2022

An Integrative Investigation of the Synechococcus A/B Clade During Adaptive Radiation at the Upper Thermal Limit of Phototrophy

Christopher L. Pierpont
University of Montana, Missoula

AN INTEGRATIVE INVESTIGATION OF THE *SYNECHOCOCCUS* A/B CLADE DURING ADAPTIVE RADIATION AT THE UPPER THERMAL LIMIT OF PHOTOTROPHY

By

CHRISTOPHER LOGAN PIERPONT

B.S. Biology, Coastal Carolina University, Conway, SC, 2017
B.S. Marine Science, Coastal Carolina University, Conway, SC, 2017

Thesis

presented in partial fulfillment of the requirements
for the degree of

Master of Science
in Cellular Molecular and Microbial Biology,
Microbial Evolution and Ecology

The University of Montana
Missoula, MT

October 2022

Approved by:

Scott Whittenburg,
Graduate School Dean

Scott R Miller, Chair
Division of Biological Sciences

Brandon S Cooper
Division of Biological Sciences

Jared T Broddrick
NASA Ames Research Center

Pierpont, Christopher, M.S., October 2022 *Cellular, Molecular, and Microbial Biology*

AN INTEGRATIVE INVESTIGATION OF THE SYNECHOCOCCUS A/B CLADE DURING ADAPTIVE RADIATION AT THE UPPER THERMAL LIMIT OF PHOTOTROPHY

Chairperson: Scott R Miller, Ph.D.

Thermophilic microorganisms have been scientifically observed since the early nineteenth century and have spurred many questions about the limits of life and the capacity of organisms to survive extreme conditions. Decades of research on thermophile proteins and genomes have yielded several proposed correlates of temperature that may contribute to adaptation of bacteria and archaea to high temperature. However, many of the generalizations reported are drawn from analyses of deeply divergent taxa or from individual case studies in isolation from mesophilic relatives. Members of the *Synechococcus* A/B (SynAB) group are the only cyanobacteria with members able to grow above 65 °C and represent the most thermotolerant phototrophs on the planet. This group exhibits extensive variation in thermal performance and appears to represent a single adaptive radiation to colonize higher temperature environments—providing an ideal opportunity to test the relative importance of proposed mechanisms of the evolution of thermophily. I have established an unparalleled collection of SynAB strains and genomes from populations in Yellowstone NP and Oregon. Phylogenomics confirmed that lineages of *Synechococcus* that have diverged in thermotolerance have a unique, ancient origin, and physiological characterization corroborates a pattern of sequential adaptation to increasingly higher temperatures. During adaptation to higher temperatures, SynAB genomes have shrunk dramatically, and I argue that this is likely due to decreases in community complexity rather than selection for smaller cell size or faster growth. Proteome adaptation at the SynAB thermal limit has included the evolution of amino acid composition (most notably, the onset of aspartate phobia) and the acquisition of new proteins from distantly related bacteria. My work also establishes a framework to tackle longstanding questions about the relative contributions of thermodynamic constraints versus biochemical adaptation during the evolution of thermal physiology. To help spur the field of thermal biology in a new direction, I present a novel integration of genomic, physiological, and metabolic modelling approaches that enables exploration of how a cellular system, not just its constituent components, responds to the factors that contribute to the thermal limit of phototrophy.

Acknowledgments

The completion of this work during some of the most tumultuous and formative moments of my life would not have been possible without the incredible network of friends, family, and colleagues that surround me. To my family and friends on the East coast, your kinship has been a constant and priceless source of warmth and love. The science herein may be unrelatable or unimportant to you, but your support has never faltered, and I consider myself fortunate for that. To the diverse array of friends that I've made in Missoula, you have made the dark times (metaphorically and seasonally) bright and easy. Your friendship has enriched my life in enumerable ways, and I am all the more grounded for it. They say that you are an amalgamation of the people you spend the most time with; I don't know if that's true, but I would be perfectly content with the result if it was. To my cat Salem, your companionship and loyalty rivals that of any dog's, and though you'll never read these words, for you I dedicate a poem by the late Ursula K. Le Guin:

A paw, a questioning nose half waken me,
and I let him get under the covers.
He curls up and purrs himself asleep.
Cats are less troublesome than lovers.

To my PI Scott Miller, the patience, encouragement, and support you've shown me during my personal and professional quarter-life course corrections have been invaluable. I don't know whether many PIs can say the same when presented with similar situations, and they should be taking notes. You have deepened my love and fascination of microorganisms and broadened my understanding of the field of evolution. The foundation of my future career is made solid in part because of you. Finally, there are several people whose experimental help directly impacted the data presented in this thesis and deserve a thousand thanks: Tim Wheeler, for your help with Nanopore library preparation and sequencing; David Xing, for use of the UM Genomic Core facilities and equipment; October Moynahan, for use of the ECOR lab's UPRtek PAR200 quantum spectrometer; Jake Baroch, for your help with the nucleotide, amino acid, and TrmH analyses; and Nigel Li, for your help in generating the standard curves I needed for normalization of the PVI experiments.

Table of Contents

Introduction

Extreme environments, from deep-sea hydrothermal vents to heavy-metal toxic wastes, have been an area of intense scientific fascination for decades due to the diverse array of organisms which colonize them (Rampelotto, 2013). Characterizing the mechanisms that enable organisms to colonize such environments has important applications across the spectrum of biological research, such as understanding how life might adapt to anthropogenic stressors, optimizing the cellular processes utilized in industrial biotechnology, or hypothesizing about the origins of life from Earth's tumultuous geochemical past (Javaux, 2006; Rampelotto, 2013). For example, thermophilic microorganisms have been scientifically observed since the early nineteenth century, capturing the attention of Charles Darwin and his contemporaries (Hass, 2000). So perplexing and alien was the idea that a living thing could persist in near-boiling water, that some saw these microorganisms as proof that life came from another planet—including Nobel laureate Svante Arrhenius (Allen, 1953; Kamminga, 1982; Hollinger and Steiner Verlag, 2016). Ultimately, these fascinating microbes spurred many fundamental questions about the limits of life and the capacity of organisms to survive extreme conditions (Allen, 1953; Hass, 2000; Javaux, 2006; Rampelotto, 2013).

Though early studies on microbial thermophiles were only capable of ecological descriptions (Allen, 1953), the next 150 years of scientific advancements enabled rigorous comparative physiological and molecular research. These studies have revealed several apparent correlates of thermophily that may shed light on the mechanisms of temperature adaptation. These include genome size, nucleic acid GC content, protein amino acid composition, and horizontally acquired genes; I briefly discuss each below.

Genome size. There is a strong negative correlation between genome size and optimal growth temperature (OGT) in an analysis of >1,500 bacterial and archaeal genomes (Sabath et al., 2013). This is proposed to be the result of selection for smaller genomes (i.e., genome streamlining; Maniloff, 1996; Kuo et al., 2009) rather than the product of drift, based on a general decrease in the proportion of intergenic DNA in the genomes of thermophiles, as well as the absence of evidence for the relaxation of selective constraints on protein-coding genes. Smaller prokaryotes tend to have smaller genomes (Shuter et al., 1983), and it has been reported that cells of thermophiles are smaller than those of their mesophilic relatives (Lamanna, 1940; Allen, 1953). Because DNA can take up a substantial fraction of cell volume in small cells (Giovannoni et al., 2005), Sabath et al. (2013) have proposed that genome size reduction in thermophiles may actually be a byproduct of selection for smaller cell size to reduce higher maintenance costs associated with high temperature environments, such as macromolecule turnover (Kuhn et al., 1980; Stouthamer and Bettenhaussen, 1980) and enhanced lipid content to reduce proton leakage (Nordström and Laakso, 1992).

Nucleic acid base composition. Stability of dsDNA and RNA secondary structure is impacted by GC content, because the extra hydrogen bond of a GC pair increases its melting temperature compared with an AT pair (Wang et al., 2015). It has been hypothesized that thermophilic organisms must have higher genomic GC content to maintain nucleic acid structure at higher temperatures (Musto et al., 2005, 2006). While

this prediction holds true for some thermophiles, it is certainly not universal. For example, *Thermus thermophilus* has an OGT of ~65 °C with a GC content of 69% (Jiang et al., 2013; Wang et al., 2015), whereas *Caldicellulosiruptor hydrothermalis* has an OGT of ~70 °C but only has a GC content of 35% (Blumer-Schuette et al., 2011; Wang et al., 2015). Though an increase in genomic GC content with temperature is an attractive hypothesis on first principles, other factors contribute to the evolution of GC content, and other physiological mechanisms exist that help stabilize the double-helix structure in prokaryotes (Galtier and Lobry, 1997; Zeldovich et al., 2007). Despite the lack of observed *genome-wide* patterns of GC content, robust correlations between OGT and the GC content of structural RNAs have been reported. Galtier and Lobry (1997) found a clear relationship between GC content and OGT in tRNAs, 5S rRNAs, and the stems of 16S and 23S rRNAs for a sample of prokaryotic thermophiles. This was further supported by Saunders et al. (2003), who found increases in tRNA GC content for archaeal thermophiles above a 60 °C threshold. Dutta and Chaudhuri (2010) also found increased GC content in tRNA and rRNA genes with increased OGT of complete thermophilic archaeal and bacterial genomes. They also point out that RNA molecules, which have intrinsically lower melting temperatures than DNA, don't permanently reside in larger protein complexes; therefore, noncovalent interactions with neighboring molecules are not sufficiently persistent to help stabilize RNA secondary structure at higher temperatures (Dutta and Chaudhuri, 2010). Interestingly, the concept of "purine loading" has been observed by several studies, which report that nucleic acids of thermophilic organisms have higher AG content (Lambros et al., 2003; Basak et al., 2004; Paz et al., 2004; Zeldovich et al., 2007; Mahale et al., 2012; Wang et al., 2015). The number of tRNA genes coded within a genome has also been observed to significantly decrease with OGT (Dutta and Chaudhuri, 2010). This observation, as well as the concept of purine loading, may be due to the relationships between protein amino acid composition and OGT, described below (Singer and Hickey, 2003; Zeldovich et al., 2007; Dutta and Chaudhuri, 2010; Wang et al., 2015).

Protein amino acid composition. Many investigations of thermophiles have focused on how their proteins are able to remain folded and functional at elevated temperatures (Allen, 1953; Angilletta, 2009; Wang et al., 2015; Somero et al., 2017). Although there are many ways to stabilize a protein, comparative genomic and proteomic studies have identified apparent, strong correlations between OGT and the frequency of nonpolar, polar, and charged amino acids, respectively, that may generally be explained by protein folding theory.

The most important driver of protein folding involves the initial hydrophobic interactions of nonpolar amino acid side chains, forming a tight hydrophobic core within globular proteins that is separated from the surrounding polar solvent environment of the cytosol (Garrett and Grisham, 2017). For this reason, it's predicted that protein thermostability increases with an increase in core hydrophobicity, in part due to increased packing density and stronger hydrophobic interactions (Schumann et al., 1993). Studies on thermophilic microorganisms have indeed reported increases in the bulky nonpolar hydrophobic residues isoleucine (Ile), valine (Val), leucine (Leu), and tryptophan (Trp) (Saunders et al., 2003; Zeldovich et al., 2007).

Charged amino acid residues may also be important for stabilization at the protein surface through electrostatic interactions between opposite charges (Perutz and Raidt, 1975; Vogt et al., 1997; Saunders et al., 2003; Garrett and Grisham, 2017). Several studies on thermophilic microorganisms have reported increases in the charged residues glutamate (Glu), arginine (Arg), and lysine (Lys) (Kreil and Ouzounis, 2001; Tekaia et al., 2002; Saunders et al., 2003; Zeldovich et al., 2007). Decreases in the histidine (His) residue have been reported in thermophilic proteins (Kreil and Ouzounis, 2001), which may be due to its disruption of secondary structures when charged (Armstrong and Baldwin, 1993; Li and Hong, 2011), which would decrease protein packing density that would otherwise increase stability (Hurley and Weiner, 1992; Kumar et al., 2000).

Finally, polar uncharged amino acids are important contributors to hydrogen bonding (Garrett and Grisham, 2017), and it might be expected that increases in protein thermostability can involve increases in hydrogen bond networks (Perutz and Raidt, 1975; Vogt et al., 1997). However, several studies report overall *decreases* in the polar uncharged resides asparagine (Asn), glutamine (Gln), serine (Ser), and threonine (Thr) (Kreil and Ouzounis, 2001; Tekaia et al., 2002; Saunders et al., 2003; Zeldovich et al., 2007). This pattern may be due to 1) the spontaneous and conformationally disruptive deamidation of Asn and Gln (Li et al., 2010; Kato et al., 2020), and 2) the conserved nucleophilic role of Thr in protease catalysis (Kisselev et al., 2000)—reactions that would both more readily proceed at higher temperatures (Wright, 1991; Saunders et al., 2003). Although protein themostability frequently increases through an increase in weak interactions such as salt bridges between opposite charges, there are also reports that some thermophiles have a higher number of covalent disulfide bridges through the polar residue cysteine (Cys) (Wang et al., 2015), though it is unclear how prevalent this mechanism is across the diversity of thermophiles.

Taken together, Zeldovich et al. (2007) found that “IVYWREL” (isoleucine, valine, tyrosine, tryptophan, arginine, glutamate, and leucine) content is the best predictor of OGT for archaea. This set contains all chemical classes of amino acids, and they argue that natural selection tunes the proportion of this set to maintain the Boltzmann energetics of protein folding in the face of changes in temperature (Zeldovich et al., 2007).

Horizontally acquired genes. Horizontal gene transfer (HGT) between distantly related organisms has long been recognized as an important driver in the evolution of genomic novelty, particularly for prokaryotic organisms (Rivera et al., 1998; Koonin et al., 2001; West-Eberhard, 2003; van Wolferen et al., 2013). The colonization of higher temperature niches by mesophilic bacteria is no exception, and some argue that thermophilic bacteria might not exist at all if not for HGT (van Wolferen et al., 2013). In fact, it has been reported that two prominent thermophilic bacteria have obtained a significant fraction of their genomes from hyperthermophilic archaea— ~16% of the genes in *Aquifex aeolicus* (Aravind et al., 1998) and ~24% in *Thermotoga maritima* (Nelson et al., 1999)—many of which are implicated in conferring growth at high temperatures. One of the best documented examples is the enzyme reverse gyrase (Déclais et al., 2000; Forterre, 2002; van Wolferen et al., 2013; Wang et al., 2015). This enzyme falls within the topoisomerase protein family and confers temperature-dependent protection to the DNA double-helix through positive DNA supercoiling or some other currently unknown thermoprotective mechanism (Perugino et al., 2009; van Wolferen et

al., 2013). Phylogenetic analyses strongly support a complex history of multiple ancient acquisitions of reverse gyrase and is now found in virtually all thermophiles described to date (Brochier-Armanet and Forterre, 2006; Gribaldo and Brochier-Armanet, 2006; van Wolferen et al., 2013).

In summary, as thermophilic microorganisms adapt to higher temperature, it is generally predicted that 1) genome size decreases, 2) GC content of structural RNAs increases, 3) protein core hydrophobicity and surface charge increase, and 4) prevalence of horizontally acquired genes increases. These correlates are joined by broad ecological observations about how organisms from across the tree of life respond to changes in temperature; however, robust models of temperature adaptation have yet to be synthesized despite this growing body of empirical data (Zeldovich et al., 2007; Angilletta, 2009; Wang et al., 2015). Ironically, a parallel yet similar sentiment was expressed during the early investigations of thermophilic bacteria:

> "[...] interest in these microorganisms, especially in their more fundamental scientific aspects, has been oddly sporadic, with the result that much the same experiments have been done over and over again" (Allen, 1953).

Predictions regarding temperature adaptation are difficult to make in part because of the broad effects that temperature has across all levels of biological organization, from biomolecules to ecosystems (Somero et al., 2017). For example, biochemical reaction rates are unequivocally temperature dependent, but so too do they depend on enzyme binding affinities, which are subject to natural selection (Clarke and Fraser, 2004; Savage et al., 2004; Gillooly et al., 2006; Angilletta et al., 2010). Additionally, many correlates of temperature, such as those described above, are generalized from analyses of deeply divergent taxa (in some cases over more than 3.5 billion years of divergence from their common ancestor) or from individual case studies in isolation from mesophilic relatives. The relationship between temperature and physiology indeed depends on a complex interplay of biochemical and cellular factors (Allen, 1953; White and Seymour, 2003; Clarke, 2006; Kingsolver and Huey, 2008; Angilletta, 2009; van Wolferen et al., 2013; Schulte, 2015; Wang et al., 2015; Somero et al., 2017), and recently there have been calls from several branches of thermal biology research to move beyond classical approaches that are fundamentally limited in their ability to characterize this relationship (Wang et al., 2015; Kingsolver and Woods, 2016; Chen et al., 2017).

In this thesis I present a novel integration of evolutionary genomics, physiology, and systems biology approaches to investigate the mechanisms of temperature adaptation within the *Synechococcus* A/B (SynAB) group, an early branching clade of thermophilic cyanobacteria (Papke et al., 2003; Shih et al., 2013; Dvořák et al., 2014; Moore et al., 2019; Jasser et al., 2022). These photosynthetic microbes inhabit alkaline geothermal gradients throughout North American hot springs between ~50 and 73 °C (Ward et al., 2012). In fact, they are the most thermotolerant phototrophs on the planet, whose members include the only cyanobacteria capable of growth above 65 °C, the most thermotolerant of which can grow at ~73 °C (Falk et al., 1996; Miller et al., 1998; Ward et al., 2012). Members of this group exhibit extensive variation in thermal performance and appear to represent a single adaptive radiation to colonize higher temperature

environments (Miller and Castenholz, 2000; Allewalt et al., 2006). The mechanisms underlying the evolution of thermophily in SynAB are poorly understood, though case studies have provided specific insights. For example, Miller et al. (2013) found that the carbon-fixing enzyme RuBisCO of a member of SynAB capable of growth at 70 °C had a much higher melting temperature compared to its lower-temperature relatives; in addition, Pedersen and Miller (2017) show, for the same strain, that photosystem II and the light-harvesting phycobilisome complex are functionally inactivated at higher temperatures compared to ancestral relatives. This system thus provides an ideal opportunity to test the relative importance of proposed mechanisms of the evolution of thermophily. To that end, here I report: 1) the largest collection of SynAB strains and genome data; 2) the first genome-wide reconstruction of their evolutionary history; 3) physiological analyses regarding the evolution of thermotolerance in the group; 4) genomic comparisons of strains that have diverged in thermal performance; and, finally, 5) a metabolic modelling approach that incorporates temperature-dependent physiological constraints. Characterizing the mechanisms of temperature adaptation in this way will provide insights on the evolutionary limitations that prevent further adaptation to even higher temperatures and potentially inspire thermal biologists to reevaluate the ways in which temperature adaptation has been historically studied.

Methods

Collection of mat samples. Microbial mat samples were collected from along the outflow channels of alkaline geothermal hot springs in Yellowstone National Park, WY (June 2018, September 2018, and July 2019) and Hunter's Hot Springs, OR (August 2018). In Yellowstone, this included Rabbit Creek (Midway Geyser Basin) and several hot springs in the White Creek thermal area of the Lower Geyser Basin (White Creek, Octopus Spring, and Five Sisters). Yellowstone samples were collected under NPS permits YELL-2018-SCI-5482 and YELL-2019-SCI-5482. For all sample sites, the collection approach was as follows. A HANNA Instruments handheld temperature probe HI-93510N was used to periodically measure the *in situ* water temperature at the water-microbial mat interface until sites at or near 50, 55, 60, 65, and 70 °C were identified. At these locations, a sterile syringe was used to transfer ~10 mL of microbial mat to a sterile 15 mL conical centrifuge tube. During the 2019 Yellowstone field trip, samples were only collected from 60, 65, and 70 °C. For all collections, the *in situ* temperature of the site at the time of collection was recorded. Samples were stored at ambient temperature in the dark until they could be processed, no later than 48 hours after collection.

Isolation and growth of SynAB strains. Unless otherwise specified, medium D was used for all culturing and isolation procedures and was prepared following Castenholz (1988), but without the addition of selenium or nickel micronutrients. For all 2018 collections, mat samples were homogenized by vortex, and *Synechococcus* cell densities were then estimated with hemocytometer counts. From these suspensions, serial dilutions were prepared with medium D such that two sets of three 2 mL dilutions at 10^{-1}, 10^{-2}, and 10^{-3} cells mL^{-1} were generated for each collection. One set of dilutions was directly transferred to individual flasks of 75 mL media. For the other, each 2 mL dilution was vacuum filtered through a sterile 47 mm diameter 0.2 µm glass fiber filter using a Whatman glass filter apparatus, and then submerged in 75 mL media. The filter apparatus was flame sterilized between dilutions from different collections, but not between dilutions of the same collection, which instead were sequentially passed through the filter from lowest to highest concentration. To maximize representation of SynAB diversity in our culture collection, the 2019 Yellowstone collections were processed as described above, except that *six* sets of three 2 mL dilutions were generated so that each collection had one set of liquid dilutions and one set of filter dilutions incubating at 60, 65, and 70 °C, respectively. All flasks were incubated in a temperature-controlled Percival I36LL growth chamber at or near their source collection temperature under ~100 µmol photons m^{-2} s^{-1} of cool white fluorescent light with a 12h/12h photoperiod.

Flasks with glass fiber filters were monitored for the presence of cyanobacterial colonies, similar to microbial isolation methods using solid media plates. Colonies would generally become apparent to the naked eye after 1-2 weeks of incubation. At this time, filters were removed from their flask and viewed under a Leica MZ6 dissecting scope. Flame-sterilized forceps were used to transfer isolated colonies to individual flasks of 75 mL media. Phase-contrast light microscopy with a Leica DME compound microscope was used to assess growth and purity of liquid-dilution and filter-isolate flasks after ~1 month of growth. The dilution-filtration approach described above was repeated at least twice for all cultures, and unialgal cultures were chosen for further analysis. Cultures

generated by this approach were subjected to 16S rRNA screening to confirm the presence of SynAB. DNA was extracted using BioRad's InstaGene Matrix protocol, and the cyanobacterial 16S rRNA locus was amplified via PCR. Each reaction contained the following reagents: 1 X Green GoTaq Flexi Buffer, 4 mM $MgCl_2$, 0.1 mM of each dNTP, 0.4 µM forward primer CYA359F (5'-GGGGAATYTTCCGCAATGGG-3'; Nübel et al., 1997), 0.4 µM reverse primer PLG2.3R (5'-CTTCAYGYAGGCGAGTTGCAGC-3'; Miller and Castenholz, 2000), and 0.025 U GoTaq Flexi DNA polymerase. Cycling conditions included 50 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s using a ThermoFisher SimpliAmp thermal cycler. Products were resolved on a 1% agarose 1X TAE gel. Successful amplicons were purified using the Zymo Research DNA Clean & Concentrator Kit, then sequenced using GeneWiz's Sanger sequencing pipeline. Cultured strains with significant 16S rRNA BLASTn hits (e-values $\leq 1e^{-10}$) to members of the SynAB lineage within NCBI's BLAST database were kept for subsequent experiments and analyses. Strains generated with these approaches were kept in culture under the growth conditions described above and were transferred to fresh D medium monthly. Three previously isolated SynAB strains from Yellowstone (RC10A2, RC10B2, WC101) were also maintained under these growth conditions.

