

The Evolution of RuBisCO Stability at the Thermal Limit of Photoautotrophy

Scott R. Miller,*¹ Michele A. McGuirl,^{1,2} and Darla Carvey¹

¹Division of Biological Sciences, University of Montana

²Center for Biomolecular Structure and Dynamics, University of Montana

*Corresponding author: E-mail: scott.miller@umontana.edu.

Associate editor: Daniel Falush

The reported sequences have been deposited in GenBank (accession nos. KC208032–KC208043).

Abstract

A long-standing question in evolutionary biology is how organisms adapt to novel environments. In North American hot springs, diversification of a clade of the cyanobacterium *Synechococcus* into hotter environments has resulted in the unique innovation of a light-driven ecosystem at temperatures up to 74°C, and temperature adaptation of photosynthetic carbon fixation with the Calvin cycle contributed to this process. Here, we investigated the evolution of thermostability of the Calvin cycle enzyme ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO) during *Synechococcus* divergence. Circular dichroism thermal scans revealed that the RuBisCO of the most thermotolerant *Synechococcus* lineage is more stable than those of other lineages or of resurrected ancestral enzymes. Using site-directed mutagenesis, we next identified four amino acid substitutions that together increased stability and activity of this enzyme at higher temperatures. These are clustered near critical subunit interfaces distant from the active site. Each of the four amino acids is also observed in a less thermostable *Synechococcus* RuBisCO, and the impact on stability of three of these appears to be epistatic. Recombination analyses that allow for recurrent mutation as well as patterns of synonymous variation surrounding these sites suggest that the evolution of a more thermostable RuBisCO may have involved homologous recombination. Our results provide insights on the molecular evolutionary processes that shape niche differentiation and ecosystem function.

Key words: adaptation, thermostability, functional synthesis, niche extension.

Introduction

Identifying the functional significance of genetic variation is a central goal of evolutionary biology, and a growing number of case studies have applied the tools of molecular biology and biochemistry to achieve insights on the mechanisms of adaptive diversification (Golding and Dean 1998; Watt and Dean 2000; Dean and Thornton 2007; Storz and Wheat 2010). Adaptation has potential impacts beyond an individual's fitness, however, including shifts in ecosystem structure and function (Harmon et al. 2009). Addressing how properties at these larger scales of biological organization are shaped by changes in the structure and function of molecules which affect organism performance remains a fundamental challenge for biologists with implications for our understanding of the origins and maintenance of diversity.

Geothermal environments provide an excellent system to investigate this issue. In most hot springs, inorganic chemicals supply energy for primary production by bacteria and/or archaea at temperatures greater than approximately 57–64°C (Ward et al. 2012). Alkaline hot springs of western North America, however, are populated by lineages of an ancient clade of the cyanobacterium *Synechococcus* with divergent thermal ecologies, the most thermotolerant of which has uniquely evolved the ability to maintain photoautotrophic growth at temperatures up to 74°C (Miller and Castenholz 2000). One consequence of niche expansion into novel

thermal environments during diversification of this group from a moderately thermophilic common ancestor has therefore been a shift in the community's energy economy at higher temperatures to one based on solar radiation.

The evolution of the temperature dependence of photoautotrophic carbon fixation by the Calvin cycle has contributed to *Synechococcus* divergence (Brock 1967; Meeks and Castenholz 1978). The first and slowest step in this cycle is catalyzed by ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO). Here, we have combined an "ancestor resurrection" approach (Thornton 2004) with site-directed mutagenesis and physical biochemical methods to investigate the evolution of RuBisCO stability and function during the diversification of *Synechococcus*. We report that four amino acid substitutions contribute to enhanced enzyme thermostability and function in the most thermotolerant *Synechococcus* lineage compared with both contemporary enzymes of less thermotolerant lineages and inferred ancestral RuBisCOs. These results suggest a lineage-specific adaptation of RuBisCO stability and function during niche extension to higher temperatures.

Results and Discussion

Study System

The *Synechococcus* A/B clade includes taxa that have diverged substantially in thermotolerance, with members of the

A clade generally exhibiting both greater fitness at higher temperatures in the laboratory than members of the B clade (Miller and Castenholz 2000; Allewalt et al. 2006) and greater abundance at higher temperatures in situ (Ward et al. 1998; Miller et al. 2009). In our sample (fig. 1A), A clade member strain OH28 was the only strain capable of growth at 70°C in laboratory culture, as compared with maximum growth temperatures at or below 65°C for A clade strains OS-A and OH2 and at or below 61°C for B clade strains OS-B' and OH20 (Miller and Castenholz 2000; Allewalt et al. 2006).

Synechococcus strain OH28 also exhibited a thermal reaction norm for photosynthetic carbon assimilation that was shifted to higher temperatures compared with less thermotolerant strains, with peak assimilation observed at 70°C (supplementary fig. S1, Supplementary Material online). This temperature was supra-optimal for all other strains, and the decline in carbon assimilation rate at temperatures greater than 60°C was faster for the B-clade strains than for strains OS-A and OH2. Photosynthetic carbon assimilation is a complex trait which integrates several processes in the cell, including light-dependent ATP production, which provides energy

for carbon flux through the Calvin cycle. For several reasons, we focused on the Calvin cycle enzyme RuBisCO for further investigation to begin to dissect the molecular mechanisms that contribute to strain OH28's unique ability to maintain photoautotrophic growth at 70°C. RuBisCO can exert substantial control over the flux of carbon through the Calvin cycle, particularly at high temperature and irradiance (Stitt and Schulze 1994). Given the agricultural importance of the enzyme, there is also an abundance of biochemical and biophysical data available for comparative analysis. Specifically, we aimed to test whether strain OH28 had evolved a more thermostable RuBisCO than those of less thermotolerant strains or inferred ancestral versions of the enzyme.

