, D. Davidi, Y. M. Bar-On, L. M. Oltrogge, R. Alves, D. Savage, R. Milo, Revisiting Trade-offs between Rubisco Kinetic Parameters. *Biochemistry* **58**, 3365–3376 (2019).

35. G. G. B. Tcherkez, G. D. Farquhar, T. J. Andrews, Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proceedings of the National Academy of Sciences* **103**, 7246–7251 (2006).

36. B. de Pins, C. Malbranke, J. Jabłońska, A. Shmuel, I. Sharon, A.-F. Bitbol, O. Mueller-Cajar, E. Noor, R. Milo, Rubisco is slow across the tree of life. *bioRxiv*, 2025.01.19.633714 (2025).

37. M. J. Harms, J. W. Thornton, Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature Reviews Genetics* **14**, 559–571 (2013).

38. J. Archer, M. Kathpalia, B. Lee, S. Li, T. Wang, Effects of chaperone selectivity on the assembly of plant Rubisco orthologs in E. coli. *Journal of Experimental Botany* **76**, 2809–2820 (2025).

39. S. Buck, T. Rhodes, M. Gionfriddo, T. Skinner, D. Yuan, R. Birch, M. V. Kapralov, S. M. Whitney, Escherichia coli expressing chloroplast chaperones as a proxy to test heterologous Rubisco production in leaves. *Journal of Experimental Botany* **74**, 664–676 (2023).

40. S. M. Whitney, P. Baldet, G. S. Hudson, T. J. Andrews, Form I Rubiscos from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *The Plant Journal* **26**, 535–547 (2001).

41. M. T. Lin, M. R. Hanson, Red algal Rubisco fails to accumulate in transplastomic tobacco expressing Griffithsia monilis RbcL and RbcS genes. *Plant Direct* **2**, e00045 (2018).

42. K. R. Moore, C. Magnabosco, L. Momper, D. A. Gold, T. Bosak, G. P. Fournier, An Expanded Ribosomal Phylogeny of Cyanobacteria Supports a Deep Placement of Plastids. *Frontiers in Microbiology* **Volume 10-2019** (2019).

43. R. I. Ponce-Toledo, P. Deschamps, P. López-García, Y. Zivanovic, K. Benzerara, D. Moreira, An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Current Biology* **27**, 386–391 (2017).

44. S. M. Whitney, R. Birch, C. Kelso, J. L. Beck, M. V. Kapralov, Improving recombinant Rubisco biogenesis, plant photosynthesis and growth by coexpressing its ancillary RAF1 chaperone. *Proceedings of the National Academy of Sciences* **112**, 3564–3569 (2015).

45. C. E. Salesse-Smith, R. E. Sharwood, F. A. Busch, J. Kromdijk, V. Bardal, D. B. Stern, Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nature Plants* **4**, 802–810 (2018).

46. M. Lynch, Evolution of the mutation rate. *Trends in Genetics* **26**, 345–352 (2010).

47. W. Sung, M. S. Ackerman, S. F. Miller, T. G. Doak, M. Lynch, Drift-barrier hypothesis and mutation-rate evolution. *Proceedings of the National Academy of Sciences* **109**, 18488–18492 (2012).

48. S. J. Gould, R. C. Lewontin, The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **205**, 581–598 (1997).

49. L. Schulz, J. Zarzycki, W. Steinchen, G. K. A. Hochberg, T. J. Erb, Layered entrenchment maintains essentiality in the evolution of Form I Rubisco complexes. *The EMBO Journal* **44**, 269–280 (2025).

50. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

51. J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, M. Porsch, M. Quint, S. A. Rensing, D. E. Soltis, P. S. Soltis, D. W. Stevenson, K. K. Ullrich, N. J. Wickett, L. DeGironimo, P. P. Edger, I. E. Jordon-Thaden, S. Joya, T. Liu, B. Melkonian, N. W. Miles, L. Pokorny, C. Quigley, P. Thomas, J. C. Villarreal, M. M. Augustin, M. D. Barrett, R. S. Baucom, D. J. Beerling, R. M. Benstein, E. Biffin, S. F. Brockington, D. O. Burge, J. N. Burris, K. P. Burris, V. Burtet-Sarramegna, A. L. Caicedo, S. B. Cannon, Z. Çebi, Y. Chang, C. Chater, J. M. Cheeseman, T. Chen, N. D. Clarke, H. Clayton, S. Covshoff, B. J. Crandall-Stotler, H. Cross, C. W. dePamphilis, J. P. Der, R. Determann, R. C. Dickson, V. S. Di Stilio, S. Ellis, E. Fast, N. Feja, K. J. Field, D. A. Filatov, P. M. Finnegan, S. K. Floyd, B. Fogliani, N. García, G. Gâteblé, G. T. Godden, F. (Qi Y. Goh, S. Greiner, A. Harkess, J. M. Heaney, K. E. Helliwell, K. Heyduk, J. M. Hibberd, R. G. J. Hodel, P. M. Hollingsworth, M. T. J. Johnson, R. Jost, B. Joyce, M. V. Kapralov, E. Kazamia, E. A. Kellogg, M. A. Koch, M. Von Konrat, K. Könyves, T. M. Kutchan, V. Lam, A. Larsson, A. R. Leitch, R. Lentz, F.-W. Li, A. J. Lowe, M. Ludwig, P. S. Manos, E. Mavrodiev, M. K. McCormick, M. McKain, T. McLellan, J. R. McNeal, R. E. Miller, M. N. Nelson, Y. Peng, P. Ralph, D. Real, C. W. Riggins, M. Ruhsam, R. F. Sage, A. K. Sakai, M. Scascitella, E. E. Schilling, E.-M. Schlösser, H. Sederoff, S. Servick, E. B. Sessa, A. J. Shaw, S. W. Shaw, E. M. Sigel, C. Skema, A. G. Smith, A. Smithson, C. N. Stewart, J. R. Stinchcombe, P. Szövényi, J. A. Tate, H. Tiebel, D. Trapnell, M. Villegente, C.-N. Wang, S. G. Weller, M. Wenzel, S. Weststrand, J. H. Westwood, D. F. Whigham, S. Wu, A. S. Wulff, Y. Yang, D. Zhu, C. Zhuang, J. Zuidof, M. W. Chase, J. C. Pires, C. J. Rothfels, J. Yu, C. Chen, L. Chen, S. Cheng, J. Li, R. Li, X. Li, H. Lu, Y. Ou, X. Sun, X. Tan, J. Tang, Z. Tian, F. Wang, J. Wang, X. Wei, X. Xu, Z. Yan, F. Yang, X. Zhong, F. Zhou, Y. Zhu, Y. Zhang, S. Ayyampalayam, T. J. Barkman, N. Nguyen, N. Matasci, D. R. Nelson, E. Sayyari, E. K. Wafula, R. L. Walls, T. Warnow, H. An, N. Arrigo, A. E. Baniaga, S. Galuska, S. A. Jorgensen, T. I. Kidder, H. Kong, P. Lu-Irving, H. E. Marx, X. Qi, C. R. Reardon, B. L. Sutherland, G. P. Tiley, S. R. Welles, R. Yu, S. Zhan, L. Gramzow, G. Theißen, G. K.-S. Wong, One Thousand Plant Transcriptomes Initiative, One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).

52. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797 (2004).

609 53. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing Large Minimum
610 Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and*
611 *Evolution* **26**, 1641–1650 (2009).

612 54. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-
613 analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

614 55. S. Q. Le, O. Gascuel, An Improved General Amino Acid Replacement Matrix.
615 *Molecular Biology and Evolution* **25**, 1307–1320 (2008).

616 56. F. Lemoine, J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila
617 Felipe, T. De Oliveira, O. Gascuel, Renewing Felsenstein's phylogenetic bootstrap
618 in the era of big data. *Nature* **556**, 452–456 (2018).

619 57. J. J. Almagro Armenteros, M. Salvatore, O. Emanuelsson, O. Winther, G. von
620 Heijne, A. Elofsson, H. Nielsen, Detecting sequence signals in targeting peptides
621 using deep learning. *Life Sci. Alliance* **2**, e201900429 (2019).

622 58. F. Figueroa-Martinez, C. Jackson, A. Reyes-Prieto, Plastid Genomes from
623 Diverse Glaucophyte Genera Reveal a Largely Conserved Gene Content and
624 Limited Architectural Diversity. *Genome Biology and Evolution* **11**, 174–188 (2019).

625 59. D. C. Price, C. X. Chan, H. S. Yoon, E. C. Yang, H. Qiu, A. P. M. Weber, R.
626 Schwacke, J. Gross, N. A. Blouin, C. Lane, A. Reyes-Prieto, D. G. Durnford, J. A. D.
627 Neilson, B. F. Lang, G. Burger, J. M. Steiner, W. Löffelhardt, J. E. Meuser, M. C.
628 Posewitz, S. Ball, M. C. Arias, B. Henrissat, P. M. Coutinho, S. A. Rensing, A.
629 Symeonidi, H. Doddapaneni, B. R. Green, V. D. Rajah, J. Boore, D. Bhattacharya,
630 Cyanophora paradoxa Genome Elucidates Origin of Photosynthesis in Algae and
631 Plants. *Science* **335**, 843–847 (2012).

632 60. K. Katoh, D. M. Standley, MAFFT Multiple Sequence Alignment Software Version
633 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**,
634 772–780 (2013).