In addition to the above SynAB strains, cells of *Synechococcus* strain Nb3U1 were generously provided by Dr Satoshi Ohkubo (Tohoku University, Japan) and Dr Hideaki Miyashita (Kyoto University, Japan). This strain was isolated from a microbial mat sample collected at approximately 50 °C from Nakabusa hot spring, in Nagano, Japan (Ohkubo and Miyashita, 2017). 16S rRNA sequence analysis identified Nb3U1 as a member of "lineage T1" (Lau et al., 2009), a cosmopolitan but poorly understood group of cyanobacteria from mesothermic to moderately high temperature habitats in alkaline geothermal environments. Notably, T1 cyanobacteria appear to be the less thermotolerant sister taxa of the SynAB clade (Ward et al., 2012); therefore, strain Nb3U1 provided an excellent outgroup for phylogenetic and physiological analyses. Nb3U1 biomass was grown in liquid medium BG11 buffered with HEPES (pH 8.0) to a final concentration of 10 mM at 45 °C under the maintenance conditions described above.

Genome sequencing, assembly, and annotation. For all strains mentioned above, as well as for frozen cell pellets of strains previously isolated from Hunter's Hot Springs (OH20 and OH30; Miller and Castenholz, 2000) and one metagenomic sample (WC10meta; unpublished), genomic DNA was extracted from growing cells using Qiagen's DNeasy PowerBiofilm Kit according to manufacturer protocols. Prior to sequencing, DNA quality and quantity was assessed by spectrophotometry (Agilent Tape Station) and fluorimetry (Qubit 2.0), respectively. Sample libraries for paired-end, short-read sequencing were prepared with a Nextera DNA flex kit and followed by 150 cycles of sequencing on an Illumina NextSeq 550 platform. Quality of all sequenced reads was checked using FastQC v0.11.9 (Babraham Bioinformatics, 2019). Illumina adapter sequences were removed with Trimmomatic v0.36 (Bolger et al., 2014), but reads were not trimmed based on Phred quality scores; during initial design of the following assembly pipeline, it was found that draft assemblies were higher quality when reads were left untrimmed. An additional set of paired Illumina reads generated from a Yellowstone *Synechococcus* strain was obtained via the NCBI Sequence Read Archive (strain 65AY6A.5F, BioProject accession number PRJNA250890).

Initial draft genomes using these Illumina sequences were assembled *de novo* with SPAdes v3.12.0, which employs a de Bruijn graph-based approach to assemble sequenced reads into contigs (Prjibelski et al., 2020). Following assembly, contigs shorter than 1 kbp were removed. Because these cultures were not axenic, a refinement pipeline using Kraken v2.1.2 (Wood et al., 2019) and BLAST+ v2.2.31 (Camacho et al., 2009) was designed to identify and pull down contigs with high sequence similarity to cyanobacteria. Contigs were extracted from an assembly if: 1) the contig returned significant local BLASTn hits (query coverage ≥ 50%, and e-values ≤ $1e^{-100}$) to two SynAB reference genomes (Bhaya et al., 2007), and/or 2) Kraken assigned the contig an NCBI taxid from the Cyanobacteria phylum, the Synechococcales order, the Synechococcaceae family, the *Synechococcus* genus, or either of the above reference genomes. Following assembly, coverage histograms were generated to identify outlier contig populations; Bandage v0.8.1 (Wick et al., 2015) was used to remove these rogue contigs, and to estimate the mean coverage of the assembly. Genomic statistics on these extracted contigs were measured using QUAST v4.5 (Gurevich et al., 2013) (Table 2). Based on these statistics, draft assemblies were immediately removed from subsequent analyses if they failed two or more of the following criteria: 1) have a coverage of at least 10 X, 2) have an N50 of 10 kbp or greater, 3) have fewer than 1,000 contigs, and 4) have a total size between 2.5 and 3.5 Mbp. The overall completeness of all assemblies was measured using BUSO v5.2.2 (Manni et al., 2021) with reference to the Synechococcales taxonomic order.

Following phylogenetic analysis (see below), six strains with the best overall genome quality and culture purity statistics (based on microscopy, Kraken, QUAST, and BUSCO analyses) were chosen as representative members across the group's evolutionary history for improved genome assembly with long-read sequencing. High molecular weight genomic DNA was extracted from these strains as well as from the Nb3U1 outgroup using Qiagen's Genomic-tip 20/G kit and protocol. Sequencing libraries were prepared without shearing or size-selection using Nanopore's Ligation Sequencing kit, and samples were sequenced for 48 hours with a Nanopore MinION sequencer using a FLO-MIN106D flow cell with R9.4.1 chemistry. An additional round of nanopore sequencing was conducted on strains W70.1, W60.3, and H60.4 using the gDNA extracted from the Genomic-tip 20/G kit. GUPPY v4.5.4 (Ueno, 2003) was used for Nanopore basecalling. Hybrid genome assemblies were generated *de novo* with SPAdes v3.12.0 using both the Illumina short-reads and Nanopore long-reads for each representative. These hybrid assemblies were then subjected to the same refinement and quality assessment pipelines described above.

Additional SynAB genomes that were sequenced and assembled prior to work conducted for this thesis were obtained through NCBI's GenBank—including JA-2-3B'a(2-13) and JA-3-3Ab, the two reference genomes mentioned above. Annotation of all genome assemblies, generated here or obtained elsewhere, was initially conducted with RAST v2.0 (Aziz et al., 2008; Overbeek et al., 2014; Brettin et al., 2015), and these annotations were used for downstream phylogenetic and genome content analyses. Assemblies generated through this thesis were additionally annotated with NCBI's Prokaryotic Genome Annotation Pipeline (Tatusova et al., 2016; Haft et al., 2018; Li et al., 2021) upon submission to NCBI.

Phylogenomics. For these annotated genomes, 404 single-copy orthologous sequences were identified using OrthoFinder v2.5.4 (Emms and Kelly, 2015, 2019). These sequences were individually aligned using MUSCLE v5.1 (Edgar, 2022)**,** then concatenated into a single alignment for each genome using MEGAX v10.2.6 (Kumar et al., 2018). A maximum likelihood species phylogeny was constructed with IQtree v2.0.6 (Nguyen et al., 2015) for 118,301 aligned amino acid sites using the JTT+F+R3 model of sequence evolution identified as the best model by AIC/BIC with ModelFinder (Kalyaanamoorthy et al., 2017). The tree was outgroup rooted with strain Nb3U1. Ultrafast bootstrap replicate analysis (Hoang et al., 2018) and SH-like approximate likelihood ratio tests (Guindon et al., 2010) were performed with 1,000 replicates using IQtree's UFBoot. Additional analyses were used to estimate the degree of congruence between the concatenated species tree and individual gene trees. A gene tree can disagree with the species tree for either technical (e.g., lack of resolution) or biological (e.g., horizontal gene transfer) reasons. These mechanisms will go undetected in large datasets that only rely on standard bootstrap approaches to estimate support—i.e., random resampling of a genome-wide phylogeny may always return high support because sampling error will be low (Salichos and Rokas, 2013). Maximum likelihood gene trees for these analyses were individually constructed using IQtree and its ModelFinder on each of the aligned orthologs identified by OrthoFinder. The IQtree software package itself can assess genealogical concordance through a metric called the gene concordance factor (gCF), defined as the percentage of a set of gene trees that contain a given bipartition in the concatenated species tree (Minh et al., 2020). RAxML v8.2.10 (Stamatakis, 2014) was used to estimate another measure, internode certainty (IC). Unlike gCF, IC simultaneously considers not just the frequency of gene trees that support a given bipartition, but also the frequency of the most prevalent *conflicting* bipartition within those gene trees, as well as the total number of disparate conflicting bipartitions in the sample; effectively, it is a measure of genealogical entropy for a given node within a phylogeny (Salichos and Rokas, 2013).

Comparative genome content analyses. Pangenomic approaches and the distinction between core and accessory genomes brought significant insights for the field of bacterial genome evolution (Guimaraes et al., 2015). The core genome is the set of loci shared between all taxa of a given dataset, and represents those genes typically associated with essential aspects of cellular homeostasis (e.g., replication, transcription, and translation) and presumably experience high selective pressure for retention (Medini et al., 2005; Tettelin et al., 2005; Lapierre and Gogarten, 2009). Alternatively, the accessory genome is the set of loci found in some but not all taxa; for bacteria, these genes are often beneficial under certain circumstances but are not required for growth or reproduction, and so are frequently gained and lost (Lapierre and Gogarten, 2009; Mira et al., 2010). To probe how core gene content has changed during SynAB diversification, a comparative pipeline was designed using the pangenomic analysis software ROARY v3.12.0 (Page et al., 2015). ROARY uses an all-versus-all BLASTp search algorithm to generate a presence-absence matrix of all features within a set of input genome annotation files. This matrix can then be queried to determine the intersections between genomes, such as the core or pangenome of the group, as well as features that are unique to a single genome. ROARY was first used to identify the core genomes of each clade across the species phylogeny. For all clades except clade I, default ROARY parameters were used,

including a 95% amino acid identity BLASTp cutoff. For clade I, a 90% cutoff was deemed appropriate given the larger number of genomes and broader range of diversity sampled within that group. The presence-absence matrices of these six analyses were used to identify loci present in all members of each group. For each clade, core genome FASTA files were generated using the sequence identifiers for these loci from a randomly chosen strain and the sequence extraction tools of BLAST+ v2.2.31 (Camacho et al., 2009). To limit false-positive hits in the downstream ROARY analysis, only single-copy sequences were extracted using this approach. Duplicate core genes were noted, and follow-up analyses of duplication events were conducted (see below).

I next conducted a second ROARY analysis for the individual core genomes together with the genome of strain Nb3U1 (Pierpont et al., 2022). To account for the larger amount of divergence within this "core-versus-core" analysis, a 70% amino acid identity cutoff was used. This cutoff was chosen based on the results of a local tBLASTx search between the nucleotide sequences of the clade VI core genome (the most diverged SynAB), and those of the Nb3U1 genome (the most basal). From this second analysis, the presence-absence matrix of these genomes was used to identify those features only belonging to a single clade. To curate these results and identify biologically interesting differences among the clades, the nucleotide sequences of these putatively unique genes were subjected to a local BLASTx search against the individual amino acid annotation files of all SynAB genomes within the species phylogeny. The BLAST results for each of these loci were parsed, and hits were considered significant if their query coverage was ≥ 50% and had an e-value ≤ $1e^{-50}$. Because the e-value metric is dependent on both subject database size and query length (Camacho et al., 2009), in cases where no significant hits were found based on these criteria, the e-value cutoff was relaxed to approximately half the e-value of the most significant hit to genomes of the query's respective clade; this cutoff was never relaxed beyond $1e^{-10}$. These loci were then categorically binned using the following criteria: 1) "unique", if no significant hits across the phylogeny; 2) "pangenomic", if represented in at least one other clade and fails other criteria; 3) "paralogous core", if represented in all other clades and duplicated in the majority of hit members; 4) "divergent core", if represented in all other clades with an observed percent identity trend; 5) "copy variant", if represented in all other clades with evidence of copy number expansions/contractions during diversification; or 6) "candidate gain/loss", if represented in at least three sequential clades and no others.

During genome assembly and phylogenetic analyses, Kraken v2.1.2, BUSCO v5.2.2, and OrthoFinder v2.5.4 (Emms and Kelly, 2015, 2019; Wood et al., 2019; Manni et al., 2021) were used to refine genome assemblies, assess genome assembly completeness, and identify orthologous sequences, respectively. The capabilities of these programs also aided in the characterization of SynAB gene content. Though BUSCO is limited by comparisons to a chosen reference lineage (Synechococcales), both BUSCO and OrthoFinder include analyses of duplications within the input set of genome sequences, and BUSCO additionally provides information on pseudogenized or missing genes. During genome assembly refinement, Kraken was used to assign NCBI taxonomic IDs to individual contigs so that contaminating sequences could be filtered out. Here, Kraken was applied to the nucleotide sequence annotations of each of the core genomes to identify genes which do not map to cyanobacterial lineages. In conjunction with the above gene content analysis with ROARY, the results of these BUSCO, OrthoFinder, and

Kraken analyses helped inform identification of novel duplication and horizontal transfer events during SynAB diversification. During these genomic analyses, the genes *trmH* and *nifHDK* were identified as having interesting and relevant evolutionary histories. Sequences of *trmH* were gathered from members of the genus *Thermus* and for high temperature members of SynAB (clades V and VI). Maximum likelihood trees were reconstructed for a ClustalW alignment of these *trmH* genes with a TPM3+F+G4 model and 1000 bootstrap replicates with IQ-TREE (Nguyen et al., 2015). The model was selected by ModelFinder (Kalyaanamoorthy et al. 2017), and bootstrap analysis was performed by UFBoot (Hoang et al., 2017). The tree is outgroup-rooted with sequence data for *Meiothermus taiwanensis* WR-220 (NCBI GenBank accession CP021130.1). For *nifHDK*, a neighbor network analysis was conducted with a concatenated alignment of 3,888 nucleotides from clades I and IV using SplitsTree v4.14.4 (Huson and Bryant, 2006). Finally, for each genome, GC content of the genome and at the three codon positions of protein-coding genes, % non-coding DNA, codon usage, amino acid composition, and protein size distributions were determined using custom Python scripts.

Phylogenetic Generalized Least Squares (PGLS) and other comparative methods. Comparative phenotypic data collected from organisms may violate the assumption of statistical dependence due to their shared evolutionary histories. To account for this possibility when investigating the relationships among traits including GC content and amino acid composition, we took a PGLS approach using the "ape" (Paradis and Schliep, 2019) and "nlme" (Pinheiro and Bates, 2022) R packages. The variance-covariance matrix of the error term in each GLS model was obtained with ape from branch-length data for the *Synechococcus* phylogeny using the "corMartins" correlation structure. This estimates matrix elements by an Ornstein-Uhlenbeck process, in which traits evolve in a drift-like manner but are pulled toward a phenotypic optimum by a restraining force, the alpha parameter (Martins and Hansen, 1997). Alpha increases in strength the further the phenotype strays from the optimum, and this constraint on phenotypic evolution results in an exponential decay of phenotypic similarity with phylogenetic distance. Inclusion of the alpha parameter in the GLS model resulted in good statistical performance in computer simulations (Martins et al., 2002), irrespective of the true evolutionary process underlying phenotypic change. In all of our models, alpha was very high, indicating that phylogenetic correlation had been erased and observations were statistically independent.

Characterization of metabolic potential. To characterize and compare the metabolic potentials of these SynAB core genomes, functional annotation was performed using eggNOG-mapper v2.0 (Cantalapiedra et al., 2021). This software pipeline is built from several widely used genomic tools and databases (e.g., Pfam, KEGG, and SMART) (Letunic et al., 2021; Mistry et al., 2021; Kanehisa et al., 2022) and comprehensively describes an input set of amino acid sequences using a reference database of orthologous groups that includes over 7,500 organisms and viruses. Notably, for any annotated feature for which orthology cannot be assigned, the pipeline will attempt to build functional descriptions based on protein domain content within the sequence (Cantalapiedra et al., 2021). It is therefore a powerful tool in the characterization of non-model organisms, such as members of SynAB. All genome assemblies generated or acquired for this study

were uploaded to the online eggNOG-mapper platform and run using default parameters. The KEGG pathway identifiers for each locus in these genomes were then collected and uploaded to the online KEGG Mapper Reconstruct module (Kanehisa et al., 2022). Pathway representation was then compared by plotting gene count against collection temperature for candidate metabolic modules that appear to underly adaptive radiations to higher temperatures.

Growth experiments and thermal performance curves. Thermal performance curves for growth rate were characterized for a representative strain from each major SynAB clade identified in the genome-wide phylogeny and for outgroup strain Nb3U1. Starting at a strain's maintenance growth temperature, triplicate flasks containing 75 mL of fresh D (or BG11 medium, in the case of Nb3U1) were inoculated with growing cells to a final OD_{750} between 0.005 and 0.010 using a Beckman DU Series 500 spectrophotometer. Every 48 hours thereafter, OD_{750} was measured using 2 mL of homogenized culture from each replicate. Flasks were grown under ~100 μmol photons m^{-2} s^{-1} of cool white fluorescent light with a 12h/12h photoperiod, and measurements were collected for at least 3 generations of growth. The generation time during exponential growth was estimated by determining $\log_{10} 2/b$, where b is the slope of logarithmically transformed A750 data regressed on time (in hours). Negative growth rates were reported as a value of 0.0. This value was transformed and reported as number of generations per day, and growth rates were averaged across replicates. The temperature of each growth experiment was continuously measured at 60 s intervals using HOBO MX Temp Logger pendants. For accurate temperature monitoring of liquid media, each logger was coated in Corning vacuum grease, sealed inside nitrile sleeves, and submerged in individual glass jars of dH_2O that were deployed alongside experimental flasks. This procedure was sequentially repeated at approximately 5 °C increments above and below the starting temperature until no growth was observed (e.g., cell bleaching, negative/stagnant growth rates). Biomass of a randomly selected replicate from the previous treatment was used to inoculate the replicate flasks of the following treatment. In certain cases, particularly towards the upper and lower limits of a strain's growth range, replicate flasks were pelleted via low-grade centrifugation (5,500 x g for 30 minutes) and resuspended in 10 mL of the appropriate media as a single, concentrated stock to obtain sufficient cell density for inoculations.

Preliminary metabolic modelling. Within the last decade, advances in computational methods have provided new ways to integrate genome-scale datasets, flux-balance analyses, and constraint-based modelling to describe and predict the metabolic capabilities of cells (Ebrahim et al., 2013; Heirendt et al., 2019). These approaches expand upon orthology-based metabolic descriptions by incorporating quantitatively measured constraints on the movement of cellular resources through the metabolic network—i.e., they describe not just metabolic potential but simulate metabolism under physiological conditions. Genome-scale constraint-based modelling has yielded robust models of phototrophic metabolism (Nogales et al., 2012; Beck et al., 2017; Gudmundsson et al., 2017) including for a well-studied member of SynAB (Ebrahim et al., 2013; Broddrick et al., 2016, 2019). Future work will utilize these approaches to simulate photoautotrophic growth for representative SynAB strains that have diverged in

thermal performance. Parameterization of these models will include common-garden, quantitative proxies of each strain's photosynthetic performance at sub-optimal, optimal, and supra-optimal temperatures (i.e., rates of oxygen evolution and carbon assimilation), and will incorporate inferred thermodynamic descriptions of enzyme structure. Effectively, this will allow prediction of cellular resource flux through metabolic sectors across each strain's TPC. Comparisons of these predictions will provide valuable insights on 1) how the metabolic system of SynAB has changed during adaptation to higher temperatures and 2) why higher temperature clades exhibit decreased optimal performance compared to their lower temperature ancestors (Figure 3).

For this study, a proof of concept of these methods was conducted using strains W60.1 (clade I) and W70.1 (clade VI) at 55 °C. The workflow first involves generation of draft metabolic models for each strain based on homology to a reference. Draft metabolic models of W60.1 and W70.1 were generated using their genome assemblies and the BiGG reference model of JA-2-3B'a(2-13) (Broddrick et al., unpublished) with the constraint-based reconstruction and analysis pipeline COBRApy v0.23.0 (Ebrahim et al., 2013). Protein structural thermodynamics were incorporated using ssbio v0.9.9.1 (Mih et al., 2018). Growth simulations of these models were parameterized using oxygen evolution and photon uptake constraints generated for each strain at 55 °C using photosynthesis versus irradiance (PVI) curves (described below) and following Broddrick et al. (2019). The biomass composition was assumed to be the same as previous modeling in *Synechococcus elongatus* PCC 7942 adapted to low light with a mass of 1.3 pg $cell^{-1}$, an inoculation density of $2x10^7$ cells per mL, and a total culture volume of 75 mL. The photon flux was set to 80 µmol m^{-2} s^{-1} with a culture light exposed surface area of 18 cm^2. Growth was simulated in 20-minute pseudo-steady state intervals for a total of 12 hours. The growth rate was determined by taking the log of the change in cell count over the growth interval (units of h^{-1}). Growth rates were converted to generations per day for comparison with the existing experimental growth data.

Standard curves relating cell density (cells mL^{-1}) to OD at 750 nm were generated for each of these strains. Sample cultures were allowed to grow under the normal growth conditions described above for approximately three weeks. Culture aliquots were gently homogenized using a flame-sterilized, glass Dounce homogenizer, pelleted via low-grade centrifugation (5,500 x g for 30 minutes), and resuspended in 1 mL D media to a final cell density of 1.0 x 10^7 cells mL^{-1}. Following Broddrick et al. 2019, a custom apparatus was designed to hold a cuvette within a temperature-controlled water bath aligned with the fiber optic measurement cable of a Walz miniPAM chlorophyll fluorometer and a Unisense OX-Eddy Clark-type oxygen sensor. Sample was allowed to acclimate to the temperature of the water bath and kept in the dark for at least 10 minutes prior to experimental measurements; sensor signal collected during this time was used to account for respiration rates during the modelling. The Light Curve (LC) program of the miniPAM was used to measure photosynthetic fluorescence yields after a rapid pulse of saturating light, followed by sequential 1-minute exposures to the following range of quantum fluxes (PPFD): 1, 2, 8, 36, 55, 86, 125, and 219 µmol photons m^{-2} s^{-1}. To prevent the settling of cells during the procedure, the sample was carefully and slowly homogenized using a sterile needle syringe before the start light curve program. The oxygen sensor signal was recorded every second during the LC using the Unisense SensorSuite Logger software. The PPFD and spectral distribution of each of the 8 LC

light treatments was quantified by averaging across three separate LCs using a UPRtek PAR200 quantum spectrometer; for both these measurements and during the experimental LCs, the fiber optic cable of the miniPAM was positioned 5 mm from the spectrometer's measuring surface and the surface of the sample cuvette, respectively. This process was repeated for each experimental temperature.