Stability of Native and Ancestral *Synechococcus* RuBisCOs

As in plants, *Synechococcus* RuBisCO is a form I hexadecamer of eight large subunits and eight small subunits encoded by *rbcL* and *rbcS*, respectively. In the *rbc* operon of most cyanobacteria, including *Synechococcus*, these loci are separated by *rbcX*, which encodes a chaperone that enables proper

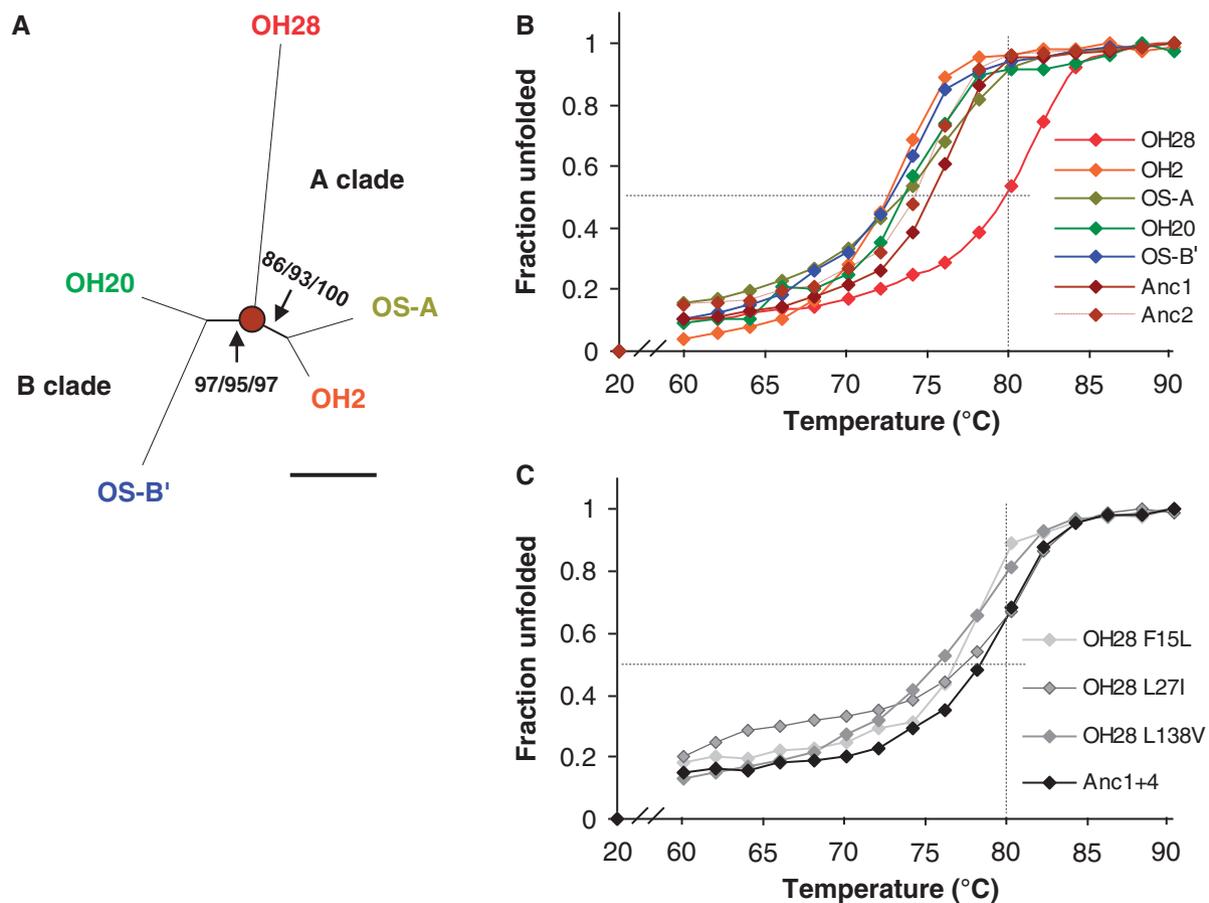


Fig. 1. Evolution of a thermostable RuBisCO in thermophilic *Synechococcus*. (A) *Synechococcus* genealogy inferred from concatenated sequence data for *rbcLXS* and upstream locus *clpP*. The brown node indicates the ancestor of the A clade in which enhanced thermotolerance evolved. ML branch lengths are shown (scale bar is 0.05 substitutions/site). Values at the nodes are bootstrap percentages for ML, parsimony, and neighbor-joining trees. (B) Plots of the fraction of secondary structure lost with increasing temperature for native and ancestor RuBisCOs obtained from CD thermal scans monitored at 222 nm. See supplementary figure S2, Supplementary Material online, for a representative spectral plot of the CD thermal scans. See text and table 1 for T_m confidence intervals. Colors are as described in (A). (C) Plots of the fraction of secondary structure lost with increasing temperature for variants that destabilized OH28 RuBisCO and for the forward construct Anc1+4.

assembly of the RbcL₈ core (Saschenbrecker et al. 2007; Liu et al. 2010). We cloned *rbclXS* of divergent *Synechococcus* for expression in *Escherichia coli* and achieved >70-fold RuBisCO purification by a combination of heat treatment, ion exchange chromatography, and gel filtration (supplementary table S1, Supplementary Material online).

We next used circular dichroism (CD) to monitor the unfolding of purified RuBisCOs with increasing temperature. These data could be modeled as a two-state folding mechanism, with melting temperature (T_m) estimated as the temperature at which 50% of helical structure is lost (supplementary fig. S2, Supplementary Material online). Purified OH28 RuBisCO exhibited greater stability ($T_m = 79.5^\circ\text{C}$; 95% CI = 79.2–79.9°C) than RuBisCOs of less thermotolerant strains (T_m estimates ranging from 72.3–73.6°C; fig. 1B; table 1).

A more thermostable RuBisCO therefore appears to have evolved since divergence from the ancestor of the *Synechococcus* A lineage (brown node in fig. 1B). To test this, we reconstructed the RuBisCO sequence of this ancestor by sequential site-directed mutagenesis for expression, purification, and analysis. Both subunits exhibited only low amounts of variation in the *Synechococcus* genealogy (21 of 474 amino acid positions in RbcL and 13 of 110 in RbcS are variable; supplementary fig. S3, Supplementary Material online). The combination of substantial phenotypic divergence yet high sequence identity makes *Synechococcus* RuBisCO an excellent model for investigating adaptive diversification with an ancestor reconstruction approach. Maximum likelihood