635 61. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A.
636 von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for
637 Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**,
638 1530–1534 (2020).

639 62. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin,
640 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature*
641 *Methods* **14**, 587–589 (2017).

642 63. S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M.
643 Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. L. Sonnhammer, L. Hirsh,
644 L. Paladin, D. Piovesan, S. C. E. Tosatto, R. D. Finn, The Pfam protein families
645 database in 2019. *Nucleic Acids Research* **47**, D427–D432 (2019).

64. Z. Yang, A space-time process model for the evolution of DNA sequences. *Genetics* **139**, 993–1005 (1995).

65. J. Soubrier, M. Steel, M. S. Y. Lee, C. Der Sarkissian, S. Guindon, S. Y. W. Ho, A. Cooper, The Influence of Rate Heterogeneity among Sites on the Time Dependence of Molecular Rates. *Molecular Biology and Evolution* **29**, 3345–3358 (2012).

66. H. Shimodaira, An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology* **51**, 492–508 (2002).

67. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**, 1586–1591 (2007).

68. D. L. Swofford, Phylogenetic analysis using parsimony. (1998).

69. S. A. Ishikawa, A. Zhukova, W. Iwasaki, O. Gascuel, A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Molecular Biology and Evolution* **36**, 2069–2085 (2019).

70. J. Ng, Z. Guo, O. Mueller-Cajar, Rubisco activase requires residues in the large subunit N terminus to remodel inhibited plant Rubisco. *Journal of Biological Chemistry* **295**, 16427–16435 (2020).

71. H. Kane, J. Viil, B. Entsch, K. Paul, M. Morell, T. Andrews, An Improved Method for Measuring the CO2/O2 Specificity of Ribulose bisphosphate Carboxylase-Oxygenase. *Functional Plant Biol.* **21**, 449–461 (1994).

72. J. Pierce, N. E. Tolbert, R. Barker, Interaction of ribulose bisphosphate carboxylase/oxygenase with transition-state analogs. *Biochemistry* **19**, 934–942 (1980).

73. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, “Protein Identification and Analysis Tools on the ExPASy Server” in *The Proteomics Protocols Handbook*, J. M. Walker, Ed. (Humana Press, Totowa, NJ, 2005; https://doi.org/10.1385/1-59259-890-0:571), pp. 571–607.

74. F. W. Studier, Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207–234 (2005).

75. D. S. Kubien, C. M. Brown, H. J. Kane, “Quantifying the Amount and Activity of Rubisco in Leaves” in *Photosynthesis Research Protocols*, R. Carpentier, Ed. (Humana Press, Totowa, NJ, 2011; https://doi.org/10.1007/978-1-60761-925-3_27), pp. 349–362.

76. B. L. Horecker, J. Hurwitz, P. K. Stumpf, “[29A] The enzymatic synthesis of ribulose-1,5-diphosphate and xylulose-5-phosphate” in *Methods in Enzymology* (Academic Press, 1957;

https://www.sciencedirect.com/science/article/pii/S0076687957033698)vol. 3, pp. 193–195.

77. F. G. Pearce, T. J. Andrews, The Relationship between Side Reactions and Slow Inhibition of Ribulose-bisphosphate Carboxylase Revealed by a Loop 6 Mutant of the Tobacco Enzyme. *Journal of Biological Chemistry* **278**, 32526–32536 (2003).

78. D. L. Edmondson, M. R. Badger, T. J. Andrews, A Kinetic Characterization of Slow Inactivation of Ribulose bisphosphate Carboxylase during Catalysis. *Plant Physiol* **93**, 1376–1382 (1990).

79. J. T. Edsall, J. Wyman, J. G. Hoffman, Thermodynamics, Electrostatics, and the Biological Significance of the Properties of Matter. *Physics Today* **11**, 50–52 (1958).

80. A. W. F. Edwards, The genetical theory of natural selection. *Genetics* **154**, 1419–1426 (2000).

81. H. A. Orr, The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* **6**, 119–127 (2005).

82. Y. Hirakawa, “Chapter Ten - Chlorarachniophytes With Complex Secondary Plastids of Green Algal Origin” in *Advances in Botanical Research*, Y. Hirakawa, Ed. (Academic Press, 2017; https://www.sciencedirect.com/science/article/pii/S006522961730054X)vol. 84, pp. 359–393.