Prior to experimental procedures, the oxygen sensor was calibrated following manufacturer's instructions for conversion from raw sensor signal (mV) to oxygen concentration (µmol O_2 L^{-1}). For each of the 8 LC light treatments, rate of oxygen evolution (µmol O_2 L^{-1} s^{-1}) was estimated by the steepest slope of the most stable increase of oxygen over 5-10 seconds, then normalized to cell density (µmol O_2 $cell^{-1}$ s^{-1}) and fit to the Platt equation to determine the QF-dependent change in oxygen evolution.

$$P = P_{max} * \left(1 - e^{-\alpha * \frac{QF}{P_{max}}}\right) * \left(e^{-\beta * \frac{QF}{P_{max}}}\right)$$

Where P is the photosynthesis rate in µmol O_2 $cell^{-1}$ s^{-1}, P_{max} is the maximum photosynthetic rate, α and β are Platt parameters describing the initial slope and photoinhibition, respectively, and QF is the quantum flux in µmol $cell^{-1}$ s^{-1}. The light harvesting complexes of photosynthetic organisms only capture a fraction of the visible spectrum of light (350-800 nm). To account for this, miniPAM irradiance measurements were converted to photon uptake rate (µmol photons $cell^{-1}$ s^{-1}) using the wavelength-specific absorption coefficient (a^*_{cell}; Moore et al. 1995) of the samples. Homogenized cultures from above were used to prepare three 2 mL dilutions at 1:1, 1:2, and 1:4 ratios. These dilutions were transferred to 8 mL of D media, then individually vacuum-filtered through a sterile 47 mm diameter 0.2 µm glass fiber filter using a Whatman glass filter apparatus; an additional filter was prepared using 10 mL D media. All filters were then placed on top of a 96-well plate with plate cover so that at least 4 wells were covered by the cell filtrate or media blank. The absorbance spectra (350-800 nm) from each filter were collected for each well using a BioTek SynergyHT microplate reader, averaged, and blank subtracted using the measurements from the D media sample. Spectrometer filter amplification was corrected for using the coefficients for *Synechococcus* WH103 (Moore et al., 1995). The density of cells deposited to each filter was calculated by dividing the circular area of the deposited filtrate (using a diameter of 3.5 cm) by the total number of cells. a^*_{cell} (cm^2 $cell^{-1}$) was calculated for each dilution by multiplying the corrected absorbance measurements at each wavelength by the density of cells found above; because raw absorbance measurements were linear with cell count, a^*_{cell} was averaged across the dilutions. Finally, excitation spectra of both strains were collected to probe electron transfer efficiency through PSII across a range of temperatures using a PTI model QM-7/2005 rapid temperature change spectrofluorometer equipped with a Quantum North-west TLC 50 thermoelectric temperature-controlled cuvette holder. Cells were excited at 1 nm intervals between 437 and 675 nm light, capturing emission of chlorophyll *a* at 685 nm. Fluorescence was measured beginning at 25 °C and then over 5 °C increments until the temperature at which complete inhibition of PSII was reached.

Results and Discussion

Genome data for the largest collection of SynAB strains. Our understanding of SynAB evolutionary history and functional diversity is primarily based on investigations of only a small number of laboratory strains and genetic loci (Miller and Castenholz, 2000; Allewalt et al., 2006; Miller et al., 2009; Inskeep et al., 2013; Nowack et al., 2015). The first two complete genomes generated from member strains of this group were published in 2007 (Bhaya et al., 2007) and were later joined by three draft assemblies in 2015 (Olsen et al., 2015), together providing new genomic details about the physiology of these organisms. For this study, I have established an unparalleled collection of SynAB laboratory strains and their genomes from populations in Yellowstone National Park, WY and Hunter's Hot Springs, OR (Table 1), as well as the first genome sequence of a member of *Synechococcus* sp. T1 clade, the sister taxon of SynAB (Ward et al., 2012; Ohkubo and Miyashita, 2017; Pierpont et al., 2022). In addition to the previously published genome data, my sample includes draft genome assemblies for 44 new laboratory strains isolated from my 2018 and 2019 mat collections, as well as 18 strains generated from previous work (Miller and Castenholz, 2000; Miller and Carvey, 2019; unpublished; Table 1). Summary statistics for draft genome assemblies and annotations are reported in Table 2 and Table 3, respectively. Most genomes appear to be complete based on BUSCO analysis, PGAP annotations, and comparison with the closed genomes reported by Bhaya et al. (2007). This collection thus provides a powerful sampling of diversity to probe longstanding questions about the evolution of thermal physiology.

The evolutionary history of the most extreme phototrophs on Earth. SynAB cyanobacteria include the most thermotolerant phototrophs on the planet: the more thermotolerant members of the clade are the only cyanobacteria able to grow above 65 °C, and some can grow above 70 °C (Ward et al., 2012). However, it has remained unresolved whether lineages of *Synechococcus* that have diverged in thermotolerance have a unique, ancient origin with subsequent dispersal, or, alternatively, whether there have been independent adaptive radiations to higher temperatures within geographically distinct populations of Western North America. This is because previous phylogenetic analyses of these bacteria have been marked by limited sampling of their geographic and/or ecological diversity, as well as largely restricted to the use of 16S rRNA gene sequences (Ferris and Ward, 1997; Miller and Castenholz, 2000; Ward et al., 2012; Olsen et al., 2015). To address the issue of SynAB evolutionary history, I reconstructed a maximum likelihood phylogeny for the collection of 68 *Synechococcus* genomes using a concatenated alignment of 118,301 amino acid sites derived from 404 single-copy orthologous gene sequences (Figure 1). The strains included in these analyses were isolated from Yellowstone NP, WY and Hunter's Hot Spring, OR. If different populations of *Synechococcus* have convergently evolved increased thermotolerance, strains would be expected to group by their source location. However, the six identified clades I-VI instead sort with respect to collection temperature rather than by geography; strains from cooler environments are more basal in the phylogeny, with clades from increasingly hotter environments nested within them (Figure 1). These clades can be further clustered into three groups based on thermal niche, with distinct YNP and OR clades within each: clades I/II from the coolest temperatures, III/IV from moderately high

temperatures and V/VI from the highest temperatures. Clade I *Synechococcus* correspond to the "B-lineage" of Ferris and Ward (1998) that appears to be absent from Oregon hot springs (Becraft et al., 2020; Figure 1). Clade II *Synechococcus*, which occupy a similar thermal niche as clade I, are the closest relatives of the more thermotolerant "A-lineage" clades III-VI.

Clades I-VI are generally well-supported by bootstrap analysis, gene tree concordance, and internode certainty values (Figure 1), but support is much weaker within clades. This is likely due to factors including recombination within clades, which has previously been demonstrated to occur at high rates among members of clade I (*Synechococcus* B) populations (Rosen et al., 2015; Miller and Carvey, 2019) and will generate discord among gene trees. In addition, the low gene concordance and internode certainty values for deeper splits in the phylogeny indicate a more complicated history for the relationships among clades than implied by a single concatenated tree. For example, clades II and III and clades III and V (both sets are from different thermal niches in Oregon) are sister taxa in 22% and 15% of gene trees, respectively. Mechanisms that could generate these patterns include continued gene flow by HGT between parapatric populations of SynAB that have diverged in thermotolerance and/or the differential fixation of alleles by descendant lineages for ancestrally polymorphic loci. Indeed, one example of the former appears to be the operon encoding nitrogen fixation genes. This set of genes is present in the outgroup strain Nb3U1, all members of clade I, and approximately half of the members of clade IV; it is missing from all other SynAB. Neighbor network analysis reveals that there has been potentially two instances of HGT at this locus between members of clade I and clade IV (Figure 2). Remarkably, this explains an observation by Bhaya et al. (2007), who report very little synteny between the complete genomes of clade I strain JA-2-3B'a(2-13) and clade IV strain JA-3-3Ab, except for a ~32 kbp region encoding the genes involved in nitrogen fixation. Future work will investigate the contributions of these different mechanisms to discord between some gene trees and the species phylogeny in Figure 1.

Finally, the phylogeny also reveals that SynAB evolutionary rate is positively associated with increasing temperature (i.e., the amount of amino acid substitutions since sharing a common ancestor is greater for lineages from higher temperature environments). Future work will investigate the possible mechanisms that contribute to this pattern (e.g., reduced effective population size and/or higher mutation rates at higher temperatures). However, because more thermotolerant lineages do not necessarily grow faster than less thermotolerant strains (see below), this observation is not due to faster generation times at higher temperatures.

Overall, these results indicate that the SynAB group diversified prior to dispersal throughout western North America. One potential dispersal mechanism may have involved the Yellowstone hot spot, which arose 45-56 MYA off the coast of northwestern California and Oregon during the subduction of several tectonic plates by the westward-advancing North American plate (Camp and Wells, 2021). The hot spot has since moved eastward to its current location as it was overtaken, inducing a tumultuous history of volcanic activity along its transect through eastern Oregon and southern Idaho (Smith and Braile, 1994; Christiansen et al., 2002; Camp and Wells, 2021; Staisch et al., 2021). An attractive speculative model would posit that the ancestral SynAB population arose and diversified in the ancient geothermal landscape of eastern Oregon (Camp and Wells,

2021); through short-range, wind- or animal-mediated dispersal events (Brock et al., 1969; Bonheyo et al., 2005; Cohan and Perry, 2007; Becraft et al., 2020), ecologically divergent members of SynAB could successively colonize new geothermal springs as they arose, following the Yellowstone hot Spot to its current location.

Increased thermotolerance with the loss of ancestral perforsance. Based on the environmental temperatures of sample collections (Figure 1), I expected that the thermotolerances of OR/YNP clade pairs I/II, III/IV and V/VI should respectively resemble each other. To investigate this, I assayed growth rate over a range of temperatures for representative strains from each clade. As predicted, strains form three groups based on well-described metrics of thermal performance: the minimum temperature of growth, CT_{min}, the optimal temperature of growth, T_{opt}, and the maximum temperature of growth, CT_{max} (Angilletta, 2009). These data indicate that there has indeed been a shift in the thermal limits of SynAB growth during diversification. Not only have members of SynAB evolved greater thermotolerance than the outgroup strain Nb3U1, CT_{min} increased from ~24 °C in clades I/II up to ~50 °C for clades V/VI; though not as dramatic, CT_{max} increased from 67 °C for clades I/II to greater than 70 °C for clades V/VI. Changes in T_{opt} were also observed, where clades I-III exhibited maximal performance near ~55 °C which increased to ~60 °C in clade IV and lies between 60 and 65 °C for clades V/VI (Figure 3). These results corroborate previous reports that increasingly thermotolerant strains branch later in SynAB phylogenies (Miller and Castenholz, 2000; Allewalt et al., 2006; Ward et al., 2012; Miller and Carvey, 2019); however, the use of genome data in this study renders my comparison much more robust. In conclusion, the repeated pattern of paraphyly of less thermotolerant SynAB (Figure 1) reveals an evolutionary trajectory of sequential adaptation to increasingly higher temperatures along these geothermal gradients; it likewise provides insights on the identity and genetic composition of the ancestors from which more thermotolerant descendants arose (see below).

Genome size shrinks with thermotolerance. We observed a strong negative correlation between genome size and environmental temperature of sample collection (R = -0.78; $F_{1,45}$ = 68.2, P < 0.0001 for a PGLS model; Figure 3). Mean genome size for highly thermotolerant clade V/VI strains was 80% of that of clade I (mean ± SE: 2.4 ± 0.03 versus 3.0 ± 0.03 Mbp). However, we can reject the cell size hypothesis of Sabath et al. (2013), because genome size reduction was not associated with a decrease in *Synechococcus* cell size (data not shown but available upon request). Alternatively, we propose a variant of the environmental stability hypothesis called the community complexity and environmental heterogeneity (CCEH) hypothesis to explain the smaller genomes of more thermotolerant *Synechococcus*. CCEH emphasizes the importance of the diversity of the biotic component of the environment for the amount of variation in conditions that an organism may experience: that is, the richness and metabolic diversity of a community impact the quality, amount, and cycling of resources in the environment, the nature of competition, cooperation, and communication among organisms, and the formation and temporal heterogeneity of chemical gradients associated with microbial metabolic activity. As discussed below, we argue that the loss of phylogenetic and

metabolic diversity from these communities with increasing temperature has contributed to the evolution of smaller genomes in more thermotolerant *Synechococcus*.

Productivity, diversity, and complexity of the microbial food web in these hot spring microbial communities are highly temperature-dependent. Peak primary production is observed for the thick, laminated microbial mats found at temperatures between about 55 to 61 °C (Revsbech and Ward, 1984), whereas *Synechococcus*-dominated biofilms near the thermal limit for photosynthesis are thin and temperature-stressed (Miller et al., 1998). There is a corresponding reduction in taxon richness along this productivity gradient (Miller et al., 2009). For two Yellowstone locations sampled in this study (White Creek, Rabbit Creek), community richness (measured for unique rRNA barcodes by the Chao1 estimator) decreased by ~16 taxa per 5°C increase in temperature to an observed minimum of fewer than 60 taxa at the thermal limit (Miller et al., 2009). This loss of diversity at the highest temperatures includes the absence of microorganisms responsible for particular metabolic processes that occur at high rates at lower temperatures, e.g., bacteria involved in sulfate reduction like *Thermodesulfovibrio* (Dillon et al., 2007) and the recently described "*Candidatus* Thermonerobacter thiotrophicus" (Bacteroidetes; (Thiel et al., 2019)). Consequently, communities at temperatures below about 65 °C are marked by a more complex web of metabolic interactions (Klatt et al., 2013), which includes the complete anaerobic decomposition of organic matter by acetogenesis and other fermentation pathways (Anderson et al., 1987; Klatt et al., 2013), sulfate reduction (Dillon et al. 2007) and methanogenesis (Ward, 1978). Rates of these anaerobic processes are generally highest near the mat surface (e.g., Ward, 1978; Dillon et al., 2007); bacteria with different physiologies are therefore mixed within the mat matrix and subject to extensive environmental heterogeneity (Stewart and Franklin, 2008). This includes dramatic fluctuations in chemical gradients of oxygen and sulfide arising from diurnal patterns of oxygenic photosynthesis by *Synechococcus* (Revsbech and Ward, 1984; Dillon et al., 2007).

SynAB genome size decreases with temperature and community complexity. Does the reduction in environmental complexity arising from the loss of phylogenetic and metabolic diversity with increasing temperature help explain the general observation of SynAB genome reduction? To explore the CCEH hypothesis proposed above, eggNOG-mapper v2.0 (Cantalapiedra et al., 2021) and the KEGG pathway reconstruction (Kanehisa et al., 2022) service was used to identify metabolic modules that have changed in composition during the adaptive radiation of the SynAB group. 178 KEGG pathway maps were analyzed with this approach, which included a set of 11 maps that summarize the total number of KEGG orthologs identified within broadly defined metabolic groups (e.g., a global map for "biosynthesis of cofactors" versus a specific map for a single cofactor) (Table A2). The metabolic potential of SynAB has shrunk during adaptation to higher temperatures, with marked reductions in secondary metabolite biosynthesis, quorum sensing, porphyrin metabolism (including B12 synthesis), nitrogen metabolism, and ABC transporters (Figure 4).

The largest observed decrease was the loss of ~30 genes classified as ATP-binding cassette (ABC) transporters. These are a ubiquitous family of transporters involved in the active uptake or excretion of substrates (Rees et al., 2009). The largest fraction of ABC transporters lost during SynAB diversification were nitrate/nitrite

transporters—clade I has 8 copies of this transporter, but clades V/VI only have 4. Taken together with the observation that there is more ammonium in alkaline hot springs at high (>60 °C) temperatures compared to nitrate/nitrite (Holloway et al., 2011; Ward et al., 2012), it may be that selection for retention of nitrate transporters was relaxed as these lineages encountered more reduced environments. Related to this observation, is the loss of nitrogenase genes in clades II, III, V, and VI. As discussed above, biofilm thickness decreases with increasing temperature (Miller et al., 1998). It is likely that high temperature communities rarely encounter anoxic conditions because of decreased community stratification, which would hinder their ability to fix nitrogen due to the oxygen-sensitivity of the nitrogenase enzyme. Overall this suggests that high temperature members of SynAB seem to have specialized on ammonium as their source of nitrogen.

Quorum sensing is an important aspect of bacterial populations that allows inter- and intra-species communication through the release of enzymes, toxins, effector proteins, and other signaling molecules, and is innately tied to the regulation of competition strategies, stress responses, and cell maintenance (Pena et al., 2019). During SynAB divergence, clades III-VI have lost a putative polyamine transport system and clades V/VI have lost an oligopeptide transport system. Homologs of these transport systems are ubiquitous and functionally diverse (Lessard and Walsh, 1999; Thomas and Thomas, 2001); however, it's clear that both are important metabolic regulators under starvation conditions, environmental stressors, and changes in community composition (Lessard and Walsh, 1999; Lee et al., 2004; Karatan et al., 2005; McGinnis et al., 2009). Without robust functional studies of these systems in *Synechococcus*, it is difficult to speculate what their purpose has been within SynAB, but future transcriptomic studies could help illuminate their function and test hypotheses relating to the importance of quorum sensing at higher temperatures.

The loss of gene content described above likely reflects the deletional bias of bacterial genomes (Mira et al., 2001). Although selection to maintain particular ancestral genetic pathways has been relaxed at higher temperatures, it is not clear whether the subsequent fixation of deletions is primarily due to drift or is selectively favored. Like host-associated endosymbionts and obligate pathogens, extremely thermotolerant *Synechococcus* live in a less heterogeneous environment than their close relatives. However, unlike the case for bacterial endosymbiont populations, which typically experience severe bottlenecks during host reproduction, we might expect drift to be much weaker compared with selection in these geothermal environments. Still, our analyses of certain genome characteristics suggest the possible importance of drift. For example, one prediction for selectively-favored genome streamlining is for the proportion of intergenic DNA to decrease with genome size; instead, we observed no clear trend (range: 11.8-13.7%). We also did not observe a negative association between genome size and cell division rate (Figure 2; Table A4), which is predicted if small genomes are selectively favored for more rapid reproduction (Mira et al., 2001). A possible role for selection may be resolved with planned future analyses of the degree of selective constraint on protein-coding genes in the different *Synechococcus* clades. Typically, endosymbionts experience faster protein evolution due to the increased probability of the fixation of deleterious mutations resulting from their low effective population size (Moran and Wernegreen, 2000); if drift is stronger in more thermotolerant SynAB, then we expect to observe a similar phenomenon (higher d_N/d_S).

Few gene content differences underlie variation in SynAB thermotolerance. From the above analysis of metabolic pathways, it's clear that more thermotolerant lineages of *Synechococcus* have shed genes as they've adapted to higher temperatures. Are there specific differences among clades that might underlie increased thermotolerance? To answer this question, the core genome for each clade in the SynAB phylogeny (Figure 1) was identified and then the intersections among them were determined. Overall, 280 loci are shared between the entire SynAB phylogeny and the outgroup Nb3U1 (Figure 5). Fewer than 50 loci are unique to individual clades of SynAB, and less than 20 are unique to each of the three thermotolerance groups (I/II, III/IV, V/VI). Many of these unique loci were annotated as hypothetical proteins by RAST and were not assigned a functional description by eggNOG; individual clades and clade intersections generally had three or fewer unique loci with functional annotations. In fact, all of the unique genes within the clade III/IV intersection and within clade V were annotated as hypothetical. The larger intersection between clades II-VI and III-VI were also probed to identify genes shared only by the A-type lineages involved in the adaptive radiation, but none were identified. Many of the unique hypothetical proteins had fewer than 100 amino acids; consequently, we cannot be certain that they are in fact protein-coding genes. Future transcriptomic and proteomic analyses with this system will resolve whether these sequences are even expressed. Peptides of this size have been assumed nonfunctional by modern proteomic approaches, but recent studies on this significantly understudied area of protein biology suggest that they may play critical roles in organismal function (Steinberg and Koch, 2021). These analyses also revealed several orthologs that have greatly diverged in amino acid identity, including an RNA polymerase sigma-70 factor and the enzyme TsaE, involved in tRNA stability and translational fidelity (Missoury et al., 2018). Future analyses of d_N/d_S will help resolve whether this divergence is due to selection for these and other interesting candidates.

HGT: a small but potentially important role in SynAB adaptation. It has been reported that some thermophilic bacteria harbor a significant number of genes obtained through HGT from hyperthermophilic archaea (Aravind et al., 1998; Nelson et al., 1999), and even argued that thermophilic bacteria may not exist without such HGT events (van Wolferen et al., 2013). There are few functionally annotated, unique genes that distinguish SynAB thermotolerance groups and individual clades; fewer have involved HGT events from distantly-related donors. Despite this, there has been at least one HGT event that is strongly associated with *Synechococcus* temperature adaptation. The *Synechococcus* V/VI ancestor obtained a copy of *trmH* from a *Thermus* bacterium (Figure 6A; the *Synechococcus* proteins are 88.1-88.7% identical to *T. aquaticus*). TrmH is a methyltransferase that methylates G18 in the D-arm of tRNAs (Swinehart and Jackman, 2015). Post-transcriptional modifications of tRNAs can play an important role in the temperature adaptation of thermophiles (Lorenz et al., 2017), and methylation of G18 may stabilize the D-arm, overall tRNA shape and the tRNA-protein interaction (Kim et al., 1974; Kawai et al., 1992; Hori et al., 2002). *Thermus* (and *Synechococcus* V/VI) TrmHs are type I TrmH enzymes that can modify all tRNA species (Ochi et al., 2013), and its activity is thermally induced in *Thermus thermophilus* (Kumagai et al., 1980), implying its role in tRNA stabilization at high temperature. Following HGT, TrmH

evolution has largely been constrained by strong purifying selection (e.g., dN/dS = 0.039 for the ancestral branch of the *Synechococcus* copies); consequently, there has been little change in nucleotide usage at first and second codon positions (Figure 6B). By contrast, the *Synechococcus* sequences have diverged greatly from *Thermus* at third codon sites: whereas GC3 is >90% in *Thermus* strains, it varies between ~70-75% in *Synechococcus* (Figure 6B). As a result of this reduction in GC3 in *Synechococcus*, GC content for *trmH* (58-59%) resembles that of *Synechococcus* genomes as a whole (see below). We conclude that HGT has not made as quantitively important a contribution to temperature adaptation as has been reported for other bacterial hyperthermophiles, but this contribution may nonetheless have been crucial.