(ML) sequence reconstruction (Yang et al. 1995) along the *rbclXS* tree (which is identical to that of the *Synechococcus* 16S rRNA species tree; Miller and Castenholz 2000) yielded two ancestors with joint posterior probability of ~85%: Anc1 (67.3%) and Anc2 (17.3%) differ only at position 36_L (Val in Anc1, Ile in Anc2; supplementary fig. S3, Supplementary Material online). The single credible RbcS reconstruction had a posterior probability of 91%. Ambiguity in the reconstructions was due to a few homoplastic sites (discussed later). Together, the ancestors differ from OH28 RuBisCO by either seven or eight substitutions in RbcL (supplementary fig. S3A, Supplementary Material online) and by eight substitutions in RbcS (supplementary fig. S3B, Supplementary Material online). Both Anc1 and Anc2 exhibited stabilities comparable with RuBisCOs of less thermotolerant strains (fig. 1B). We obtained similar results for ancestors reconstructed in different *rbclX* backgrounds (using OH28 vs. OH2 operons, respectively, as the templates for mutagenesis), indicating that results were not dependent upon the specific sequence of RbcX. Therefore, with the exception of the OH28 lineage, RuBisCO stability shows little evidence of evolution. This lineage-specific, substantial increase in OH28 RuBisCO stability suggests that this trait was not under selection during much of *Synechococcus* diversification but has contributed to the unique extension of the thermal niche of *Synechococcus* strain OH28 to higher temperatures.

Genetic Basis of Enhanced RuBisCO Stability at High Temperature

Several of the sites that differ between the strain OH28 enzyme and its inferred ancestors represent potential candidates for RuBisCO adaptation based on biochemical studies of cyanobacterial and plant model systems or on statistical tests of molecular adaptation. Mutations at positions 52–54_S of the RbcS βA – βB loop can reduce enzyme stability and carboxylation activity (Lee et al. 1991; Spreitzer et al. 2001), as can changes within the RbcS N-terminal arm, a conserved region including site 15_S which contacts RbcL helix α8 (Lee et al. 1991; Paul et al. 1991; Kostov et al. 1997; Genkov and Spreitzer 2009). In addition, a branch-site model of codon evolution (Yang and Nielsen 2002) identified the substitutions of highly conserved alanines by isoleucine at positions 389_L and 438_L along the terminal branch of the *rbclXS* genealogy leading to strain OH28 as potentially the result of positive selection ($2\Delta L = 12.02$, $P = 0.025$; supplementary fig. S4, Supplementary Material online). The posterior probabilities that these substitutions belonged to the positively selected site class ($\omega = 3.89$) were 0.94 and 0.86, respectively; in contrast, these codon sites were estimated to be under strong purifying selection along the other branches of the genealogy ($\omega = 0.01$; supplementary fig. S4, Supplementary Material online). These changes result in the addition of six additional methyl or methylene groups to the protein interior, which could potentially enhance thermal stability by facilitating tighter packing and van der Waals interactions. A similar shift to bulkier hydrophobic side-chains has been previously noted

Table 1. The Unfolding Temperatures (T_m) of Native, Ancestral, and Variant RuBisCOs and Their Confidence Intervals.

RuBisCO	T_m (95% CI)
OH28	79.5 (79.2–79.9)
OH2	72.3 (71.5–73.1)
OS-A	73.6 (72.8–74.3)
OH20	73.5 (72.8–74.1)
OS-B'	72.9 (72.1–73.8)
Anc1	75.5 (74.4–76.8)
Anc2	73.6 (73.2–74.1)
OH28 F15 _L	76.2 (75.5–77.0)
OH28 L138 _L V	75.4 (75.1–75.7)
OH28 L27 _L I	77.2 (75.7–78.6)
Anc1 + 4 (Anc1 V36 _L I V138 _L L L15 _S F I27 _S L)	78.2 (78.0–78.4)
OH28 I389 _L A I438 _L A	79.5 (78.7–80.4)
OH28 V465 _L I	80.4 (79.0–81.8)
OH28 D52 _S V H53 _S R F54 _S Y	79.7 (78.4–81.0)
OH28 V206 _L I M212 _L Q	78.4 (77.3–79.4)
OH28 I389 _L A I438 _L A V465 _L I	81.0 (79.6–82.4)
OH28 I389 _L A I438 _L A V465 _L I V206 _L I M212 _L Q	80.4 (79.6–81.2)
OH28 K8 _L Q	80.0 (78.5–81.5)
OH28 A105 _S V	81.4 (79.3–83.5)
OH28 L75 _S V	79.6 (78.9–80.3)
OH28 G64 _S S	79.3 (78.5–80.2)
Anc1 + L15 _S F	74.0 (72.6–75.4)
Anc2 + I27 _S L	73.4 (72.5–74.2)

in the proteins of thermophilic methanogens (Haney et al. 1999).

To identify which amino acid substitutions contributed to the increased thermal stability of OH28 RuBisCO, we mutated OH28 *rbcl* and *rbcS* at sites that differ between OH28 and the two reconstructed ancestors (either singly or in combination). Among the candidates noted earlier, only the F15_SL mutation resulted in a reduction in stability ($T_m = 76.2^\circ\text{C}$; fig. 1C; table 1), whereas neither the D52_SV H53_SR F54_SY nor the I389_LA I438_LA variants differed from the native OH28 enzyme (T_m s are 79.7°C and 79.5°C , respectively). Our observation that sites 389_L and 438_L did not impact stability emphasizes the importance of experimentally testing hypotheses generated by tests for positive selection. We have not tested whether these or other substitutions may be adaptive with respect to some other property of RuBisCO, such as specificity for carbon dioxide over oxygen.

In addition to F15_SL, changes at only three of the other positions resulted in a phenotype that differed from OH28 RuBisCO (table 1). L27_SI ($T_m = 77.2^\circ\text{C}$) and L138_LV ($T_m = 75.4^\circ\text{C}$) exhibited reduced stability (fig. 1C), whereas I36_LV was insoluble despite repeated purification attempts (we could, however, purify soluble enzyme following reversion of 36_L in the Anc2 background to create Anc1; discussed earlier). A forward construct of the derived OH28 residues at these four sites in an Anc1 background (Anc1+4 = Anc1 V36_LI V138_LL L15_SF I27_SL) is sufficient to nearly recapitulate native OH28 RuBisCO thermostability ($T_m = 78.2^\circ\text{C}$; 95% CI = 78.0 – 78.4°C ; fig. 1C). Purified OH28 and Anc1+4 RuBisCOs also exhibited greater carboxylation specific activity at 70°C than Anc1 and Anc2 (fig. 2), though the consequences of individually mutating OH28 RuBisCO at sites 138_L, 15_S, and 27_S varied in these short-term (5 min) end-point functional assays.