83. B. A. Curtis, G. Tanifuji, F. Burki, A. Gruber, M. Irimia, S. Maruyama, M. C. Arias, S. G. Ball, G. H. Gile, Y. Hirakawa, J. F. Hopkins, A. Kuo, S. A. Rensing, J. Schmutz, A. Symeonidi, M. Elias, R. J. M. Eveleigh, E. K. Herman, M. J. Klute, T. Nakayama, M. Oborník, A. Reyes-Prieto, E. V. Armbrust, S. J. Aves, R. G. Beiko, P. Coutinho, J. B. Dacks, D. G. Durnford, N. M. Fast, B. R. Green, C. J. Grisdale, F. Hempel, B. Henrissat, M. P. Höppner, K.-I. Ishida, E. Kim, L. Kořený, P. G. Kroth, Y. Liu, S.-B. Malik, U. G. Maier, D. McRose, T. Mock, J. A. D. Neilson, N. T. Onodera, A. M. Poole, E. J. Pritham, T. A. Richards, G. Rocap, S. W. Roy, C. Sarai, S. Schaack, S. Shirato, C. H. Slamovits, D. F. Spencer, S. Suzuki, A. Z. Worden, S. Zauner, K. Barry, C. Bell, A. K. Bharti, J. A. Crow, J. Grimwood, R. Kramer, E. Lindquist, S. Lucas, A. Salamov, G. I. McFadden, C. E. Lane, P. J. Keeling, M. W. Gray, I. V. Grigoriev, J. M. Archibald, Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* **492**, 59–65 (2012).

84. O. Mueller-Cajar, S. M. Whitney, Evolving improved Synechococcus Rubisco functional expression in Escherichia coli. *Biochemical Journal* **414**, 205–214 (2008).

85. N. Prywes, N. R. Phillips, L. M. Oltrogge, S. Lindner, L. J. Taylor-Kearney, Y.-C. C. Tsai, B. de Pins, A. E. Cowan, H. A. Chang, R. Z. Wang, L. N. Hall, D. Bellieny-Rabelo, H. M. Nisonoff, R. F. Weissman, A. I. Flamholz, D. Ding, A. Y. Bhatt, O. Mueller-Cajar, P. M. Shih, R. Milo, D. F. Savage, A map of the rubisco biochemical landscape. *Nature* **638**, 823–828 (2025).

721 86. M. Gionfriddo, R. Birch, T. Rhodes, S. Buck, T. Skinner, I. Andersson, S.
722 Whitney, Laboratory evolution of Rubisco solubility and catalytic switches to
723 enhance plant productivity. *Nature Plants* **11**, 1939–1950 (2025).

724 87. L.-Y. Xia, Y.-L. Jiang, W.-W. Kong, H. Sun, W.-F. Li, Y. Chen, C.-Z. Zhou,
725 Molecular basis for the assembly of RuBisCO assisted by the chaperone Raf1. *Nat.*
726 *Plants* **6**, 708–717 (2020).

727 88. L. Li, S. Wang, H. Wang, S. K. Sahu, B. Marin, H. Li, Y. Xu, H. Liang, Z. Li, S.
728 Cheng, T. Reder, Z. Çebi, S. Wittek, M. Petersen, B. Melkonian, H. Du, H. Yang, J.
729 Wang, G. K.-S. Wong, X. Xu, X. Liu, Y. Van de Peer, M. Melkonian, H. Liu, The
730 genome of Prasinoderma coloniale unveils the existence of a third phylum within
731 green plants. *Nature Ecology & Evolution* **4**, 1220–1231 (2020).

732 89. N. Rodríguez-Ezpeleta, H. Philippe, H. Brinkmann, B. Becker, M. Melkonian,
733 Phylogenetic Analyses of Nuclear, Mitochondrial, and Plastid Multigene Data Sets
734 Support the Placement of Mesostigma in the Streptophyta. *Molecular Biology and*
735 *Evolution* **24**, 723–731 (2007).

736

Acknowledgments

We thank Dr. Oliver Mueller-Cajar and members of the Hochberg and Erb labs for critical reading of the manuscript.

Funding:

This research was co-funded by the European Union (ERC, EVOCATION, 101040472 (GKAH, JZYN, DW). The views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. GKAH, TJE, JZYN, LS, AMK, DW, AH and CG were supported by the Max Planck Society. GKAH gratefully acknowledges funding from a LOEWE Excellence professorship. DW acknowledges funding from the International Max Planck Research School for Principles of Microbial Life: From molecules to cells, from cells to interactions, IMPRS-µLife. AMK was funded by a Long Term Fellowship by the European Molecular Biology Organization, (ALTF 684-2022) and a Marie Skłodowska-Curie fellowship (ECOfix, 101106795).

Author contributions:

Conceptualization: GKAH, TJE, JZYN, LS

Methodology: JZYN, LS, AMK

Investigation: JZYN, DW, LS, AH, CG

Visualization: JZYN

Funding acquisition: GKAH, TJE

Project administration: GKAH

Supervision: GKAH

Writing – original draft: JZYN, GKAH

Writing – review & editing: all authors.

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All raw data for kinetic traces, phylogenetic trees, alignments, and ancestral sequences are deposited in the Edmund repository of the Max-Planck Institute and will be made available upon peer-review and publication.