Changes in GC content are not associated with increased thermotolerance. The relationship between nucleic acid GC content and optimal growth temperature has long been debated. The additional hydrogen bond within GC base pairs increases the melting temperature of dsDNA (Wang et al. 2015), and therefore it has been hypothesized that thermophilic organisms must have higher genomic GC content to maintain nucleic acid structure (Musto et al., 2005, 2006). Similar to other studies (Galtier and Lobry, 1997; Zeldovich et al. 2007; Dutta and Chaudhuri 2010), I did not observe a general increase in GC content during SynAB temperature adaptation. Rather, clade I genome-wide GC (58.5%; 95% CI = 58.4,58.6) was slightly lower than that of the other clades, all of which were approximately 60% (all CIs overlap 60%) and did not differ significantly from each other ($F_{1,32} = 0.19$, $P = 0.66$). Similarly, neither GC3, AG content nor GC of RNAs are associated with temperature adaptation (not shown). Clades I and II occupy similar thermal habitats (Fig. 1) and have not diverged in thermotolerance (Fig. 2). The divergence in GC content that we observe between them therefore did not impact thermotolerance in a clear way. However, we cannot rule out that higher GC (or correlates such as codon usage or amino acid composition) in the clade II-VI ancestor predisposed subsequent adaptation to higher temperatures. This could have occurred, for example, through the shifts in amino acid composition associated with a change in GC (see below).

Evolution of amino acid composition. Because GC content did not change during adaptation to higher temperatures within Clades II-VI, any observed changes in amino acid composition during SynAB temperature adaptation cannot be the result of changes in GC. The evolution of amino acid composition during *Synechococcus* diversification was not as simple as the predicted increases in both charged and bulky aliphatic residues. First, we did not observe a general increase in charged amino acids (Asp, Glu, Arg, Lys) with increased thermotolerance. Rather, only Clade VI *Synechococcus* exhibited a subtle increase in the frequency of these residues (20.9%) compared with other clades (20.6-20.7%; all adjusted *P* values in Tukey HSD tests were less than 0.002 for pairwise comparisons between Clade VI and other clades). Similarly, clades have not diverged in the frequency of the bulky aliphatic amino acids isoleucine, leucine, valine, and methionine (26.0% of residues in Clade I *Synechococcus* B strains versus 26.1% in extremely thermotolerant Clade V/VI *Synechococcus* A strains). However, we do see a decrease overall in polar uncharged (NQST): 18.6% in Clade I, 18.0% in Clades V/VI.

We next took a multivariate approach to better understand how the composition of specific amino acids has evolved during *Synechococcus* diversification. The relationships among amino acid variables could generally be fit well by linear models and are therefore well-suited for principal components analysis (PCA). The first two principal components explained most (> 80%) of the variation in the data. Variables strongly associated with PC1 (64.3% of variance) included both positive and negative correlates of environmental temperature (percentage of glutamate, arginine, and proline for the former; aspartate, isoleucine, methionine and serine for the latter). Based on pairwise comparisons among clades by Tukey HSD tests, the composition of several of these (leucine, isoleucine, methionine, phenylalanine and serine) clustered into three groups separated along PC1, consisting of clades I, II/III/IV, and V/VI, respectively. By contrast, the strongest correlates of PC2 (16.1%) exhibited different patterns. These included % lysine (which is lowest for Clades II/III/IV), % valine and % tyrosine (which increase only in the most thermotolerant clades V/VI), and % glutamine and % glycine (which decrease in the most thermotolerant clades). A discriminant analysis with stepwise variable addition could assign all *Synechococcus* strains to the correct clade (Entropy $R^2 = 0.98$) based just on the three variables with the greatest *F* ratios in ANCOVA tests: % aspartate ($F = 296.2$, $P = 0$), % alanine ($F = 59.0$, $P = 0$), and % valine ($F = 35.8$, $P = 0$).

To address which specific changes in amino acid composition are most strongly associated with increased thermotolerance, we focused on clades II-VI, which have diverged in CT_{max} but do not exhibit potentially confounding differences in GC content (see above). Most notably, there was a marked reduction in the use of aspartate with increasing environmental temperature, particularly for the more thermotolerant strains of clades III-VI isolated from samples collected at or above 62 °C ($R^2 = 0.73$; $F_{1,31} = 82.1$, $P < 0.0001$; slope = -0.014 % Asp per °C). This is in accord with the recent proposal that extreme thermophiles exhibit "aspartate phobia" to avoid the deleterious consequences of temperature-dependent protein damage at aspartate residues (Villain et al., 2022). Asp is particularly prone to the spontaneous hydrolysis of peptide bonds (Partridge and Davis, 1950; Inglis, 1983), which results in the irreversible truncation and impaired function of proteins (Ahern and Klibanov, 1985; Zale and Klibanov, 1986); Asp isomerization is also a major contributor to protein degradation (Geiger and Clarke, 1987; Capasso et al., 1993; Cacia et al., 1996). Loss of aspartate in *Synechococcus* proteins is mirrored by similar gains in glutamate ($R^2 = 0.80$; $F_{1,31} = 125.7$, $P < 0.0001$), which increases at the same rate as aspartate declines (0.017 % Glu per °C). This indicates that aspartate has been replaced by glutamate in more thermotolerant lineages, thereby potentially mitigating protein damage while maintaining net charge. In addition to being more rapid at higher temperature, the rates of these Asp modifications have been shown to vary with the identity of the carboxyl adjacent amino acid in the protein: Asp-Gly is most prone to degradation, while larger residues appear to sterically interfere with the formation of reaction intermediates (Geiger and Clarke, 1987; Sydow et al., 2014). In agreement with these observations, we observed a striking decrease in the probability of Gly following Asp in more thermotolerant strains (7.3% in clade I vs. 6.9 % in clades V/VI) compared with the proteome-wide frequency of Gly (7.8 vs. 7.7%); by contrast, Leu (the most common amino acid to follow Asp) increased in frequency at the same rate as the proteome (14.2% in clade I vs. 14.8 % in V/VI). This suggests that natural selection is

acting in multiple ways to reduce the rate of Asp modification in the proteins of more thermotolerant *Synechococcus*.

Preliminary metabolic modelling provides promising insights. Flux balance analysis (FBA) of phototrophic genome-scale models has an established set of parameters that can successfully predict intracellular metabolism with a limited set of parameters (Nogales et al., 2012; Ebrahim et al. 2013; Broddrick et al., 2016; Beck et al. 2017; Gudmundsson et al., 2017; Broddrick et al., 2019). This set consists of the whole cell absorption spectrum (optical absorption cross-section), the oxygen evolution rate (Figure 11), and the irradiance spectrum of the experimental set up. This approach will be applied to the SynAB system to probe 1) how the metabolic system of SynAB has changed during adaptation to higher temperatures and 2) why higher temperature clades exhibit decreased optimal performance compared to their lower temperature ancestors (Figure 3). Future work will incorporate sub-optimal, optimal, and supra-optimal temperature conditions for representative members across the SynAB phylogeny, effectively allowing prediction of cellular resource flux through metabolic sectors across each strain's TPC. Here an initial proof of concept was designed at 55 °C for W60.1 (clade I) and W70.1 (clade VI), SynAB members at opposite ends of the thermotolerance spectrum.

Using the electron transport efficiency coefficients, we compared the fraction of available PAR that can be routed to PSII for the PVI curve determination versus the light used for growth experiments. Combining the irradiance spectrum and the cellular absorption spectrum, for a PPFD irradiance of 100 µmol photons m^{-2} s^{-1}, cultures under growth conditions capture 18% more light than the aliquot under the PVI conditions (8.8 versus 7.4 µmol photons m^{-2} s^{-1}, respectively). However, when the efficiency coefficient from the action spectrum was applied, the cultures under the growth conditions could route 49% *less* light to PSII than cultures under the PVI conditions (2.6 versus 5.0 µmol photons m^{-2} s^{-1}, respectively). We incorporated this correction by reducing the oxygen evolution constraint on the model to 51% of the value derived from the PVI curve. Future work will include a more mechanistic inclusion of these data into the modelling construct.

After correcting for exciton transport efficiency, the model predicted growth rates of 3.0 generations day^{-1} and 1.4 generations day^{-1} for W60.1 and W70.1, respectively, at 55 °C. Though the result that W60.1 grows approximately twice as fast as W70.1 is consistent with experimental results (Figure 3), the model greatly overestimated growth for both strains. There are several possible reasons for this observation. First, these simulations assume an axenic, planktonic culture. The quantity of excreted carbon used to sustain the microbial population that supplements the cofactor needs of these species may account for the difference in predicted versus observed growth rates. Incorporation of approximate, non-SynAB biomass in these models will help with this. Second, the current model doesn't incorporate maintenance costs. Incorporating dark respiration rate will begin to account for the difference in predicted versus observed growth rates. Experiments are underway that will quantify carbon assimilation rates for these strains under PVI temperatures.

Overall, the decrease in optimal performance for W70.1 (Figure 3) doesn't appear to be an issue with excitation energy transport efficiency, as 70 °C showed a near-optimal action spectrum (Figure 12). Thus, quantifying maintenance energy cost would be

informative when trying to ascribe a mechanism behind the reduced growth rate of W70.1. To this end, experiments are underway to quantify carbon assimilation rates at desired PVI temperatures for these and other strains. An additional observation is the W70.1 strain has roughly the same optical absorption cross section in the phycocyanin portion of the spectrum (~620 nm) but nearly twice the chlorophyll (~685 nm) and carotenoid (~425 nm) cross section, compared to W60.1 (data not shown but available upon request). Most of the chlorophyll in the cyanobacterial photosynthetic apparatus is in photosystem I. It is possible this strain maintains a higher pool of photosynthetic complexes compared to W60.1; potentially, high temperature strains of SynAB may keep a larger pool of these complexes since a larger fraction of them will presumably be undergoing repair and not participating in light harvesting. This would increase the per cell biosynthetic cost to build and maintain these complexes. Currently, the model assumes the biomass compositions between the strains are identical.

Conclusions

Using phylogenomic approaches, I have confirmed that lineages of *Synechococcus* that have diverged in thermotolerance have a unique, ancient origin and were subsequently dispersed across Western North America (Figure 1), and characterization of thermal performance curves supports a pattern of sequential adaptation to increasingly higher temperatures (Figure 2). Genome size does indeed shrink with increasing thermotolerance (Figure 4) but does not follow the cell size prediction of Sabath et al. (2013) because genome size reduction was not associated with a decrease in *Synechococcus* cell size. Instead, we hypothesize that the smaller genomes of more thermotolerant SynAB can be explained by the reduction of community complexity at higher temperatures, like the genomes of host-associated endosymbionts and obligate pathogens. Changes in GC content are not associated with increased thermotolerance at the genomic or RNA level. Though slight increases in GC content may have predisposed the clade II-VI ancestor to subsequent temperature adaptation, it does not appear to have been a crucial driver of continued increases in thermotolerance. Rather, stabilization of tRNAs through other mechanisms may have played a larger role during SynAB adaptation at the upper thermal limit, given the acquisition of TrmH by the ancestor of clades V and VI (Figure 6), and the putatively adaptive divergence of TsaE. The evolution of amino acid composition during SynAB diversification was not as simple as the predicted increases in both charged and bulky aliphatic residues, however there was a notable reduction in the use of aspartate with increasing environmental temperature (Figure 10). This observation is supported by the recent proposal that extreme thermophiles exhibit "aspartate phobia" to avoid the deleterious consequences of temperature-dependent protein damage at aspartate residues (Villain et al., 2022). Overall, there are relatively few unique genes that distinguish SynAB thermotolerance groups and individual clades. Further, there have been far fewer HGT events during SynAB adaptation compared to reports that up to 24% of a thermophilic bacterial genome consists of horizontally acquired genes from archaea (Nelson et al. 1999). Though HGT events with distantly related taxa were not identified in SynAB, exchange of genetic material with other bacterial community members (Figure 2)—including between *Synechococcus* that occupy different but overlapping thermal niches (Figure 6)—has been involved in the evolution of the SynAB genome.

The work presented in this thesis provides a wealth of new genomic and physiological insights on a group of thermophilic cyanobacteria that have been studied since the 1960s. Not only do these data provide a more robust analysis of the mechanisms underlying thermal adaptation in bacteria, but they also establish a framework that can be utilized to probe long-standing questions about the relative contributions of thermodynamic constraints versus biochemical adaptation in shaping the evolution of thermal physiology (Angilletta 2010). Finally, there have been calls from several branches of thermal biology research to move beyond classical approaches that are fundamentally limited in their ability to describe the effects of temperature on physiology (Wang et al., 2015; Kingsolver and Woods, 2016; Chen et al., 2017). To answer this call and help spur the field in a new direction, I present a novel integration of genomic, physiological, and computational approaches that enables exploration of how a cellular system, not just its constituent components, responds to the effects of temperature.

Table 1. Collection metadata of all study strains.

Strain name	Geographic location	Source name	Temperature (°C)	Reported by [a]	Status [b]
Nb3U1	Nakabusa, Nagano, Japan	Nakabusa Hot Spring	49.9	Ohkubo & Miyashita 2017	Active
R50.1	Yellowstone NP, WY, USA	Rabbit Creek	50.3	This study	Active
R3-13	Yellowstone NP, WY, USA	Rabbit Creek	51.0	Miller & Carvey 2018	–
R55.1	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.2	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.3	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.4	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.5	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.6	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	Active
R55.7	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
R55.8	Yellowstone NP, WY, USA	Rabbit Creek	54.0	This study	–
OH2	Lakeview, OR, USA	Hunter's Hot Spring	55.0	Miller & Castenholz 2000	–
OH20	Lakeview, OR, USA	Hunter's Hot Spring	55.0	Miller & Castenholz 2000	–
R5-12	Yellowstone NP, WY, USA	Rabbit Creek	55.0	Miller & Carvey 2018	–
R5-13	Yellowstone NP, WY, USA	Rabbit Creek	55.0	Miller & Carvey 2018	–
R5-15	Yellowstone NP, WY, USA	Rabbit Creek	55.0	Miller & Carvey 2018	–
R5-16	Yellowstone NP, WY, USA	Rabbit Creek	55.0	Miller & Carvey 2018	–
H55.1	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.2	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.4	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.5	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.6	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.7	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.8	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.9	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	–
H55.10	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	Active
H55.11	Lakeview, OR, USA	Hunter's Hot Spring	55.6	This study	Active

a Publication in which strain was first reported

b Indication of whether the strain is being actively maintained in culture; dash = inactive/unknown

Table 1 (continued). Collection metadata of all study strains.

Strain name	Geographic location	Source name	Temperature (°C)	Reported by [a]	Status [b]
JA-2-3B'a(2-13)	Yellowstone NP, WY, USA	Octopus Spring	56.0	Bhaya et al. 2007	–
W55.1	Yellowstone NP, WY, USA	White Creek	56.5	This study	Active
W55.2	Yellowstone NP, WY, USA	White Creek	56.5	This study	–
R60.1	Yellowstone NP, WY, USA	Rabbit Creek	58.7	This study	Active
R60.2	Yellowstone NP, WY, USA	Rabbit Creek	58.7	This study	–
R60.3	Yellowstone NP, WY, USA	Rabbit Creek	58.7	This study	Active
60AY4M2	Yellowstone NP, WY, USA	Mushroom Spring	60.0	Olsen et al. 2015	–
R6-5	Yellowstone NP, WY, USA	Rabbit Creek	61.0	Miller & Carvey 2018	–
R6-6	Yellowstone NP, WY, USA	Rabbit Creek	61.0	Miller & Carvey 2018	–
R6-7	Yellowstone NP, WY, USA	Rabbit Creek	61.0	Miller & Carvey 2018	–
R6-10	Yellowstone NP, WY, USA	Rabbit Creek	61.0	Miller & Carvey 2018	–
W60.1	Yellowstone NP, WY, USA	White Creek	61.3	This study	Active
JA-3-3Ab	Yellowstone NP, WY, USA	Octopus Spring	61.5	Bhaya et al. 2007	–
H60.1	Lakeview, OR, USA	Hunter's Hot Spring	62.0	This study	–
H60.2	Lakeview, OR, USA	Hunter's Hot Spring	62.0	This study	–
H60.3	Lakeview, OR, USA	Hunter's Hot Spring	62.0	This study	–
H60.4	Lakeview, OR, USA	Hunter's Hot Spring	62.0	This study	Active
R60.4	Yellowstone NP, WY, USA	Rabbit Creek	62.7	This study	–
W60.2	Yellowstone NP, WY, USA	White Creek	62.8	This study	Active
B60.1	Lakeview, OR, USA	Hunter's Hot Spring	63.4	This study	Active
B60.2	Lakeview, OR, USA	Hunter's Hot Spring	63.4	This study	–
R8-2	Yellowstone NP, WY, USA	Rabbit Creek	63.5	Miller & Carvey 2018	–
R65.1	Yellowstone NP, WY, USA	Rabbit Creek	64.7	This study	Active
W60.3	Yellowstone NP, WY, USA	White Creek	64.8	This study	Active
65AY6A.5F	Yellowstone NP, WY, USA	Mushroom Spring	65.0	BioProject PRJNA250890	–
65AY6A5	Yellowstone NP, WY, USA	Mushroom Spring	65.0	Olsen et al. 2015	–
65AY6Li	Yellowstone NP, WY, USA	Mushroom Spring	65.0	Olsen et al. 2015	–

a Publication in which strain was first reported

b Indication of whether the strain is being actively maintained in culture; dash = inactive/unknown

Table 1 (continued). Collection metadata of all study strains.

Strain name	Geographic location	Source name	Temperature (°C)	Reported by [a]	Status [b]
H65.1	Lakeview, OR, USA	Hunter's Hot Spring	65.1	This study	–
W65.1	Yellowstone NP, WY, USA	White Creek	66.1	This study	Active
H70.1	Lakeview, OR, USA	Hunter's Hot Spring	67.5	This study	Active
H70.2	Lakeview, OR, USA	Hunter's Hot Spring	67.5	This study	Active
RC10A2	Yellowstone NP, WY, USA	Rabbit Creek	69.0	Miller Lab	–
RC10B2	Yellowstone NP, WY, USA	Rabbit Creek	69.0	Miller Lab	–
R70.1	Yellowstone NP, WY, USA	Rabbit Creek	69.9	This study	Active
OH30	Lakeview, OR, USA	Hunter's Hot Spring	70.0	Miller & Castenholz 2000	–
F70.1	Yellowstone NP, WY, USA	Five Sisters Springs	70.4	This study	–
O70.2	Yellowstone NP, WY, USA	Octopus Spring	71.6	This study	–
W70.1	Yellowstone NP, WY, USA	White Creek	71.7	This study	Active
O70.1	Yellowstone NP, WY, USA	Octopus Spring	72.1	This study	–
WC101	Yellowstone NP, WY, USA	White Creek	72.5	Miller Lab	–
WC10meta	Yellowstone NP, WY, USA	White Creek	72.5	Miller Lab	–

a Publication in which strain was first reported

b Indication of whether the strain is being actively maintained in culture; dash = inactive/unknown

Table 2. QUAST genome statistics and BUSCO scores.

Strain name	Coverage	N50 (bp)	№ contigs	Assembly size (bp)	BUSCO notation [a]
Nb3U1	30.0	326,094	25	3,504,380	C:93.3% [S:99.0%, D:1.0%], F:0.4%, M:5.3%, n:791
R50.1	20.5	44,184	84	2,744,089	C:87.9% [S:87.8%,D:0.1%], F:0.4%, M:11.7%, n:788
R3-13	10.5	8,213	530	2,834,688	C:88.6% [S:100.0%, D:0.0%], F:3.6%, M:7.9%, n:787
R55.1	12.8	6,564	651	2,873,751	C:82.9% [S:82.5%,D:0.4%], F:4.1%, M:13.0%, n:788
R55.2	17.0	25,000	218	2,761,692	C:91.8% [S:99.0%, D:1.0%], F:1.4%, M:5.8%, n:791
R55.3	24.6	28,866	165	3,036,368	C:93.8% [S:93.7%, D:0.1%], F:0.5%, M:5.7%, n:788
R55.4	39.8	43,419	92	2,980,801	C:93.6% [S:93.5%, D:0.1%], F:0.3%, M:6.1%, n:788
R55.5	24.6	34,070	123	2,911,736	C:93.0% [S:92.9%, D:0.1%], F:0.3%, M:6.7%, n:788
R55.6	39.7	42,396	96	2,955,413	C:93.2% [S:93.1%, D:0.1%], F:0.3%, M:6.5%, n:788
R55.7	50.5	10,180	533	2,707,366	C:87.4% [S:99.7%, D:0.3%], F:2.4%, M:9.9%, n:788
R55.8	41.5	19,895	199	2,619,785	C:86.0% [S:99.7%, D:0.3%], F:0.9%, M:12.8%, n:788
OH2	220.0	39,262	126	2,734,292	C:87.7% [S:99.7%, D:0.3%], F:3.8%, M:8.2%, n:788
OH20	50.0	92,216	61	2,845,005	C:94.3% [S:99.7%, D:0.3%], F:0.3%, M:5.2%, n:788
R5-12	28.5	15,925	305	3,079,860	C:93.3% [S:99.0%, D:1.0%], F:0.4%, M:5.3%, n:791
R5-13	34.5	22,294	219	2,931,709	C:94.0% [S:99.7%, D:0.3%], F:0.6%, M:5.1%, n:788
R5-15	29.5	23,659	200	2,901,212	C:94.2% [S:99.7%, D:0.3%], F:0.3%, M:5.3%, n:788
R5-16	29.5	23,569	201	2,910,342	C:93.9% [S:99.7%, D:0.3%], F:0.6%, M:5.2%, n:788
H55.1	46.2	98,780	52	2,761,084	C:93.8% [S:93.7%, D:0.1%], F:0.4%, M:5.8%, n:788
H55.2	59.8	20,449	246	2,756,440	C:93.5% [S:99.7%, D:0.3%], F:0.8%, M:5.5%, n:788
H55.4	49.4	51,425	105	2,826,811	C:92.6% [S:98.7%, D:1.3%], F:0.5%, M:5.7%, n:792
H55.5	60.0	108,873	47	2,728,066	C:91.2% [S:99.7%, D:0.3%], F:0.1%, M:8.4%, n:788
H55.6	70.5	113,044	52	2,799,196	C:94.3% [S:99.7%, D:0.3%], F:0.1%, M:5.3%, n:788
H55.7	45.3	115,727	38	2,739,010	C:93.4% [S:93.3%, D:0.1%], F:0.1%, M:6.5%, n:788
H55.8	35.5	84,108	58	2,727,731	C:90.7% [S:99.7%, D:0.3%], F:0.1%, M:8.9%, n:788
H55.9	19.6	23,503	214	2,702,316	C:89.6% [S:89.5%, D:0.1%], F:2.2%, M:8.2%, n:788
H55.10	25.5	45,408	316	3,128,005	C:92.3% [S:97.7%, D:2.3%], F:0.3%, M:5.2%, n:796
H55.11	32.5	18,474	363	2,964,952	C:93.5% [S:99.7%, D:0.3%], F:0.6%, M:5.6%, n:788

[a] BUSCO notation quickly summarizes counts of benchmarking universal single-copy orthologs, here in reference to the Synechococcales order. C = complete, S = single-copy, D = duplicated, F = fragmented/pseudogenized, M = missing, n = total counts (Manni et al. 2021)

Table 2 (continued). QUAST genome statistics and BUSCO scores.