None of the four amino acid replacements that contribute to increased RuBisCO thermostability is unique to strain OH28. Each is also observed in a less thermostable enzyme: Ile36_L in OH2, Leu138_L in OH20, and Phe15_S and Leu27_S in OS-A (supplementary fig. S3, Supplementary Material online). This raises the possibility that the effect on stability of one or more of these amino acids may be epistatic. Although a comprehensive analysis of epistasis and its potential role in shaping which evolutionary pathways were selectively accessible (Weinreich et al. 2005) during *Synechococcus* RuBisCO diversification awaits further investigation, mutants were available for three of the four sites for an initial appraisal of whether a

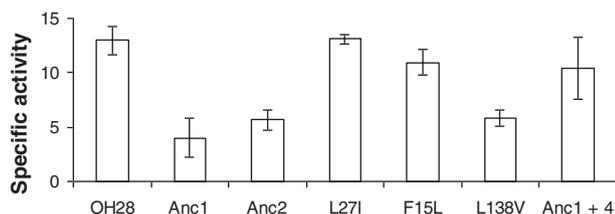


Fig. 2. Specific carboxylation activities (nmol ¹⁴C fixed mg protein⁻¹ min⁻¹) of native, ancestral, and variant RuBisCOs. Activity was assayed for 5 min at 70°C . Bars represent two standard errors.

substitution had the same effect on stability in different genetic backgrounds. Both F15_SL and L27_SI mutations are destabilizing in an OH28 background (fig. 1B and C; table 1), but neither affects stability in an ancestral background: that is, T_m s do not differ either between Anc1 and the variant Anc1 + L15_SF ($T_m = 74.0^\circ\text{C}$) or between Anc2 and the variant Anc2 + I27_SL ($T_m = 73.4^\circ\text{C}$), respectively (table 1). In addition, although V36_LI appears to be stabilizing in the OH28 background (as OH28 I36_LV failed to assemble properly), this substitution was slightly destabilizing (i.e., exhibits sign epistasis) in a reconstructed ancestral background: T_m of Anc1 + V36_LI (=Anc2) is lower than that of Anc1 (fig. 1B; table 1). The greater thermostabilities of OH28 and Anc1+4 RuBisCOs therefore appear to involve epistatic interactions among residues.

Structural Context of RuBisCO Adaptation

A homology model of the strain OH28 RuBisCO shows that the stabilizing amino acid changes are clustered in regions that ring the holo-enzyme center and poles (fig. 3A). Site 15_S is at an RbcS–RbcL interface and is a hydrogen bond donor (main chain–side chain) to Glu425_L of helix $\alpha 8$ in previously solved crystal structures (fig. 3B). Site 27_S is also near this interface with helix $\alpha 8$ (25_S is hydrogen-bonded to 429_L and 433_L of $\alpha 8$). Sites 36_L and 138_L are located on adjacent anti-parallel β -strands, contact each other through their side-chains and are buried in a region of extensive RbcL intradimer and interdimer interactions (fig. 3C). 138_L is adjacent to Asp137_L, which may play a critical role in holo-enzyme assembly and stability, as it is one of the most conserved sites of the RbcL superfamily (Tabita et al. 2007). Asp137_L and Lys316_L form an intrasubunit salt bridge, and 138_L is a hydrogen bond acceptor (main chain–side chain) with Lys316_L (fig. 3C).

It is noteworthy that the stabilizing substitution sites are located near 5 of the 11 distinct intersubunit salt bridges in form I RuBisCO, as the high content of contacts between charged residues at subunit interfaces compared with hydrophobic interactions is unusual for an oligomeric protein (Knight et al. 1990). These include interactions between RbcL and RbcS (Lys164_L–Glu13_S and Arg167_L–Glu13_S; fig. 3B) as well as within (Glu109_L–Arg253_L and Glu110_L–Arg213_L) and between (Glu110_L–Lys146_L) RbcL dimers (fig. 3C). The likely importance of the salt bridges for RuBisCO stability has long been recognized (Knight et al. 1990; Curmi et al. 1992) yet has not been extensively investigated. Replacement of Glu13_S with Val disrupts holo-enzyme assembly (Fitchen et al. 1990), as do several mutations in the vicinity of the salt bridges near 36_L and 138_L, including Ser112_L to Phe (Avni et al. 1989), Leu37_L to Pro (Smith and Tabita 2003), and Phe108_L to Leu (Smith and Tabita 2003). The latter site has an extensive side chain–side chain contact with 36_L (fig. 3C). Our inability to purify the OH28 I36_LV variant further highlights the importance of interactions within this region.

Although the mechanism(s) of increased RuBisCO thermostability remains to be determined, the structural context suggests some possibilities. For example, the greater van der

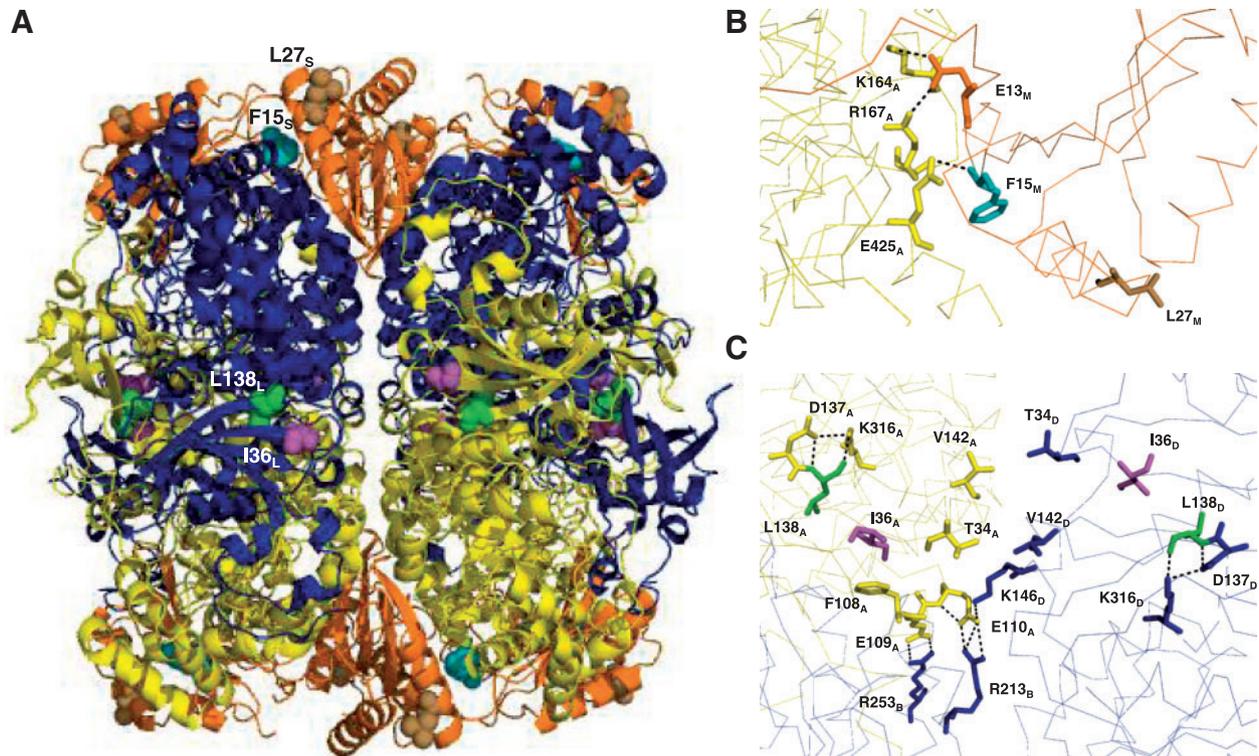


Fig. 3. Structural context of adaptive amino acid substitutions in RuBisCO. (A) Homology model of the *Synechococcus* strain OH28 RuBisCO. Individual large subunit monomers (blue and yellow, respectively) form four L_2 dimers capped at each pole by four small subunits (orange). Sites impacting thermostability are depicted as space-filling models: Phe15_s (teal); Leu27_s (brown); Ile36_L (magenta); and Val138_L (green). (B) RbcS sites impacting thermostability. Ribbon colors are as in (A) with chain label subscripts as in the *Synechococcus* PCC 6301 model template (PDB ID 1rbl). Dashed lines indicate salt bridges and hydrogen bonds discussed in the text. (C) RbcL sites impacting thermostability are located near the hydrophobic patch between L_2 dimers including symmetry-related residues Thr34 and Val142 from chains A and D. Dashed lines indicate polar contacts involving chains A, B, and D. For clarity, chain C from the CD L_2 dimer is not shown, and the residues from A and D involved in symmetry-related salt bridges with chain C are not highlighted.

Waals volume of the surrounding protein environment in the vicinity of intersubunit salt bridges in the OH28 RuBisCO may strengthen the electrostatic interactions between subunits. The bulkier side chain of phenylalanine compared with leucine at 15_s may also stabilize the RbcS–RbcL interface through the exclusion of additional water molecules.

Possible Evolutionary Origins of a More Thermostable RuBisCO

The possible explanations for the observed pattern of homoplastic amino acid sites that are stabilizing in the OH28 RuBisCO background are recurrent mutation with selection, recombination, or a combination of the two. Because synonymous divergence is high in the sample (Jukes–Cantor corrected $\pi_s = 0.45$ for *rbcL* and 0.64 for *rbcS*), to test for the presence of recombination with a gene conversion model appropriate for bacteria, we used a coalescent-based method (McVean et al. 2002) that employs a finite-sites model of mutation to allow for recurrent mutations to have occurred at a nucleotide position. The model was developed specifically for bacterial and viral data sets for which the assumption of an infinite-sites mutation model may not be appropriate. Simulations have demonstrated that this

estimator performs well even when most sites analyzed have experienced multiple mutations (McVean et al. 2002), including values of sequence diversity (per site $\theta = 0.5$) five-times greater than that of our sample (per site $\theta = 0.1$). Analyses of viral data sets of Worobey (2001; per site $\theta = 0.32$) and Woelk et al. (2001; per site $\theta = 0.09$) by the method have further indicated that a high level of sequence diversity does not in itself produce high or statistically significant estimates of γ , the population rate of recombination caused by gene conversion (per site $\gamma = 0.84$ and 3.0, respectively; McVean et al. 2002).

Per site γ for the *rbcLXS* data was estimated to be low relative to the mutation rate ($\gamma = 0.002$; $\gamma/\theta = 0.02$, for an estimated 10 recombination events during *Synechococcus rbcLXS* diversification; supplementary fig. S5, Supplementary Material online) but was very highly significant ($P = 0$) by a likelihood permutation test that permutes by location to evaluate whether sites are exchangeable (i.e., whether location of a site in the sequence matters). With recombination, closely linked sites will have correlated genealogies and will therefore not be exchangeable, whereas sites are expected to be exchangeable with recurrent mutation in the absence of recombination. We obtained a similar result if we analyze only the synonymous variants in the data set ($\gamma = 0.003$; $P = 0$),

which removes any potential for false positives due to selection on recurrent nonsynonymous substitutions.

The permutation tests indicate a general sample-wide correlation of genealogies among physically linked variants consistent with a history of recombination. We next investigated the specific patterns of codon usage for the four stabilizing sites in the OH28 RuBisCO and nearby synonymous variants. Retention of identical synonymous variants between OH28 and a donor-like sequence from a less thermotolerant strain would represent powerful evidence in favor of recombination, although such a molecular signature would also be expected to erode over time given the potential for synonymous-site divergence following a past recombination event from an ancestral donor. A data pattern consistent with recombination is strongest for nucleotide tracts surrounding sites 36_L and 138_L. *rbcl* sites 32–40 are identical between OH28 and OH2, including at the two parsimony-informative synonymous variants flanking the amino acid substitution (34_L and 37_L; [supplementary fig. S6A, Supplementary Material](#) online). Further, the use of ACT at 34_L is extremely rare in OH28 *rbcl* (2 of 30 Thr codons, for a relative synonymous codon usage [RCSU] equal to 0.27; $RCSU_{max} = 4$ for Thr) but less so for OH2 (its RCSU of 0.67 is highest among the *Synechococcus* sequences). By contrast, OH28 and OH20 are identical at three of four parsimony-informative synonymous variants in the region spanning codons 137_L–145_L (137_L, 141_L, and 145_L; the exception is 140_L, for which OH28 and OS-B' share ATT; [supplementary fig. S6B, Supplementary Material](#) online) and at six of the eight informative variants between codons 131_L and 159_L (they are identical at 131_L, 133_L, and 155_L; OH28, OH2, and OS-A are GAG at 136_L). The use of TTG at 138_L itself as well as at 145_L is very unusual for OH28 (4 of 40 Leu codons in *rbcl*, for a RCSU equal to 0.6; $RCSU_{max} = 6$ for Leu) but not so for OH20 (RCSU = 1.54). The nucleotides surrounding 36_L and 138_L therefore appear to have different evolutionary histories. The data are less clear for sites 15_S and 27_S. OH28 and OS-A share identity at two of the three parsimony-informative synonymous variants near 15_S (14_S and 18_S; the exception is 17_S, for which OH28 and OS-B' share TAT; [supplementary fig. S6C, Supplementary Material](#) online). The only parsimony-informative site between codons 19_S and 32_S, however, is the nonsynonymous site 27_S itself ([supplementary fig. S6D, Supplementary Material](#) online).