Strain name	Coverage	N50 (bp)	№ contigs	Assembly size (bp)	BUSCO notation [a]
JA-2-3B'a(2-13)	NR [b]	3,046,682	1	3,046,682	C:94.3% [S:99.5%, D:0.5%], F:0.3%, M:4.9%, n:789
W55.1	20.9	108,188	48	3,000,410	C:94.5% [S:94.4%, D:0.1%], F:0.3%, M:5.2%, n:788
W55.2	10.7	38,967	122	2,914,768	C:94.6% [S:94.3%, D:0.3%], F:0.4%, M:5.0%, n:788
R60.1	21.5	28,000	162	2,905,532	C:94.4% [S:94.3%, D:0.1%], F:0.4%, M:5.2%, n:788
R60.2	12.3	8,141	452	2,556,584	C:80.4% [S:80.3%, D:0.1%], F:2.8%, M:16.8%, n:788
R60.3	44.5	21,382	328	3,086,673	C:93.8% [S:99.2%, D:0.8%], F:0.4%, M:5.1%, n:790
60AY4M2	35.0	3,142,301	6	3,162,818	C:93.1% [S:98.7%, D:1.3%], F:0.1%, M:5.5%, n:793
R6-5	55.0	13,258	413	2,883,433	C:93.4% [S:99.7%, D:0.3%], F:1.0%, M:5.3%, n:788
R6-6	21.5	6,737	618	2,801,882	C:85.9% [S:99.7%, D:0.3%], F:5.2%, M:8.6%, n:788
R6-7	24.5	7,741	551	2,804,242	C:88.8% [S:99.7%, D:0.3%], F:4.1%, M:6.9%, n:788
R6-10	12.5	6,826	764	2,994,040	C:86.9% [S:99.7%, D:0.3%], F:6.1%, M:6.7%, n:788
W60.1	7.82	174,744	15	2,396,226	C:76.0% [S:75.9%, D:0.1%], F:1.1%, M:22.9%, n:788
JA-3-3Ab	NR [b]	2,932,766	1	2,932,766	C:92.2% [S:97.7%, D:2.3%], F:0.1%, M:5.4%, n:796
H60.1	35.5	22,664	292	2,811,816	C:92.9% [S:99.0%, D:1.0%], F:0.3%, M:5.8%, n:791
H60.2	9.89	31,653	146	2,709,982	C:93.6% [S:93.5%, D:0.1%], F:0.6%, M:5.8%, n:788
H60.3	9.13	20,553	216	2,666,742	C:92.6% [S:92.5%, D:0.1%], F:1.1%, M:6.3%, n:788
H60.4	24.1	120,468	38	2,798,208	C:94.0% [S:93.9%, D:0.1%], F:0.3%, M:5.7%, n:788
R60.4	17.5	26,183	168	2,652,680	C:91.0% [S:90.6%, D:0.4%], F:1.5%, M:7.5%, n:788
W60.2	11.1	17,123	234	2,575,073	C:88.0% [S:85.8%, D:2.2%], F:2.5%, M:9.5%, n:788
B60.1	16.2	21,885	227	2,755,066	C:92.9% [S:92.5%, D:0.4%], F:1.0%, M:6.1%, n:788
B60.2	15.7	44,633	87	2,689,978	C:94.3% [S:94.2%, D:0.1%], F:0.3%, M:5.4%, n:788
R8-2	47.5	5,312	775	2,809,343	C:86.2% [S:99.7%, D:0.3%], F:5.5%, M:8.1%, n:788
R65.1	11.7	91,742	40	2,260,421	C:91.8% [S:91.5%, D:0.3%], F:0.8%, M:7.4%, n:788
W60.3	12.6	67,926	85	2,634,779	C:84.0% [S:81.0%, D:3.0%], F:0.9%, M:15.1%, n:788
65AY6A.5F	1,682.3	58,733	88	2,852,378	C:90.2% [S:95.5%, D:4.5%], F:0.4%, M:5.0%, n:805
65AY6A5	22.0	2,508,234	9	2,981,827	C:93.2% [S:99.0%, D:1.0%], F:0.3%, M:5.6%, n:791
65AY6Li	16.0	2,795,989	2	2,933,219	C:92.1% [S:97.7%, D:2.3%], F:0.1%, M:5.5%, n:796

[a] BUSCO notation quickly summarizes counts of benchmarking universal single-copy orthologs, here in reference to the Synechococcales order.
C = complete, S = single-copy, D = duplicated, F = fragmented/pseudogenized, M = missing, n = total counts (Manni et al. 2021)
[b] NR = not reported

Table 2 (continued). QUAST genome statistics and BUSCO scores.

Strain name	Coverage	N50 (bp)	№ contigs	Assembly size (bp)	BUSCO notation [a]
H65.1	22.9	179,977	20	2,370,985	C:92.7% [S:92.6%, D:0.1%], F:0.1%, M:7.2%, n:788
W65.1	51.2	130,261	31	2,236,887	C:92.0% [S:91.9%, D:0.1%], F:0.5%, M:7.5%, n:788
H70.1	13.5	34,803	191	2,463,748	C:90.5% [S:98.7%, D:1.3%], F:0.9%, M:7.3%, n:792
H70.2	9.74	239,022	19	2,347,479	C:92.8% [S:92.4%, D:0.4%], F:0.3%, M:6.9%, n:788
RC10A2	16.5	47,547	174	3,893,066	C:91.0% [S:96.5%, D:3.5%], F:0.4%, M:5.1%, n:801
RC10B2	60.0	62,332	132	4,070,176	C:90.2% [S:95.5%, D:4.5%], F:0.2%, M:5.1%, n:805
R70.1	18.5	63,677	116	2,301,405	C:91.5% [S:99.5%, D:0.5%], F:0.3%, M:7.7%, n:789
OH30	26.6	278,810	16	2,355,518	C:92.9% [S:92.8%, D:0.1%], F:0.1%, M:7.0%, n:788
F70.1	33.0	65,604	153	2,340,200	C:91.3% [S:99.2%, D:0.8%], F:0.5%, M:7.5%, n:790
O70.2	11	228,043	13	2,216,138	C:92.2% [S:91.9%, D:0.3%], F:0.4%, M:7.4%, n:788
W70.1	15.5	202,408	36	2,295,909	C:91.9% [S:99.5%, D:0.5%], F:0.3%, M:7.4%, n:789
O70.1	10.7	156,181	33	2,245,644	C:91.7% [S:91.6%, D:0.1%], F:0.5%, M:7.8%, n:788
WC101	78.6	12,595	308	2,529,723	C:86.9% [S:86.8%, D:0.1%], F:2.8%, M:10.3%, n:788
WC10meta	30.8	91,742	40	2,260,421	C:83.6% [S:83.1%, D:0.5%], F:0.6%, M:15.8%, n:788
Median Statistics	24.6	43,802	128	2,798,702	

[a] BUSCO notation quickly summarizes counts of benchmarking universal single-copy orthologs, here in reference to the Synechococcales order. C = complete, S = single-copy, D = duplicated, F = fragmented/pseudogenized, M = missing, n = total counts (Manni et al. 2021)

Table 3. NCBI PGAP annotation summaries.

Strain name	№ Genes	№ CDSs	№ rRNAs [ab]			№ tRNAs	№ ncRNAs	№ Pseudogenes	№ CRISPRs
			5S	16S	23S				
Nb3U1	3,300	3,209	1	1	1	42	4	42	0
R50.1	2,570	2,500	0	0	0	39	4	27	7
R3-13	2,878	2,825	0	0	(1)	27	3	22	0
R55.1	2,875	2,793	3 + (1)	3 + (1)	1 + (4)	40	2	27	22
R55.2	3,198	3,082	4	1 + (5)	4 + (2)	51	5	44	22
R55.3	2,858	2,781	1	(1)	0	43	4	28	10
R55.4	2,760	2,687	1	0	0	40	4	28	10
R55.5	2,717	2,634	0	2	(3)	45	4	29	9
R55.6	2,733	2,659	0	(1)	0	40	4	29	9
R55.7	2,700	2,636	1	1	1	37	3	21	3
R55.8	2,460	2,397	1	1	1	36	3	21	3
OH2	2,770	2,671	1	(1)	(3)	42	3	49	6
OH20	2,640	2,561	2	2	2	45	3	25	8
R5-12	2,863	2,778	2	1 + (2)	1 + (2)	43	3	31	8
R5-13	2,750	2,676	(1)	1	1	41	4	26	11
R5-15	2,737	2,656	1	1	1	41	3	34	6
R5-16	2,754	2,670	1	1	1	41	3	37	11
H55.1	2,563	2,502	0	0	0	40	3	18	8
H55.2	2,632	2,562	1	1	1	42	3	22	5
H55.4	2,641	2,573	0	(2)	(1)	43	3	19	9
H55.5	2,539	2,468	1	1	1	42	3	23	8
H55.6	2,599	2,530	1	1	1	42	3	21	7
H55.7	2,539	2,478	0	0	0	39	3	19	8
H55.8	2,543	2,473	1	1	(2)	40	3	23	5
H55.9	2,640	2,577	(1)	1	1	36	3	21	1

[a] (#) = Partial annotation

[b] [#] = Total number of genes reported when complete/partial distinction not reported

Table 3 (continued). NCBI PGAP annotation summaries.

Strain name	№ Genes	№ CDSs	№ rRNAs [ab]			№ tRNAs	№ ncRNAs	№ Pseudogenes	№ CRISPRs
			5S	16S	23S				
H55.10	2,968	2,867	3 + (1)	4 + (4)	4 + (2)	47	3	33	33
H55.11	2,793	2,707	2 + (1)	2 + (4)	2 + (5)	45	3	22	40
JA-2-3B'a(2-13)	2,812	2,722	2	2	2	44	3	37	6
W55.1	2,771	2,675	1 + (1)	1 + (3)	1 + (2)	42	4	41	13
W55.2	2,745	2,677	0	0	0	41	3	24	6
R60.1	2,719	2,645	0	2	1 + (1)	44	3	24	11
R60.2	2,602	2,530	0	1	(2)	36	3	30	5
R60.3	2,905	2,823	3	3	3	43	3	27	37
60AY4M2	2,709	2,622	[	3	]	47	1	36	0
R6-5	2,797	2,717	0	(2)	(1)	41	3	33	0
R6-6	2,852	2,780	1	(1)	(5)	37	4	24	0
R6-7	2,790	2,720	0	(1)	(3)	37	4	25	0
R6-10	3,164	3,079	1	1	1 + (1)	47	3	31	8
W60.1	2,270	2,189	1	1	1	32	2	44	6
JA-3-3Ab	2,825	2,701	2	2	2	47	4	67	7
H60.1	2,662	2,552	3	4 + (2)	4 + (1)	45	3	48	20
H60.2	2,549	2,451	1	1	1	42	3	50	7
H60.3	2,573	2,499	0	0	0	38	3	33	6
H60.4	2,616	2,516	1	1	1	42	3	52	7
R60.4	2,624	2,532	1	(1)	(1)	39	3	47	8
W60.2	2,592	2504	1	(1)	(1)	38	3	44	8
B60.1	2,638	2,552	0	0	(1)	40	3	42	8
B60.2	2,521	2,432	0	0	0	40	3	46	4
R8-2	2,906	2,841	0	(3)	(2)	36	3	21	0
R65.1	2,207	2,136	1	1	0	39	4	26	2

[a] (#) = Partial annotation

[b] [#] = Total number of genes reported when complete/partial distinction not reported

Table 3 (continued). NCBI PGAP annotation summaries.

Strain name	№ Genes	№ CDSs	№ rRNAs [ab]			№ tRNAs	№ ncRNAs	№ Pseudogenes	№ CRISPRs
			5S	16S	23S				
W60.3	2,555	2,457	1	1	1	39	2	54	6
65AY6A.5F	2,725	2,670	2	2	1 + (2)	45	3	58	8
65AY6A5	2,684	2,597	[	9	]	47	1	30	0
65AY6Li	2,741	2,632	[	11	]	51	1	46	0
H65.1	2,286	2,204	0	1	(1)	43	4	33	4
W65.1	2,179	2,104	1	1	1	40	4	28	2
H70.1	2,422	2,321	3	3 + (2)	2 + (5)	46	4	36	4
H70.2	2,263	2,189	0	0	0	41	4	29	4
RC10A2	3,708	3,557	4	3 + (1)	3 + (2)	75	4	59	12
RC10B2	3,751	3,605	4	3 + (1)	3 + (3)	66	5	61	10
R70.1	2,304	2,204	0	2 + (1)	3 + (7)	48	4	35	7
OH30	2,268	2,191	0	0	(1)	40	4	32	4
F70.1	2,359	2,252	4	3 + (6)	1 + (11)	49	4	29	2
O70.2	2,178	2,108	0	0	0	40	4	26	2
W70.1	2,261	2,172	3	3 + (2)	2 + (2)	48	4	25	1
O70.1	2,211	2,125	0	1	(1)	44	4	36	2
WC101	2,532	2,447	1	0	(1)	37	3	43	10
WC10meta	2,064	1,996	1 + (1)	2	1	36	4	23	2

[a] (#) = Partial annotation

[b] [#] = Total number of genes reported when complete/partial distinction not reported

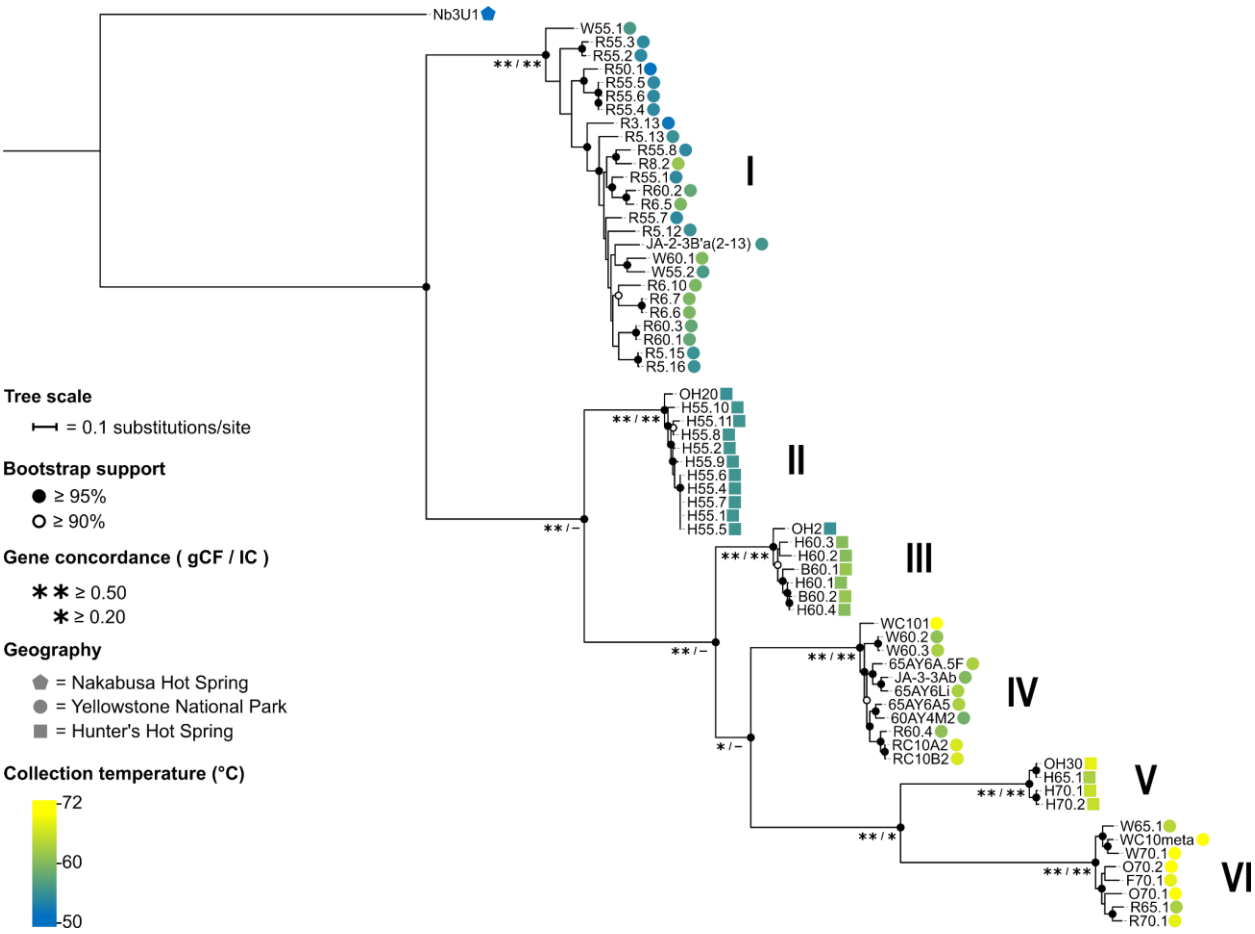

Figure 1. Maximum likelihood phylogeny of the SynAB group. A maximum likelihood phylogeny of a collection of 68 *Synechococcus* genomes reconstructed using a concatenated alignment of 118,301 amino acid sites derived from 404 single-copy orthologous gene sequences. The strains included in these analyses were isolated from Yellowstone NP, WY and Hunter's Hot Spring, OR, and were rooted with a strain of *Synechococcus* sp. T1 (Nb3U1). Strains from geographically isolated populations form six distinct clades that are generally well-supported by bootstrap analysis and sort with respect to collection temperature rather than by geography; strains from cooler environments are more basal in the phylogeny, with clades from increasingly hotter environments nested within them. gCF = gene concordance factor; IC = internode certainty.

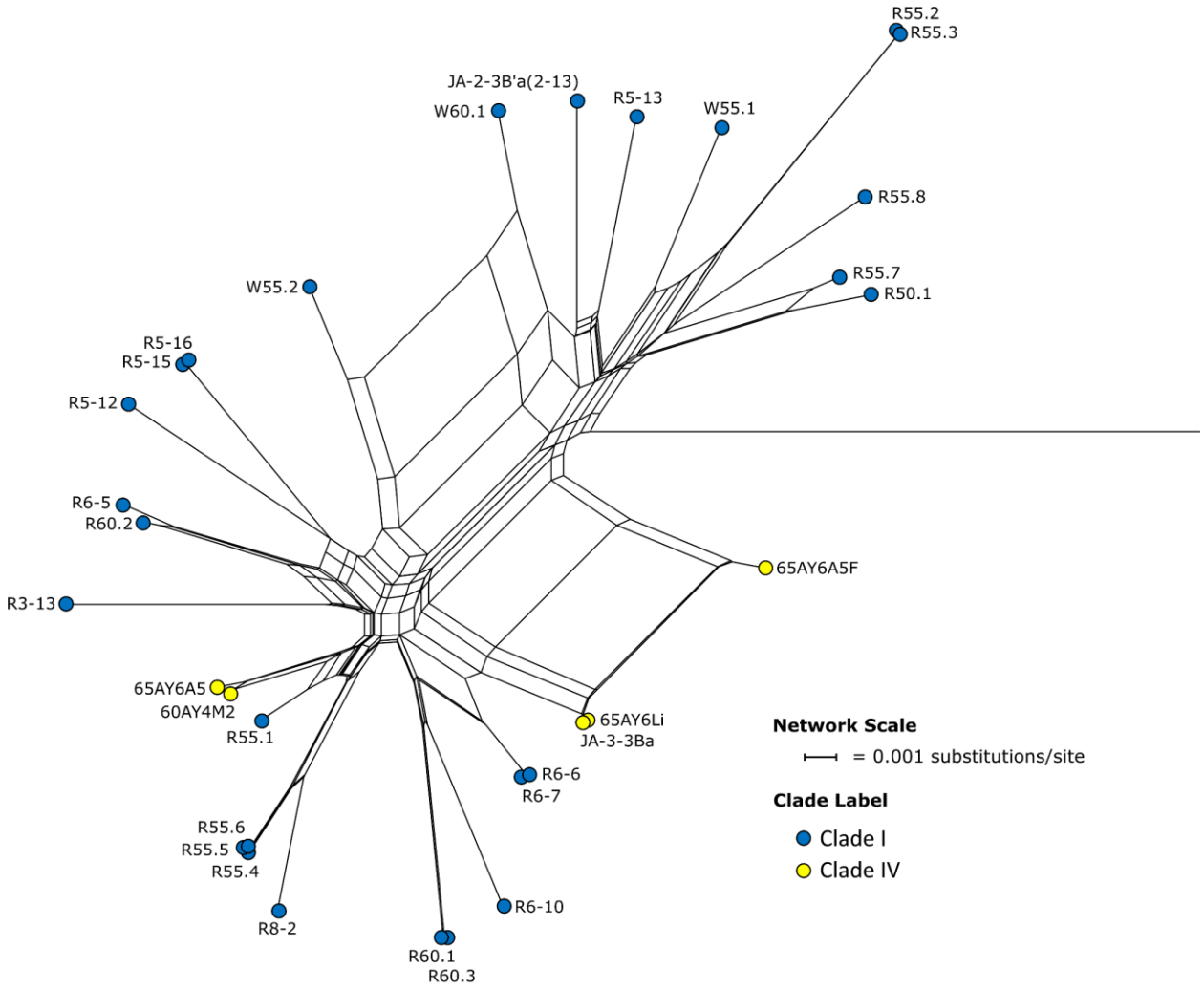

Figure 2. Neighbor network analysis of *nifHDK* from clades I and IV. A neighbor network analysis was conducted with a concatenated alignment of 3,888 nucleotides from clades I and IV using SplitsTree v4.14.4 (Huson and Bryant, 2005). There appears to have been at least two instances of HGT between members of clade I and clade IV.