Though we cannot completely rule out that recurrent mutation with selection is solely responsible for generating the patterns of variation for these divergent sequences, we believe that the above analyses strongly suggest the contribution of recombination to the evolutionary origins of a more thermostable RuBisCO. Many examples of the importance of homologous recombination for bacterial adaptation have been reported, most notably for pathogens. For example, recombination can be a source of novel surface antigen variation that may enable evasion of the host immune system (Millman et al. 2001; Andrews and Gojobori 2004; Qiu et al. 2004) as well as foster the evolution of antibiotic resistance through the diversification of antibiotic targets (Spratt et al. 1992). Given the impact of *Synechococcus* niche expansion for community energetics at higher temperatures in alkaline

geothermal environments, our study may represent a new and interesting example of the importance of the recombination of standing variation during the adaptive diversification of bacteria.

Materials and Methods

Thermal Reaction Norms for Photosynthetic Carbon Assimilation

Batches (300 ml) of exponentially growing cells were cultured as previously described (Miller and Castenholz 2000), harvested by filtration, and resuspended in D medium (Castenholz 1988) to a concentration of 0.1 $\mu\text{g Chl } a \text{ ml}^{-1}$. For a series of temperatures ranging between 45 and 70°C, triplicate 5 ml aliquots of cells were incubated for 1 h in the presence of 0.06 $\mu\text{Ci ml}^{-1}$ [¹⁴C]-sodium bicarbonate at saturating irradiance (200 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$) provided by cool white fluorescent lamps. Assays were terminated by the addition of 300 μl formalin, and radiolabeled carbon assimilation was estimated with a Beckman Coulter LS6500 liquid scintillation counter as previously described (Miller et al. 1998).

Cloning of *rbclXS* and Site-Directed Mutagenesis

Native RuBisCOs were amplified by polymerase chain reaction and cloned using the Champion pET Directional TOPO Expression Kit (Invitrogen) according to manufacturer instructions. Mutants were constructed from isolated plasmid DNA (Plasmid Midi Kit; Qiagen) using the QuickChange XL (Stratagene) and cloned into *E. coli* XL10 Gold following manufacturer instructions. Following plasmid DNA isolation from individual transformants, sequencing of the entire *rbclXS* operon was performed for each construct at the University of Montana Murdock Sequencing Facility to confirm that only the specific desired mutation(s) was present. Plasmids were then cloned in *E. coli* strain BL21 (DE3).

Protein Expression and Purification

Shake cultures (240 rpm) of transformed *E. coli* strain BL21 (DE3) were grown at 37°C in seven flasks of 1 l LB broth containing 1 M sorbitol and 2.5 mM betaine and 75 $\mu\text{g ml}^{-1}$ carbenicillin until reaching an OD at 600 nm between 0.5 and 1.0. Protein expression was induced by the addition of IPTG to a final concentration of 0.75 mM, followed by incubation for an additional 12 h. Cells were pelleted and stored at –20°C for protein extraction. The cell pellet was resuspended in Buffer A (20 mM Tris–HCl, pH 8.0, 10 mM MgCl₂) at a ratio of 40% (w/v), and β -mercaptoethanol was added to a final concentration of 5 mM. Cells were disrupted by three rounds of sonication, with subsequent addition of PMSF, DNase and lysozyme. Following incubation at 4°C for 30 min, cells were pelleted by centrifugation at 8,000 rpm for 20 min. To remove thermolabile host protein, the supernatant was next heated at 65°C for 1 h and then centrifuged at 8,000 rpm for 20 min. The supernatant was concentrated to approximately 4 ml with a 15 ml Amicon Ultra centrifugal filter unit with 100,000 molecular weight cutoff (Amicon) at 4,000 $\times g$ in a swinging bucket rotor. Conductivity and pH of the

concentrate were adjusted with sterile $3\times$ dH₂O and either NaOH or HCl to match Buffer A (265 μ S, pH 8.0) for anion-exchange chromatography on an FPLC 15 Q ion-exchange column. Protein was eluted with a gradient of Buffer B (same as Buffer A plus 1 M NaCl). RuBisCO-containing fractions (between 20% and 23% Buffer B) identified by carboxylation assays (discussed later) were pooled and concentrated with an Amicon Ultra filter unit as above, then loaded on an FPLC Superose 12 gel filtration column with a fixed gradient of Buffer A containing 0.1 M NaCl for a total of 1.5 column volumes at 0.5 ml min⁻¹. Fractions with carboxylation activity were pooled, and the presence of only two proteins with MW conforming to Rbcl and RbcS, respectively, was confirmed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (Laemmli 1970) on a PhastSystem unit (GE Healthcare).

Circular Dichroism

CD spectra of 2 ml samples of 75 nM RuBisCO solution were recorded between 190 and 300 nm with a JASCO J-810 CD spectrometer equipped with Peltier PFD-425S temperature-controlled sample compartment at 20, 30, and 40°C and for a thermal scan between 50 and 90°C at a rate of 2°C min⁻¹, with three measurements of ellipticity taken at each temperature point. To estimate T_m , replicate thermal scans for independent preparations of a given RuBisCO were pooled and analyzed by the regression of ellipticity at 222 nm on temperature over the linear interval. T_m and 95% confidence intervals were estimated by solving for y (the fraction unfolded) = 0.5.

Carboxylase Assays

We used a modification of the assay of Tabita et al. (1978) to measure carboxylase activity of purified RuBisCO. Reactions (500 μ l) containing 10 μ g of enzyme, 150 mM MOPS/KOH (pH 7.8), 20 mM MgCl₂ · 6H₂O, and 0.4 μ Ci Na¹⁴CO₃ were pre-incubated at 65°C (the temperature empirically determined to maximize enzyme activation) for 30 min to activate the enzyme. The assay was initiated by the addition of 20 μ l of 25 mM ribulose-1,5-bisphosphate and terminated after 5 min by the addition of 100 μ l 2 M HCl. Acid-stable ¹⁴C was counted on a Beckman Coulter LS6500 liquid scintillation counter. A modification of this assay (in which case 200 μ l of protein sample were assayed at 65°C) was also used to monitor different stages of the purification process (including following cell disruption, heat denaturation, Amicon concentration, and anion exchange chromatography).

Phylogenetics, Ancestor Reconstruction, and Statistical Tests of Molecular Adaptation

A total of 2,595 bp of concatenated sequence data of *Synechococcus* *rbclXS* and upstream *clpP* were aligned with CLUSTAL W (Thompson et al. 1994). Phylogeny reconstruction was performed with PAUP* (Swofford 1996). The most parsimonious unweighted tree was found with the branch-and-bound method. The GTR + G model of sequence evolution was used for the ML analysis, as selected by both hierarchical likelihood ratio tests and the Akaike information

criterion implemented in Modeltest (Posada and Crandall 1998). To find the ML tree, a starting tree was obtained by random sequence addition followed by branch-swapping with the TBR algorithm. One thousand bootstrap pseudoreplicates were obtained for the ML analysis, and 10,000 pseudoreplicates were obtained for parsimony and neighbor-joining trees.

Amino acid sequences at ancestral nodes along the *rbclXS* genealogy were reconstructed by the likelihood approach of Yang et al. (1995) implemented in PAML (Yang 1997). Similar results were obtained irrespective of the substitution rate matrix used.

ML models of codon evolution implemented in PAML (Yang and Nielsen 2002) were used to identify amino acid replacements that were potentially the product of positive selection and which might contribute to the enhanced thermostability of strain OH28 RuBisCO. For each locus in the *rbclXS/clpP* data set and for each branch of the phylogeny, we implemented the following two models: 1) a branch-site model (Yang and Nielsen 2002) that tests for adaptive evolution at a few key sites along a lineage by allowing ω (i.e., d_N/d_S) to vary both among codon sites and among branches; and 2) a null model that constrains ω between 0 and 1 (i.e., neutral evolution). These models were statistically compared with a likelihood ratio test, and codons potentially under diversifying selection were identified by Bayes Empirical Bayes analysis. The amount of genetic variation among strains was calculated as nucleotide diversity, the average number of nucleotide differences per site between two randomly chosen sequences, for a sliding window of 50 nucleotides and a step size of 10 nucleotides.

Recombination Analyses

Population-scaled recombination rate for 2,196 nucleotides of the *rbclXS* operon was estimated by a modification of the composite-likelihood method of Hudson (2001) that accommodates a finite-sites model of sequence evolution using LDhat (McVean et al. 2002). The latter model allows for recurrent mutations to have occurred at a nucleotide position, a possibility that becomes more probable as sequences diverge. The method assumes that the gene genealogies of the loci can be modeled as a coalescent process according to the neutral Wright–Fisher model, but the estimates and significance tests appear to be robust to minor deviations from this model (McVean et al. 2002). We used a gene-conversion model of recombination, which estimates the recombination parameter γ (i.e., the population rate of recombination between two distantly linked loci caused by gene conversion). The gene conversion tract length was set to 100 nt, the length that maximized the composite likelihood for this data set (not shown). The number of recombination events was estimated by multiplication of γ/θ (where θ is Watterson's [1975] estimator of the population-scaled mutation rate) by the number of inferred mutation events in the *rbclXS* data. A likelihood permutation test (McVean et al. 2002) based on the null hypothesis that nucleotide site data are exchangeable in the absence of recombination was used to test whether the

estimated recombination rate was significantly greater than zero. We rejected the null hypothesis of no recombination if fewer than 5% of 1,000 permutated data sets had a composite likelihood score equal to or higher than the ML estimate for the original data.

Homology Modeling

The crystal structure of *Synechococcus* PCC 6301 RuBisCO (PDB ID 1rbl) was used as the template for SCWRL4 (Krivov et al. 2009) prediction of all atom models of individual subunit chains in the strain OH28 RuBisCO hexadecamer with original conformation retained. Strain OH28 and the template are 85% identical in RbcL amino acid sequence and 67% identical in RbcS amino acid sequence. At these levels of sequence identity, generally high quality model prediction accuracy for both the protein backbone and side-chain packing is expected (Martí-Renom et al. 2000; Wallner and Elofsson 2005). Graphics were produced from the resultant PDB files with PyMOL (www.pymol.org).

Supplementary Material

Supplementary material, figures S1–S6, and table S1 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Chris Wall, John Gerdes, and Michael Braden for discussions regarding homology modeling and also thank John McCutcheon, Stephen Sprang, Joe Thornton, and three anonymous reviewers for comments on an earlier version of the manuscript. This work was supported by the National Science Foundation (grant numbers MCB-0347627 and EF-0801999 to S.R.M.).