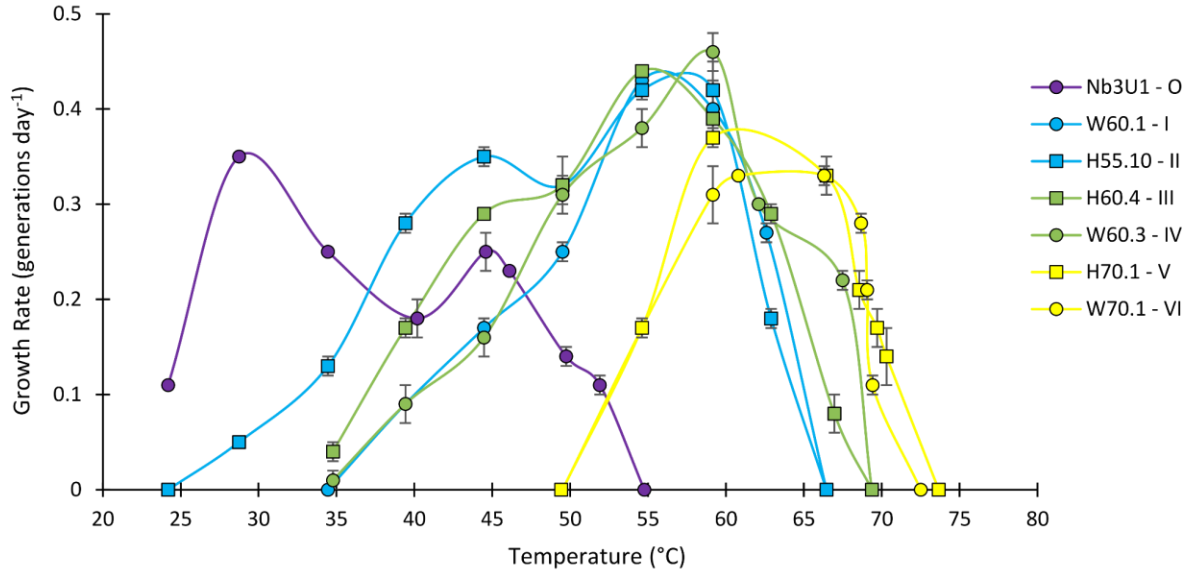

Figure 3. Thermal performance curves of representative SynAB strains. TPCs were characterized for representative *Synechococcus* A/B cells by assaying growth rate over a range of temperatures between 25 and 75 °C. The generation time during exponential growth was estimated after three generations of growth by determining log10 2/b, where b is the slope of logarithmically transformed A750 data regressed on time (in hours). This value was transformed and reported as number of generations per day, and growth rates were averaged across triplicates. CT_{min} increases from ~24 °C in clades I/II up to ~50 °C for clades V/VI; though not as dramatic, CT_{max} increases from below 67 °C for clades I/II to greater than 70 °C for clades V/VI. Changes in T_{opt} were also observed, where clades I-III exhibited maximal performance near ~55 °C which increased to ~60 °C in clade IV and lies between 60 and 65 °C for clades V/VI. Curves are colored based on similar values of CT_{max}.

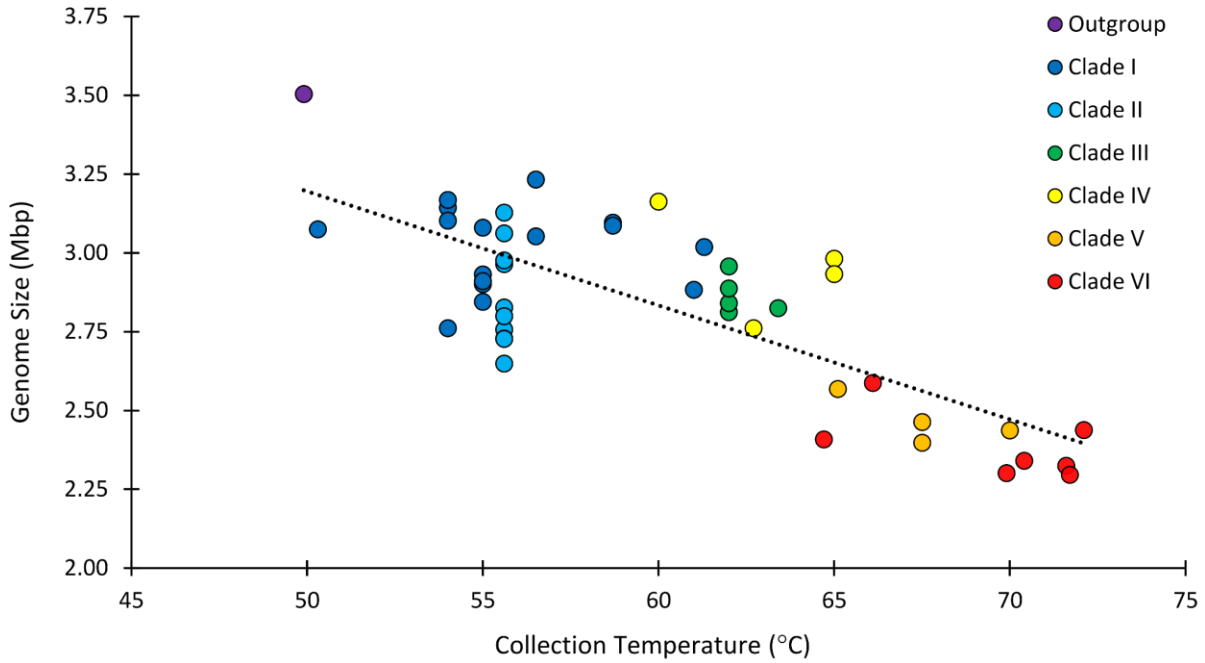

Figure 4. Genome size and collection temperature. Genomic statistics on *de novo* SynAB genome assemblies were measured using QUAST v4.5 (Gurevich et al., 2013) following a refinement pipeline. Genome size was plotted against environmental collection temperature. There is a strong negative correlation between genome size and environmental temperature of sample collection ($R = -0.78$; $F_{1,45} = 68.2$, $P < 0.0001$ for a PGLS model). Mean genome size for highly thermotolerant clade V/VI strains was 80% of that of clade I (mean ± SE: 2.4 ± 0.03 versus 3.0 ± 0.03 Mbp).

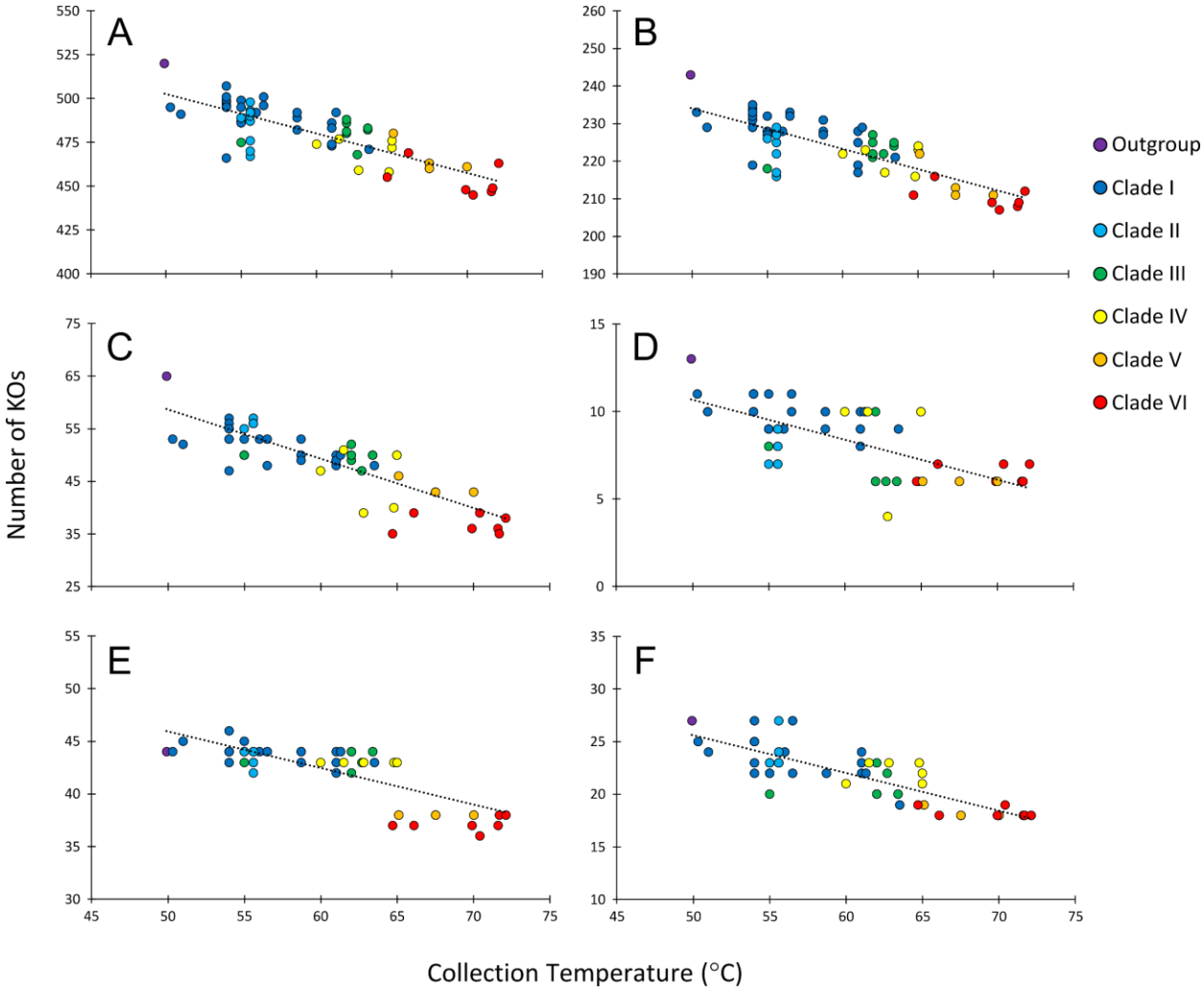

Figure 5. Example metabolic trends with collection temperature. Functional annotation was performed using eggNOG-mapper v2.0 and the KEGG pathway identifiers for each locus in these genomes were uploaded to the online KEGG Mapper Reconstruct module. Pathway representation was compared by plotting KEGG ortholog count against collection temperature for candidate metabolic modules that appear to underly the SynAB radiation to higher temperatures and potentially explain the reduction in SynAB genome size. **A.** global metabolic pathways; **B.** biosynthesis of secondary metabolites; **C.** ABC transporters; **D.** nitrogen metabolism; **E.** porphyrin metabolism; **F.** quorum sensing. KOs = KEGG orthologs.

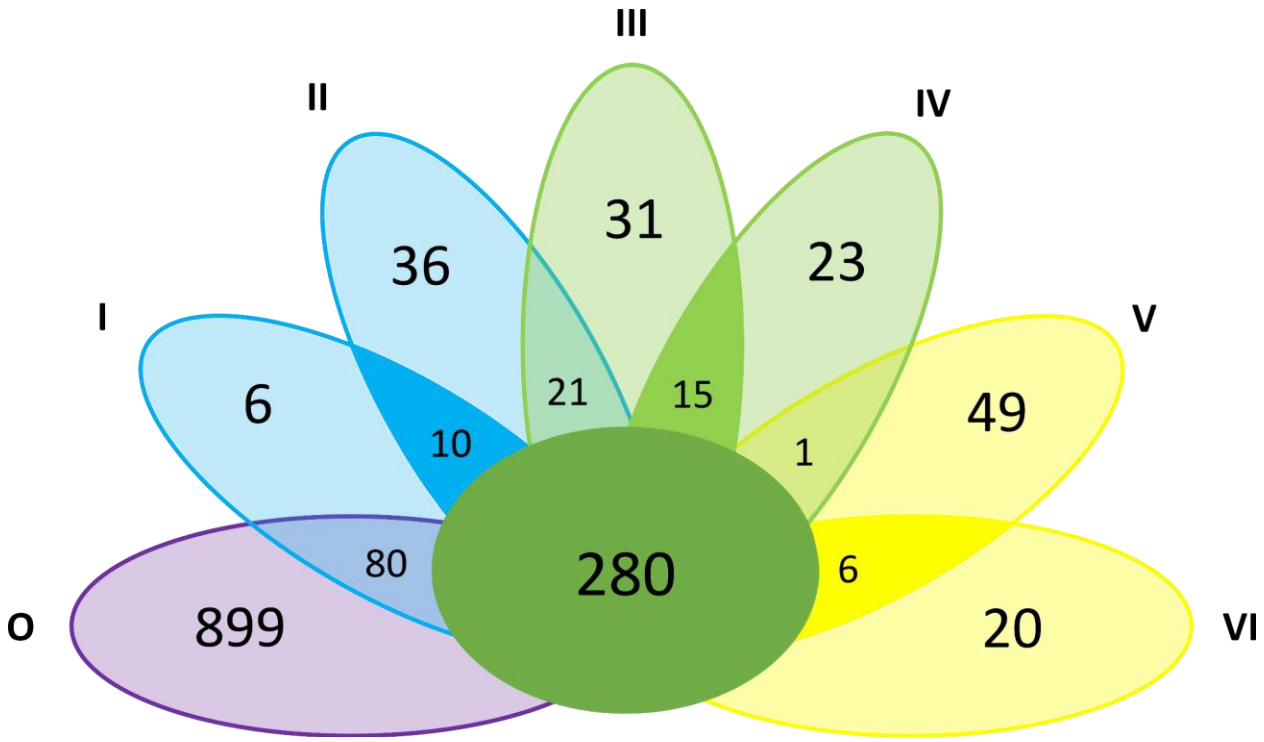

Figure 6. Core genome intersections of SynAB clades. The core genomes of each clade were identified using ROARY, then compared to identify genomic features only belonging to a single clade and the relevant clade intersections shown. Results were curated via a local BLASTx search against the individual amino acid annotation files of all SynAB genomes within the species phylogeny to identify biologically interesting differences among the clades. Overall, 280 loci are shared between the entire SynAB phylogeny and the outgroup Nb3U1. Fewer than 50 loci are unique to individual clades of SynAB, and less than 20 are unique to each of the three thermotolerance groups (I/II, III/IV, V/VI). Roman numerals indicate clade as in Figure 1 and clades were colored based on CT_{max} values as in Figure 2; O = outgroup.

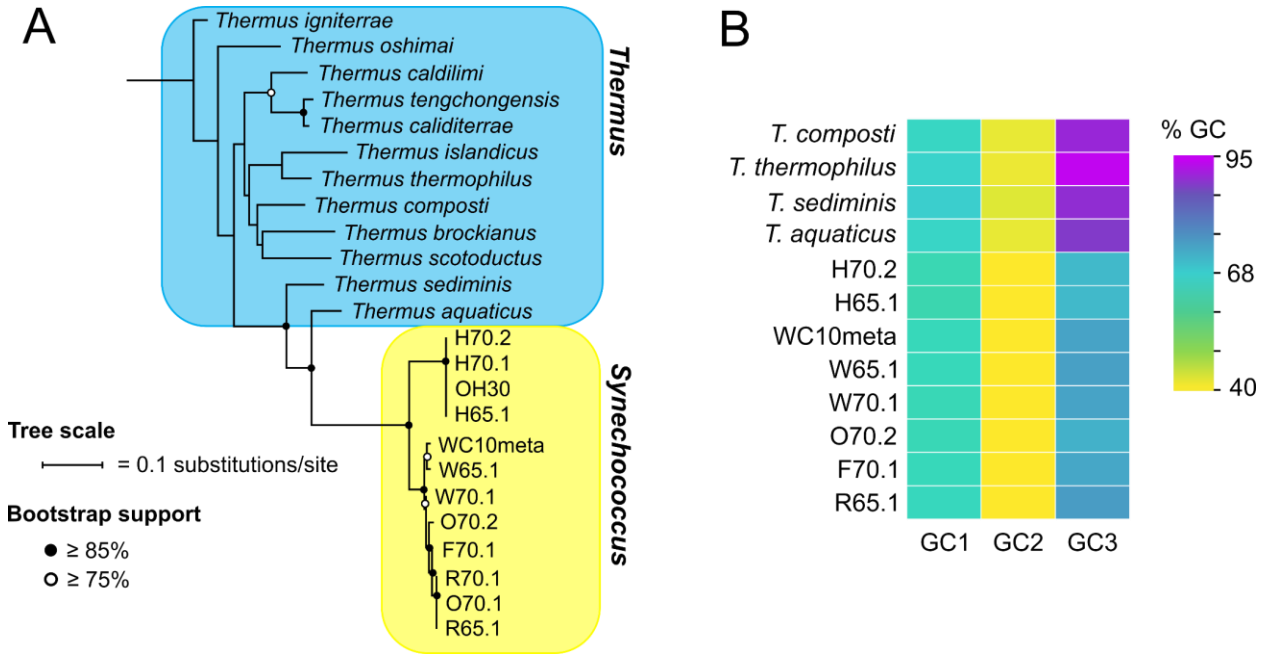

Figure 7. HGT of tRNA methyltransferase from *Thermus* to *Synechococcus*. There are very few examples of HGT during *Synechococcus* adaptation above 70 °C but obtaining a tRNA modification enzyme from *Thermus* appears to be important. **A.** Maximum likelihood *trmH* phylogeny of selected *Thermus* strains and *Synechococcus* A/B clades V and VI. Maximum likelihood trees were reconstructed for a ClustalW alignment of *trmH* genes with a TPM3+F+G4 model and 1000 bootstrap replicates with IQ-TREE (Nguyen et al., 2015). **B.** Percentage of GC base pairs at each codon site within the *trmH* reading frame. There has been little change in nucleotide usage at first and second codon positions. By contrast, the *Synechococcus* sequences have diverged greatly from *Thermus* at third codon sites: whereas GC3 is >90% in *Thermus* strains, it varies between ~70-75% in *Synechococcus*.

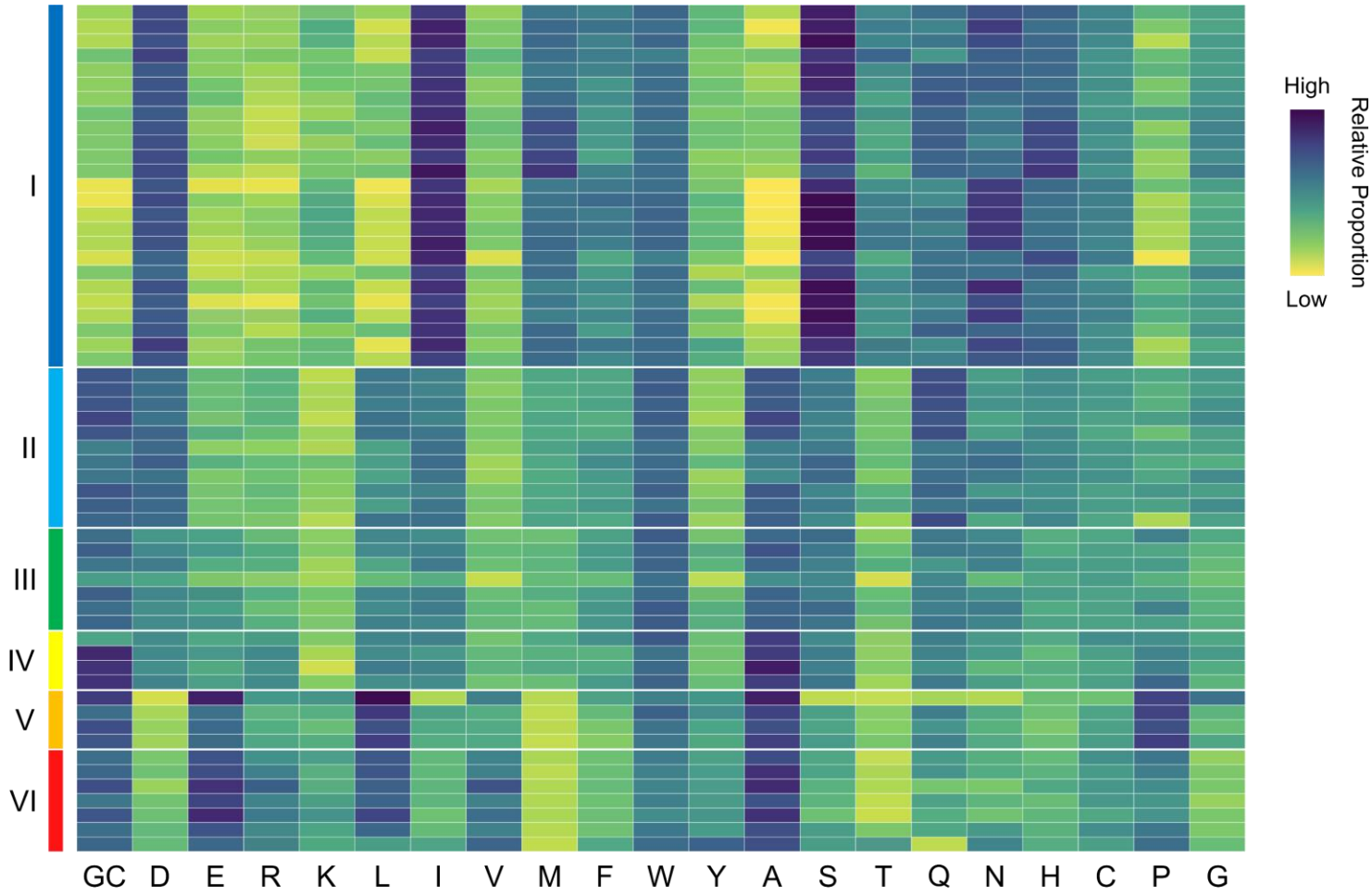

Figure 8. Relative proportions of GC and amino acid content in SynAB. There is not a general increase in charged amino acids (DERK) with increased thermotolerance. but clade VI Synechococcus exhibited a subtle increase in the frequency of these residues (20.9%) compared with other clades (20.6-20.7%). Similarly, clades have not diverged in the frequency of the bulky aliphatic amino acids (ILVM). However, there is an overall decrease in polar uncharged residues (NQST): 18.6% in Clade I, 18.0% in Clades V/VI. The absolute ranges for each category are as follows: GC (58-61%); D (4-4.5); E (6.2-6.7); R (7-7.5); K (2.7-3.2); L (12.3-13); I (4.4-4.9); V (6.8-7.3); M (1.6-1.9); F (3.3-3.6); W (1.6-1.9); Y (2.5-2.8); A (9.7-10.3); S (5.3-5.8); T (4.4-4.9); Q (5.4-5.9); N (2.3-2.6); H (1.8-2.1); C (0.9-1.2); P (6.2-6.5); G (7.6-8.1).

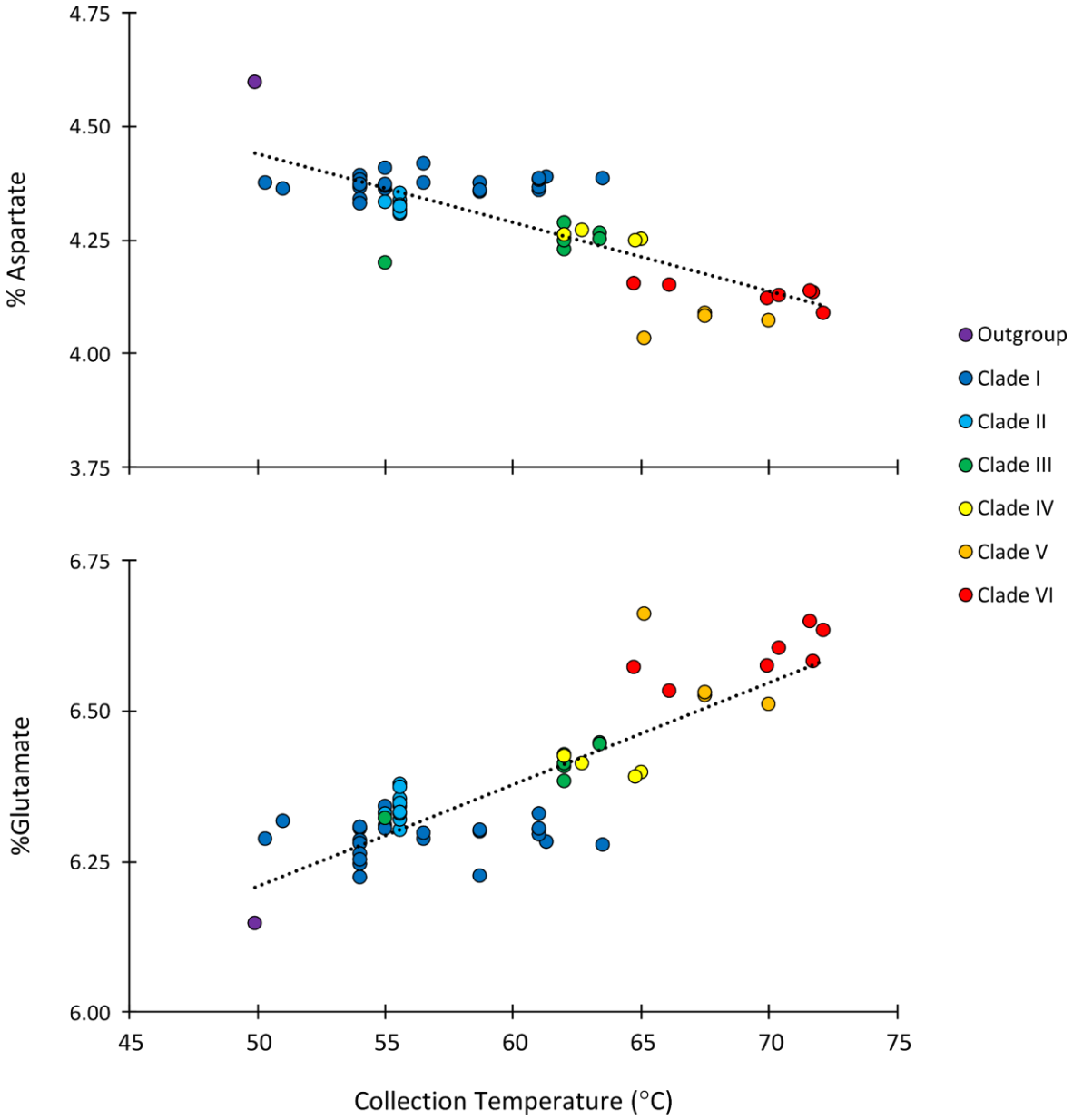

Figure 9. Changes in aspartate and glutamate content during SynAB divergence. There was a marked reduction in the use of aspartate with increasing environmental temperature, especially for the more thermotolerant strains of clades III-VI isolated from samples collected at or above 62 °C (R^2 = 0.73; $F_{1,31}$ = 82.1, P < 0.0001; slope = -0.014 % Asp per °C). Loss of aspartate in *Synechococcus* proteins is mirrored by similar gains in glutamate (R^2 = 0.80; $F_{1,31}$ = 125.7, P < 0.0001), which increases at the same rate as aspartate declines (0.017 % Glu per °C).

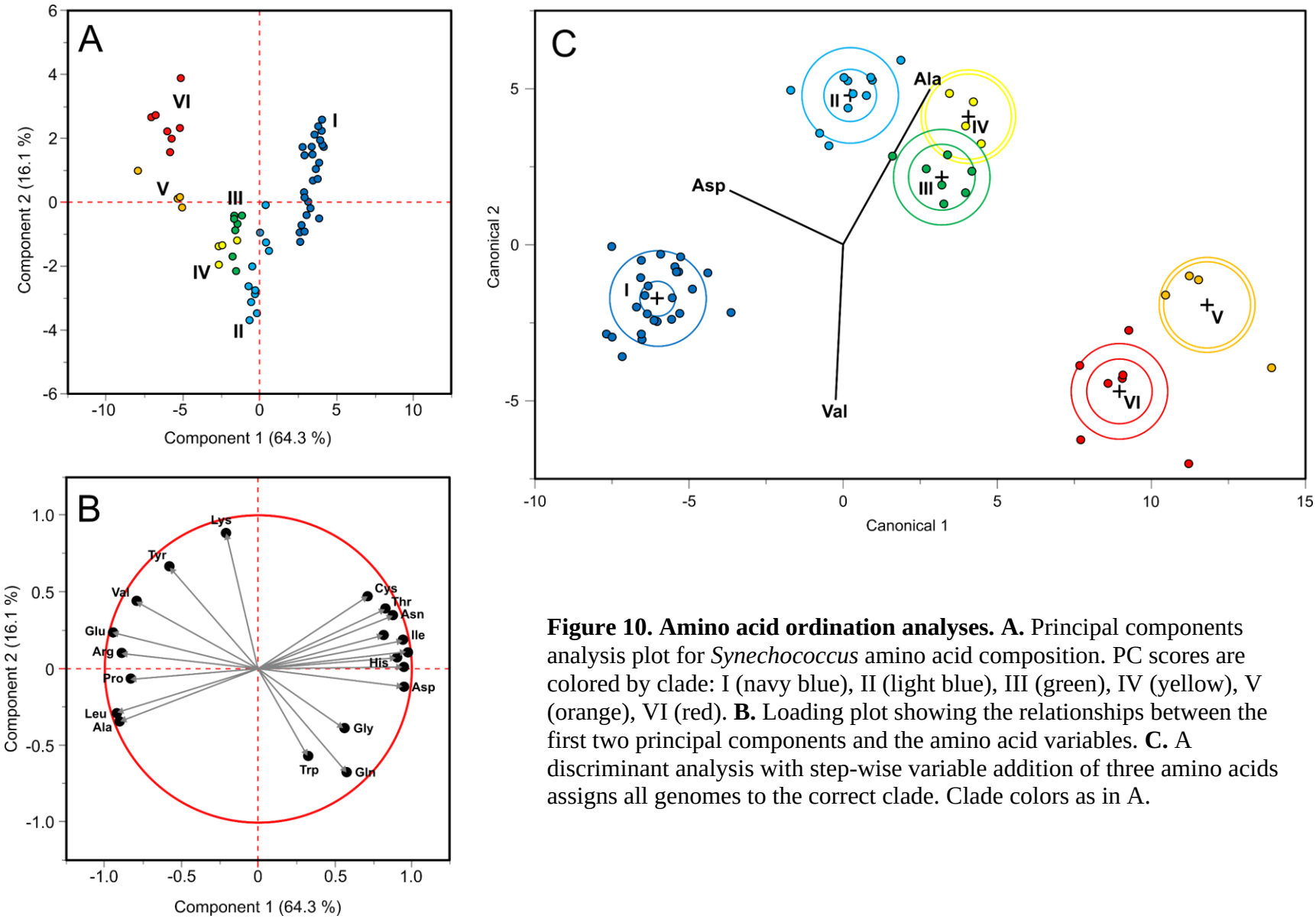

Figure 10. Amino acid ordination analyses. A. Principal components analysis plot for *Synechococcus* amino acid composition. PC scores are colored by clade: I (navy blue), II (light blue), III (green), IV (yellow), V (orange), VI (red). **B.** Loading plot showing the relationships between the first two principal components and the amino acid variables. **C.** A discriminant analysis with step-wise variable addition of three amino acids assigns all genomes to the correct clade. Clade colors as in A.

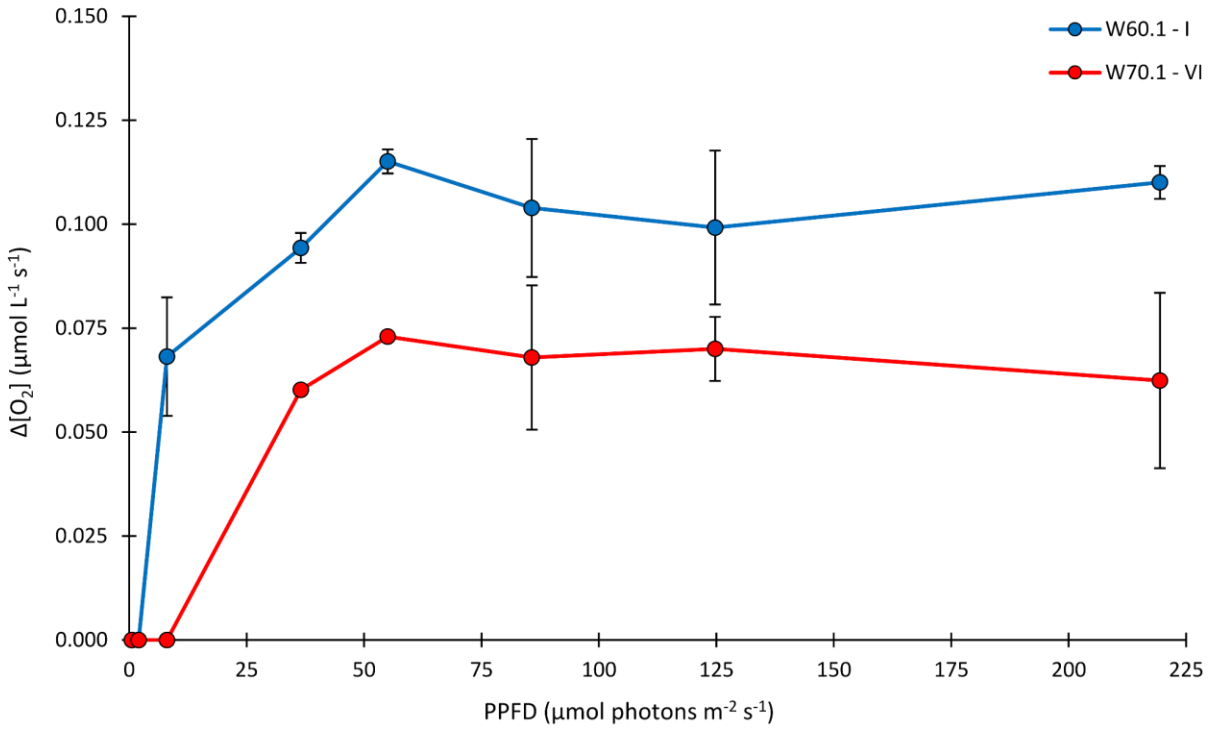

Figure 11. PVI curves for W60.1 and W70.1 at 55 °C. Growth simulations using metabolic models were parameterized in part by rates of oxygen evolution for each strain at 55 °C following Broddrick et al. (2019). The lower temperature strain, W60.1 (clade I) has a higher maximum rate of oxygen evolution compared to the high temperature W70.1 (clade VI).

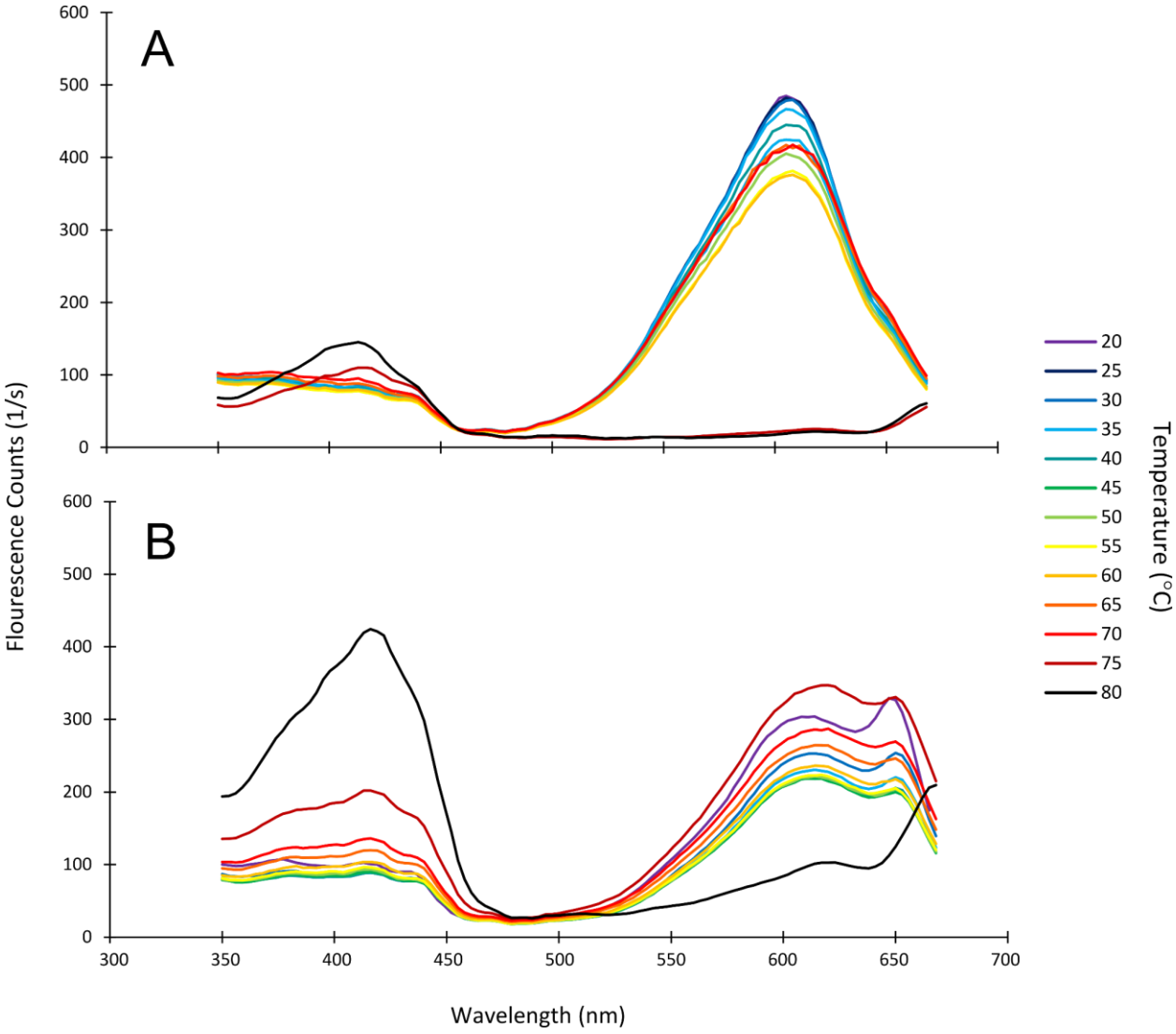

Figure 12. Excitation spectra of W60.1 and W70.1. Excitation spectra of both strains were collected to probe electron transfer efficiency through PSII across a range of temperatures using a Photon Technology International model QM-7/2005 rapid temperature change spectrofluorometer equipped with a Quantum North-west TLC 50 thermoelectric temperature-controlled cuvette holder. The decrease in fitness above 60 °C for W70.1 (Figure 3) doesn't appear to be an issue with excitation energy transport efficiency, as 70 °C showed a near-optimal action spectrum. **A.** Spectra for W60.1; inactivation of PSII occurs at 75 °C. **B.** Spectra for W70.1; inactivation of PSII occurs at 80 °C.

Literature Cited

Ahern, T.J. and Klibanov, A.M. 1985. The mechanism of irreversible enzyme inactivation at 100°C. *Science*, 228(4705): 1280–1284. https://doi.org/10.1126/science.4001942

Allen, M.B. 1953. The thermophilic aerobic sporeforming bacteria. *Bacteriological Reviews*, 17(2): 125–173. https://doi.org/10.1128/mmbr.17.2.125-173.1953

Allewalt, J.P., Bateson, M.M., Revsbech, N.P., Slack, K. and Ward, D.M. 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. *Applied and Environmental Microbiology*, 72(1): 544–550. https://doi.org/10.1128/AEM.72.1.544-550.2006

Anderson, K.L., Tayne, T.A. and Ward, D.M. 1987. Formation and fate of fermentation products in hot spring cyanobacterial mats. *Applied and Environmental Microbiology*, 53(10): 2343–2352. https://doi.org/10.1128/aem.53.10.2343-2352.1987

Angilletta, M.J., Huey, R.B. and Frazier, M.R. 2010. Thermodynamic effects on organismal performance: Is hotter better? *Physiological and Biochemical Zoology*, 83(2): 197–206. https://doi.org/10.1086/648567

Angilletta, M. Jr. 2009. *Thermal adaptation: A theoretical and empirical synthesis*. First edition. New York, NY, Oxford University Press Inc.

Aravind, L., Tatusov, R.L., Wolf, Y.I., Walker, D.R. and Koonin, E. v. 1998. Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genetics*, 14(11): 442–444. https://doi.org/10.1016/S0168-9525(98)01553-4

Armstrong, K.M. and Baldwin, R.L. 1993. Charged histidine affects alpha-helix stability at all positions in the helix by interacting with the backbone charges. *Proceedings of the National Academy of Sciences*, 90(23): 11337–11340. https://doi.org/10.1073/pnas.90.23.11337

Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. and Zagnitko, O. 2008. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, 9. https://doi.org/10.1186/1471-2164-9-75

Babraham Bioinformatics. 2019. *FastQC*. 0.11.9. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Basak, S., Banerjee, T., Gupta, S.K. and Ghosh, T.C. 2004. Investigation on the causes of codon and amino acid usages variation between thermophilic *Aquifex aeolicus* and

mesophilic *Bacillus subtilis*. *Journal of Biomolecular Structure and Dynamics*, 22(2): 205–214. https://doi.org/10.1080/07391102.2004.10506996

Beck, A.E., Bernstein, H.C. and Carlson, R.P. 2017. Stoichiometric network analysis of cyanobacterial acclimation to photosynthesis-associated stresses identifies heterotrophic niches. *Processes*, 5(2). https://doi.org/10.3390/pr5020032

Becraft, E.D., Wood, J.M., Cohan, F.M. and Ward, D.M. 2020. Biogeography of American northwest hot spring A/B′-lineage *Synechococcus* populations. *Frontiers in Microbiology*, 11(2). https://doi.org/10.3389/fmicb.2020.00077

Beeby, M., O'Connor, B.D., Ryttersgaard, C., Boutz, D.R., Perry, L.J. and Yeates, T.O. 2005. The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biology*, 3(9): 1549–1558. https://doi.org/10.1371/journal.pbio.0030309

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M.C., Bateson, M.M., Ward, D.M. and Heidelberg, J.F. 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME Journal*, 1(8): 703–713. https://doi.org/10.1038/ismej.2007.46

Blumer-Schuette, S.E., Ozdemir, I., Mistry, D., Lucas, S., Lapidus, A., Cheng, J.F., Goodwin, L.A., Pitluck, S., Land, M.L., Hauser, L.J., Woyke, T., Mikhailova, N., Pati, A., Kyrpides, N.C., Ivanova, N., Detter, J.C., Walston-Davenport, K., Han, S., Adams, M.W.W. and Kelly, R.M. 2011. Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. *Journal of Bacteriology*, 193(6): 1483–1484. https://doi.org/10.1128/JB.01515-10

Bolger, A.M., Lohse, M. and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bonheyo, G., Frias-Lopez, J. and Fouke, B. 2005. A test for airborne dispersal of thermophilic bacteria from hot springs. In: *Geothermal Biology and Geochemistry in Yellowstone National Park*. First edition. Bozeman, MT. Montana State University Publications.

Boutz, D.R., Cascio, D., Whitelegge, J., Perry, L.J. and Yeates, T.O. 2007. Discovery of a Thermophilic Protein Complex Stabilized by Topologically Interlinked Chains. *Journal of Molecular Biology*, 368(5): 1332–1344. https://doi.org/10.1016/j.jmb.2007.02.078

Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R. and Xia, F. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5. https://doi.org/10.1038/srep08365

Brochier-Armanet, C. and Forterre, P. 2006. Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea*, 2(2): 83–93. https://doi.org/10.1155/2006/582916

Brock, M.L., Wiegert, R.G. and Brock, T.D. 1969. Feeding by Paracoenia and Ephydra (Diptera>Ephydridae) on the microorganisms of hot springs. *Ecology*, 50(20). https://doi.org/10.2307/1934846

Broddrick, J.T., Rubin, B.E., Welkie, D.G., Du, N., Mih, N., Diamond, S., Lee, J.J., Golden, S.S. and Palsson, B.O. 2016. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. *Proceedings of the National Academy of Sciences*, 113(51): E8344–E8353. https://doi.org/10.1073/pnas.1613446113

Broddrick, J.T., Welkie, D.G., Jallet, D., Golden, S.S., Peers, G. and Palsson, B.O. 2019. Predicting the metabolic capabilities of *Synechococcus* elongatus PCC 7942 adapted to different light regimes. *Metabolic Engineering*, 52(November 2018): 42–56. https://doi.org/10.1016/j.ymben.2018.11.001

Cacciapuoti, G., Moretti, M.A., Forte, S., Brio, A., Camardella, L., Zappia, V. and Porcelli, M. 2004. Methylthioadenosine phosphorylase from the archaeon *Pyrococcus furiosus*: Mechanism of the reaction and assignment of disulfide bonds. *European Journal of Biochemistry*, 271(23–24): 4834–4844. https://doi.org/10.1111/j.1432-1033.2004.04449.x

Cacia, J., Keck, R., Presta, L.G. and Frenz, J. 1996. isomerization of an aspartic acid residue in the complementarity-determining regions of a recombinant antibody to human IgE: Identification and effect on binding affinity. *Biochemistry*, 35(6): 1897–1903. https://doi.org/10.1021/bi951526c

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics*, 10. https://doi.org/10.1186/1471-2105-10-421

Camp, V.E. and Wells, R.E. 2021. The case for a long-lived and robust Yellowstone hotspot. *GSA Today*, 31(1): 4–10. https://doi.org/10.1130/GSATG477A.1

Cantalapiedra, C.P., Herṇandez-Plaza, A., Letunic, I., Bork, P. and Huerta-Cepas, J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12): 5825–5829. https://doi.org/10.1093/molbev/msab293

Capasso, S., Mazzarella, L., Sica, F., Zagari, A. and Salvadori, S. 1993. Kinetics and mechanism of succinimide ring formation in the deamidation process of asparagine residues. *Journal of the Chemical Society, Perkin Transactions 2*(4): 679. https://doi.org/10.1039/p29930000679

Carlos Guimaraes, L., Benevides de Jesus, L., Vinicius Canario Viana, M., Silva, A., Thiago Juca Ramos, R., de Castro Soares, S. and Azevedo, V. 2015. Inside the pan-genome: Methods and software overview. *Current Genomics*, 16(4): 245–252. https://doi.org/10.2174/1389202916666150423002311

Castenholz, R.W. 1988. Culturing methods for cyanobacteria. In: *Methods in Enzymology*. pp. 68–93. Vol. 167. San Diego, CA, Academic Press, Inc.