References

- Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM. 2006. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. *Appl Environ Microbiol.* 72:544–550.
- Andrews TD, Gojobori T. 2004. Strong positive selection and recombination drive the antigenic variation of the *pilE* protein of the human pathogen *Neisseria meningitidis*. *Genetics* 166:25–32.
- Avni A, Edelman M, Rachailovich I, Aviv D, Fluhr R. 1989. A point mutation in the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase affects holoenzyme assembly in *Nicotiana tabacum*. *EMBO J.* 8:1915–1918.
- Brock TD. 1967. Micro-organisms adapted to high temperatures. *Nature* 214:882–885.
- Castenholz RW. 1988. Culturing methods for cyanobacteria. *Methods Enzymol.* 167:68–93.
- Curmi PM, Cascio D, Sweet RM, Eisenberg D, Schreuder H. 1992. Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 2.0-Å resolution. *J Biol Chem.* 267:16980–16989.
- Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet.* 8:675–688.
- Fitch JH, Knight S, Andersson I, Brändén CI, McIntosh L. 1990. Residues in three conserved regions of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase are required for quaternary structure. *Proc Natl Acad Sci U S A.* 87:5768–5772.
- Genkov T, Spreitzer RJ. 2009. Highly conserved small subunit residues influence Rubisco large subunit catalysis. *J Biol Chem.* 284:30105–30112.
- Golding GB, Dean A. 1998. The structural basis of adaptation. *Mol Biol Evol.* 15:355–369.
- Haney PJ, Badger JH, Buldak GL, Reich CI, Woese CR, Olsen GJ. 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc Natl Acad Sci U S A.* 96:3578–3583.
- Harmon LJ, Matthews B, Des Roches S, Chase JM, Shurin JB, Schluter D. 2009. Evolutionary diversification in stickleback affects ecosystem functioning. *Nature* 458:1167–1170.
- Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics* 159:1805–1817.
- Knight S, Andersson I, Brändén C-I. 1990. Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution: subunit interactions and active site. *J Mol Biol.* 215:113–160.
- Kostov RV, Small CL, McFadden BA. 1997. Mutations in a sequence near the N-terminus of the small subunit alter the CO₂/O₂ specificity factor for ribulose bisphosphate carboxylase/oxygenase. *Photosynthesis Res.* 54:127–134.
- Krivov GC, Shapovalov MV, Dunbrack RL. 2009. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77:778–795.
- Laemmli U. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680–685.
- Lee B, Berka RM, Tabita FR. 1991. Mutations in the small subunit of cyanobacterial ribulose-1,5-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits. *J Biol Chem.* 266:7417–7422.
- Liu C, Young AL, Starling-Windhof A, et al. (12 co-authors). 2010. Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* 463:197–202.
- Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. 2000. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* 29:291–325.
- McVean G, Awadalla P, Fearnhead P. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* 160:1231–1241.
- Meeks JC, Castenholz RW. 1978. Photosynthetic properties of the extreme thermophile *Synechococcus lividus*. II. Stoichiometry between oxygen evolution and CO₂ assimilation. *J Thermal Biol.* 3:19–24.
- Miller SR, Castenholz RW. 2000. Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. *Appl Environ Microbiol.* 66:4222–4229.
- Miller SR, Strong AL, Jones KL, Ungerer MC. 2009. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl Environ Microbiol.* 75:4565–4572.
- Miller SR, Wingard CE, Castenholz RW. 1998. Effects of visible light and UV radiation on photosynthesis in a population of a hot spring cyanobacterium, a *Synechococcus* sp., subjected to high-temperature stress. *Appl Environ Microbiol.* 64:3893–3899.
- Millman KL, Tavaré S, Dean D. 2001. Recombination in the *ompA* gene but not the *omcB* gene of *Chlamydia* contributes to serovar-specific differences in tissue tropism, immune surveillance, and persistence of the organism. *J Bacteriol.* 183:5997–6008.
- Paul K, Morell MK, Andrews TJ. 1991. Mutations in the small subunit of ribulosebisphosphate carboxylase affect subunit binding and catalysis. *Biochemistry* 30:10019–10026.
- Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Qiu W-G, Schutzer SE, Bruno JF, Attie O, Xu Y, Dunn JJ, Fraser CM, Casjens SR, Luft BJ. 2004. Genetic exchange and plasmid transfers in *Borrelia burgdorferi sensu stricto* revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A.* 101:14150–14155.

- Saschenbrecker S, Bracher A, Rao KV, Rao BV, Hartl FU, Hayer-Hartl M. 2007. Structure and function of *rbcX*, an assembly chaperone for hexadecameric Rubisco. *Cell* 129:1189–1200.
- Smith SA, Tabita FR. 2003. Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Mol Biol* 331:557–569.
- Spratt BG, Bowler LD, Zhang Q-Y, Zhou J, Maynard Smith J. 1992. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J Mol Evol* 34:115–125.
- Spreitzer RJ, Esquivel MG, Du Y-C, McLaughlin PD. 2001. Alanine-scanning mutagenesis of the small-subunit β A- β B loop of chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase: substitution at Arg-71 affects thermal stability and CO₂/O₂ specificity. *Biochemistry* 40:5615–5621.
- Stitt M, Schulze D. 1994. Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. *Plant Cell Environ* 17:465–487.
- Storz JF, Wheat CW. 2010. Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* 64:2489–2509.
- Swofford DL. 1996. Paup 3.1.1. Sunderland (MA): Sinauer Associates.
- Tabita F, Caruso P, Whitman W. 1978. Facile assay of enzymes unique to the Calvin cycle in intact cells, with special reference to ribulose 1,5-bisphosphate carboxylase. *Anal Biochem* 84:462–472.
- Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. 2007. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev* 71:576–599.
- Thompson J, Higgins D, Gibson T. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
- Thornton JW. 2004. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat Rev Genet* 5:366–375.
- Wallner B, Elofsson A. 2005. All are not equal: a benchmark of different homology modeling programs. *Protein Sci* 14:1315–1327.
- Ward DM, Castenholz RW, Miller SR. 2012. Cyanobacteria in geothermal habitats. In: Whitton BA, editor. Ecology of cyanobacteria II: their diversity in space and time. Dordrecht (The Netherlands): Springer. p. 39–63.
- Ward DM, Ferris MJ, Nold SC, Bateson MM. 1998. A natural view of microbial diversity within hot spring cyanobacterial mat communities. *Microbiol Mol Biol Rev* 62:1353–1370.
- Watt WB, Dean AM. 2000. Molecular-functional studies of adaptive genetic variation in prokaryotes and eukaryotes. *Annu Rev Genet* 34:593–622.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Pop Biol* 7:256–276.
- Weinreich DM, Watson RA, Chao L. 2005. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* 59:1165–1174.
- Woelk CH, Li J, Holmes EC, Brown DWG. 2001. Immune and artificial selection in the hemagglutinin (h) glycoprotein of measles virus. *J Gen Virol* 82:2463–2474.
- Worobey M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria and mitochondria. *Mol Biol Evol* 18:1425–1434.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13:555–556.
- Yang Z, Kumar S, Nei M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* 141:1641–1650.
- Yang Z, Nielsen R. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917.