Chen, K., Gao, Y., Mih, N., O'Brien, E.J., Yang, L. and Palsson, B.O. 2017. Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proceedings of the National Academy of Sciences*, 114(43): 11548–11553. https://doi.org/10.1073/pnas.1705524114

Christiansen, R.L., Foulger, G.R. and Evans, J.R. 2002. Upper-mantle origin of the Yellowstone hotspot. *Bulletin of the Geological Society of America*, 114(10): 1245–1256. https://doi.org/10.1130/0016-7606(2002)114<1245:UMOOTY>2.0.CO;2

Clarke, A. and Fraser, K.P.P. 2004. Why does metabolism scale with temperature? *Functional Ecology*, 18(2): 243–251. https://doi.org/10.1111/j.0269-8463.2004.00841.x

Cohan, F.M. and Perry, E.B. 2007. a systematics for discovering the fundamental units of bacterial diversity. *Current Biology*, 17(10): R373-R386. https://doi.org/10.1016/j.cub.2007.03.032

Déclais, A.C., Marsault, J., Confalonieri, F., Bouthier De La Tour, C. and Duguet, M. 2000. Reverse gyrase, the two domains intimately cooperate to promote positive supercoiling. *Journal of Biological Chemistry*, 275(26): 19498–19504. https://doi.org/10.1074/jbc.M910091199

Dillon, J.G., Fishbain, S., Miller, S.R., Bebout, B.M., Habicht, K.S., Webb, S.M. and Stahl, D.A. 2007. High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Applied and Environmental Microbiology*, 73(16): 5218–5226. https://doi.org/10.1128/AEM.00357-07

Dutta, A. and Chaudhuri, K. 2010. Analysis of tRNA composition and folding in psychrophilic, mesophilic and thermophilic genomes: Indications for thermal adaptation. *FEMS Microbiology Letters*, 305(2): 100–108. https://doi.org/10.1111/j.1574-6968.2010.01922.x

Dvořák, P., Casamatta, D.A., Poulíčková, A., Hašler, P., Ondřej, V. and Sanges, R. 2014. Synechococcus: 3 billion years of global dominance. *Molecular Ecology*, 23(22): 5538–5551. https://doi.org/10.1111/mec.12948

Ebrahim, A., Lerman, J.A., Palsson, B.O. and Hyduke, D.R. 2013. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7. https://doi.org/10.1186/1752-0509-7-74

Edgar, R.C. 2022. High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. https://doi.org/10.1101/2021.06.20.449169

Emms, D.M. and Kelly, S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1). https://doi.org/10.1186/s13059-015-0721-2

Emms, D.M. and Kelly, S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1). https://doi.org/10.1186/s13059-019-1832-y

Falk, S., Maxwell, D.P., Laudenbach, D.E. and Huner, N.P.A. 1996. Photosynthetic adjustment to temperature. In: *Photosynthesis and the Environment*. pp. 367–385. Dordrecht, Netherlands, Springer. https://doi.org/10.1007/0-306-48135-9_15

Ferris, M.J. and Ward, D.M. 1997. Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology*, 63(4): 1375–1381. https://doi.org/10.1128/aem.63.4.1375-1381.1997

Forterre, P. 2002. A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends in Genetics*, 18(5): 236–237. https://doi.org/10.1016/S0168-9525(02)02650-1

Galtier, N. and Lobry, J.R. 1997. Revisiting the relationships between genomic G + C Content, RNA secondary structures, and optimal growth temperature. *Journal of Molecular Evolution*, 44: 632–636. https://doi.org/10.1007/pl00006186

Garrett, R.H. and Grisham, C.M. 2017. *Biochemistry*. Sixth edition. Boston, MA, Cengage Learning.

Geiger, T. and Clarke, S. 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *Journal of Biological Chemistry*, 262(2): 785–794. https://doi.org/10.1016/s0021-9258(19)75855-4

Gillooly, J.F., Allen, A.P., Savage, V.M., Charnov, E.L., West, G.B. and Brown, J.H. 2006. Response to Clarke and Fraser: Effects of temperature on metabolic rate. *Functional Ecology*, 20(2): 400–404. https://doi.org/10.1111/j.1365-2435.2006.01110.x

Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., Bibbs, L., Eads, J., Richardson, T.H., Noordewier, M., Rappé, M.S., Short, J.M., Carrington,

J.C. and Mathur, E.J. 2005. Genetics: Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, 309(5738): 1242–1245. https://doi.org/10.1126/science.1114057

Gribaldo, S. and Brochier-Armanet, C. 2006. The origin and evolution of Archaea: A state of the art. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1470): 1007–1022. https://doi.org/10.1098/rstb.2006.1841

Gudmundsson, S., Agudo, L. and Nogales, J. 2017. Applications of genome-scale metabolic models of microalgae and cyanobacteria in biotechnology. In: *Microalgae-based Biofuels and Bioproducts*. Sawston, UK. Woodhead Publishing. https://doi.org/10.1016/B978-0-08-101023-5.00004-2

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3): 307–321. https://doi.org/10.1093/sysbio/syq010

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. 2013. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, 29(8): 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haft, D.H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., Gwadz, M., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Zheng, C., Thibaud-Nissen, F., Geer, L.Y., Marchler-Bauer, A. and Pruitt, K.D. 2018. RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Research*, 46(D1): D851–D860. https://doi.org/10.1093/nar/gkx1068

Hass, J.W. 2000. The Reverend Dr William Henry Dallinger, F.R.S. (1839-1909). *Notes and Records of the Royal Society of London*, 54: 53–65. https://doi.org/10.1098/rsnr.2000.0096

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S.N., Richelle, A., Heinken, A., Haraldsdóttir, H.S., Wachowiak, J., Keating, S.M., Vlasov, V., Magnusdóttir, S., Ng, C.Y., Preciat, G., Žagare, A., Chan, S.H.J., Aurich, M.K., Clancy, C.M., Modamio, J., Sauls, J.T., Noronha, A., Bordbar, A., Cousins, B., el Assal, D.C., Valcarcel, L. v., Apaolaza, I., Ghaderi, S., Ahookhosh, M., ben Guebila, M., Kostromins, A., Sompairac, N., Le, H.M., Ma, D., Sun, Y., Wang, L., Yurkovich, J.T., Oliveira, M.A.P., Vuong, P.T., el Assal, L.P., Kuperstein, I., Zinovyev, A., Hinton, H.S., Bryant, W.A., Aragón Artacho, F.J., Planes, F.J., Stalidzans, E., Maass, A., Vempala, S., Hucka, M., Saunders, M.A., Maranas, C.D., Lewis, N.E., Sauter, T., Palsson, B., Thiele, I. and Fleming, R.M.T. 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3): 639–702. https://doi.org/10.1038/s41596-018-0098-2

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q. and Vinh, L.S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2): 518–522. https://doi.org/10.1093/molbev/msx281

Hollinger, M. and Steiner Verlag, F. 2016. Life from elsewhere–Early history of the maverick theory of panspermia. *Sudhoff's Archive*, 100(2): 188–205. https://www.jstor.org/stable/24913787

Holloway, J.A.M., Nordstrom, D.K., Böhlke, J.K., McCleskey, R.B. and Ball, J.W. 2011. Ammonium in thermal waters of Yellowstone National Park: Processes affecting speciation and isotope fractionation. *Geochimica et Cosmochimica Acta*, 75(16): 4611–4636. https://doi.org/10.1016/j.gca.2011.05.036

Hori, H., Suzuki, T., Sugawara, K., Inoue, Y., Shibata, T., Kuramitsu, S., Yokoyama, S., Oshima, T. and Watanabe, K. 2002. Identification and characterization of tRNA (Gm18) methyltransferase from *Thermus thermophilus* HB8: domain structure and conserved amino acid sequence motifs. *Genes to Cells*, 7(3): 259–272. https://doi.org/10.1046/j.1365-2443.2002.00520.x

Hurley, T.D. and Weiner, H. 1992. Crystallization and preliminary X-ray investigation of bovine liver mitochondrial aldehyde dehydrogenase. *Journal of Molecular Biology*, 227(4): 1255–1257. https://doi.org/10.1016/0022-2836(92)90536-S

Huson, D.H. and Bryant, D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2): 254–267. https://doi.org/10.1093/molbev/msj030

Inglis, A.S. 1983. Cleavage at aspartic acid. In: *Methods in Enzymology*. pp. 324–332. Vol. 91. https://doi.org/10.1016/S0076-6879(83)91030-3

Inskeep, W.P., Jay, Z.J., Tringe, S.G., Herrgård, M.J. and Rusch, D.B. 2013. The YNP metagenome project: Environmental parameters responsible for microbial distribution in the yellowstone geothermal ecosystem. *Frontiers in Microbiology*, 4(MAY): 1–15. https://doi.org/10.3389/fmicb.2013.00067

Jasser, I., Panou, M., Khomutovska, N., Sandzewicz, M., Panteris, E., Niyatbekov, T., Łach, Ł., Kwiatowski, J., Kokociński, M. and Gkelis, S. 2022. Cyanobacteria in hot pursuit: Characterization of cyanobacteria strains, including novel taxa, isolated from geothermal habitats from different ecoregions of the world. *Molecular Phylogenetics and Evolution*, 170: 107454. https://doi.org/10.1016/j.ympev.2022.107454

Javaux, E.J. 2006. Extreme life on Earth: Past, present and possibly beyond. *Research in Microbiology*, 157: 37–48. https://doi.org/10.1016/j.resmic.2005.07.008

Jiang, L., Lin, M., Li, X., Cui, H., Xu, X., Li, S. and Huang, H. 2013. Genome sequence of *Thermus thermophilus* ATCC 33923, a thermostable trehalose-producing strain. *Genome Announcements*, 1(4). https://doi.org/10.1128/genomeA.00493-13

Jorda, J. and Yeates, T.O. 2011. widespread disulfide bonding in proteins from thermophilic archaea. *Archaea*, 2011: 1–9. https://doi.org/10.1155/2011/409156

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermiin, L.S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6): 587–589. https://doi.org/10.1038/nmeth.4285

Kamminga, H. 1982. Life from space: A history of panspermia. *Vistas in Astronomy*, 26(2): 67–86. https://doi.org/10.1016/0083-6656(82)90001-0

Kanehisa, M., Sato, Y. and Kawashima, M. 2022. KEGG mapping tools for uncovering hidden features in biological data. *Protein Science*, 31(1): 47–53. https://doi.org/10.1002/pro.4172

Karatan, E., Duncan, T.R. and Watnick, P.I. 2005. NspS, a predicted polyamine sensor, mediates activation of Vibrio cholerae biofilm formation by norspermidine. *Journal of Bacteriology*, 187(21): 7434–7443. https://doi.org/10.1128/JB.187.21.7434-7443.2005

Kato, K., Nakayoshi, T., Kurimoto, E. and Oda, A. 2020. Mechanisms of deamidation of asparagine residues and effects of main-chain conformation on activation energy. *International Journal of Molecular Sciences*, 21(19): 1–14. https://doi.org/10.3390/ijms21197035

Kawai, G., Yamamoto, Y., Kamimura, T., Masegi, T., Sekine, M., Hata, T., Iimori, T., Watanabe, T., Miyazawa, T. and Yokoyama, S. 1992. Conformational rigidity of specific pyrimidine residues in tRNA arises from posttranscriptional modifications that enhance steric interaction between the base and the 2'-hydroxyl group. *Biochemistry*, 31(4): 1040–1046. https://doi.org/10.1021/bi00119a012

Kim, S.H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C. and Rich, A. 1974. three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science*, 185(4149): 435–440. https://doi.org/10.1126/science.185.4149.435

Kingsolver, J.G. and Woods, H.A. 2016. Beyond thermal performance curves: Modeling time-dependent effects of thermal stress on ectotherm growth rates. *American Naturalist*, 187(3): 283–294. https://doi.org/10.1086/684786

Kisselev, A.F., Songyang, Z. and Goldberg, A.L. 2000. Why does threonine, and not serine, function as the active site nucleophile in proteasomes? *Journal of Biological Chemistry*, 275(20): 14831–14837. https://doi.org/10.1074/jbc.275.20.14831

Klatt, C.G., Inskeep, W.P., Herrgard, M.J., Jay, Z.J., Rusch, D.B., Tringe, S.G., Parenteau, M.N., Ward, D.M., Boomer, S.M., Bryant, D.A. and Miller, S.R. 2013. Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. *Frontiers in Microbiology*, 4(JUN). https://doi.org/10.3389/fmicb.2013.00106

Koonin, E. v., Makarova, K.S. and Aravind, L. 2001. horizontal gene transfer in prokaryotes: Quantification and classification. *Annual Review of Microbiology*, 55(1): 709–742. https://doi.org/10.1146/annurev.micro.55.1.709

Kreil, D.P. and Ouzounis, C.A. 2001. Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Research*, 29(7): 1608–1615. https://doi.org/10.1093/nar/29.7.1608

Kuhn, H.J., Cometta, S. and Fiechter, A. 1980. Effects of growth temperature on maximal specific growth rate, yield, maintenance, and death rate in glucose-limited continuous culture of the thermophilic *Bacillus caldotenax*. *European Journal of Applied Microbiology and Biotechnology*, 10: 303–315. https://doi.org/10.1300/J123v27n02_11

Kumagai, I., Watanabe, K. and Oshima, T. 1980. Thermally induced biosynthesis of 2'-O-methylguanosine in tRNA from an extreme thermophile, *Thermus thermophilus* HB27. *Proceedings of the National Academy of Sciences*, 77(4): 1922–1926. https://doi.org/10.1073/pnas.77.4.1922

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6): 1547–1549. https://doi.org/10.1093/molbev/msy096

Kumar, S., Tsai, C.-J. and Nussinov, R. 2000. Factors enhancing protein thermostability and starch processing, production of high fructose corn syrup. *Protein Engineering*, 13(3): 179–191. https://doi.org/10.1093/protein/13.3.179

Kuo, C.H., Moran, N.A. and Ochman, H. 2009. The consequences of genetic drift for bacterial genome complexity. *Genome Research*, 19(8): 1450–1454. https://doi.org/10.1101/gr.091785.109

Lamanna, C. 1940. Relation between temperature growth range and size in the genus *Bacillus*. *Journal of Bacteriology*, 39(5): 593–596. https://doi.org/10.1128/jb.39.5.593-596.1940

Lambros, R.J., Mortimer, J.R. and Forsdyke, D.R. 2003. Optimum growth temperature and the base composition of open reading frames in prokaryotes. *Extremophiles*, 7(6): 443–450. https://doi.org/10.1007/s00792-003-0353-4

Lapierre, P. and Gogarten, J.P. 2009. Estimating the size of the bacterial pan-genome. *Trends in Genetics*, 25(3): 107–110. https://doi.org/10.1016/j.tig.2008.12.004

Lau, M.C.Y., Aitchison, J.C. and Pointing, S.B. 2009. Bacterial community composition in thermophilic microbial mats from five hot springs in central Tibet. *Extremophiles*, 13(1): 139–149. https://doi.org/10.1007/s00792-008-0205-3

Lee, E.M., Ahn, S.H., Park, J.H., Lee, J.H., Ahn, S.C. and Kong, I.S. 2004. Identification of oligopeptide permease (opp) gene cluster in Vibrio fluvialis and characterization of biofilm production by oppA knockout mutation. *FEMS Microbiology Letters*, 240(1): 21–30. https://doi.org/10.1016/j.femsle.2004.09.007

Lessard, I.A.D. and Walsh, C.T. 1999. VanX, a bacterial D-alanyl-D-alanine dipeptidase: Resistance, immunity, or survival function? *Proceedings of the National Academy of Sciences, USA*, 96: 11028–11032. https://doi.org/10.1073/pnas.96.20.11028

Letunic, I., Khedkar, S. and Bork, P. 2021. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1): D458–D460. https://doi.org/10.1093/nar/gkaa937

Li, S. and Hong, M. 2011. Protonation, tautomerization, and rotameric structure of histidine: A comprehensive study by magic-angle-spinning solid-state NMR. *Journal of the American Chemical Society*, 133(5): 1534–1544. https://doi.org/10.1021/ja108943n

Li, W., O'Neill, K.R., Haft, D.H., Dicuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M.K., Durkin, A.S., Gonzales, N.R., Gwadz, M., Lanczycki, C.J., Song, J.S., Thanki, N., Wang, J., Yamashita, R.A., Yang, M., Zheng, C., Marchler-Bauer, A. and Thibaud-Nissen, F. 2021. RefSeq: Expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic Acids Research*, 49(D1): D1020–D1028. https://doi.org/10.1093/nar/gkaa1105

Li, X., Lin, C. and O'connor, P.B. 2010. Glutamine deamidation: Differentiation of glutamic acid and γ-glutamic acid in peptides by electron capture dissociation. *Analytical Chemistry*, 82(9): 3606–3615. https://doi.org/10.1021/ac9028467

Littlechild, J.A., Guy, J.E. and Isupov, M.N. 2004. Hyperthermophilic dehydrogenase enzymes. *Biochemical Society Transactions*, 32(2): 255–258. https://doi.org/10.1042/bst0320255

Lorenz, C., Lünse, C.E. and Mörl, M. 2017. tRNA modifications: Impact on structure and thermal adaptation. *Biomolecules*, 7(2):35. https://doi.org/10.3390/biom7020035

Mahale, K.N., Kempraj, V. and Dasgupta, D. 2012. Does the growth temperature of a prokaryote influence the purine content of its mRNAs? *Gene*, 497(1): 83–89. https://doi.org/10.1016/j.gene.2012.01.040

Mallick, P., Boutz, D.R., Eisenberg, D. and Yeates, T.O. 2002. Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds. *Proceedings of the National Academy of Sciences*, 99(15): 9679–9684. https://doi.org/10.1073/pnas.142310499

Maniloff, J. 1996. The minimal cell genome: “on being the right size”. *Proceedings of the National Academy of Sciences*, 93(19): 10004–10006. https://doi.org/10.1073/pnas.93.19.10004

Manni, M., Berkeley, M.R., Seppey, M. and Zdobnov, E.M. 2021. BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, 1(12). https://doi.org/10.1002/cpz1.323

Martins, E.P. and Hansen, T.F. 1997. Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *The American Naturalist*, 149(4): 646–667. https://www.jstor.org/stable/2463542

Martins, P., Alexandre, J., Diniz-Filho, F. and Housworth, E.A. 2002. Adaptive constraints and the phylogenetic comparative method: A computer simulation test. *Evolution*, 56(1): 1–13. https://doi.org/10.1554/0014-3820(2002)056[0001:ACATPC]2.0.CO;2

McGinnis, M.W., Parker, Z.M., Walter, N.E., Rutkovsky, A.C., Cartaya-Marin, C. and Karatan, E. 2009. Spermidine regulates *Vibrio cholerae* biofilm formation via transport and signaling pathways. *FEMS Microbiology Letters*, 299(2): 166–174. https://doi.org/10.1111/j.1574-6968.2009.01744.x

Medini, D., Donati, C., Tettelin, H., Masignani, V. and Rappuoli, R. 2005. The microbial pan-genome. *Current Opinion in Genetics and Development*, 15(6): 589–594. https://doi.org/10.1016/j.gde.2005.09.006

Mih, N., Brunk, E., Chen, K., Catoiu, E., Sastry, A., Kavvas, E., Monk, J.M., Zhang, Z. and Palsson, B.O. 2018. Ssbio: A Python framework for structural systems biology. *Bioinformatics*, 34(12): 2155–2157. https://doi.org/10.1093/bioinformatics/bty077

Miller, S.R. and Carvey, D. 2019. Ecological divergence with gene flow in a thermophilic cyanobacterium. *Microbial Ecology*, 78(1): 33–41. https://doi.org/10.1007/s00248-018-1267-0

Miller, S.R. and Castenholz, R.W. 2000. Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. *Applied and Environmental Microbiology*, 66(10): 4222–4229. https://doi.org/10.1128/AEM.66.10.4222-4229.2000

Miller, S.R., McGuirl, M.A. and Carvey, D. 2013. The evolution of RuBisCO stability at the thermal limit of photoautotrophy. *Molecular Biology and Evolution*, 30(4): 752–760. https://doi.org/10.1093/molbev/mss327

Miller, S.R., Strong, A.L., Jones, K.L. and Ungerer, M.C. 2009. Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied and Environmental Microbiology*, 75(13): 4565–4572. https://doi.org/10.1128/AEM.02792-08

Miller, S.R., Wingard, C.E. and Castenholz, R.W. 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. *Applied and Environmental Microbiology*, 64(10): 3893–3899. https://doi.org/10.1128/aem.64.10.3893-3899.1998

Minh, B.Q., Hahn, M.W. and Lanfear, R. 2020. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution*, 37(9): 2727–2733. https://doi.org/10.1093/molbev/msaa106

Mira, A., Martín-Cuadrado, A.B., D'Auria, G. and Rodríguez-Valera, F. 2010. The bacterial pan-genome: A new paradigm in microbiology. *International Microbiology*, 13(2): 45–57. https://doi.org/10.2436/20.1501.01.110

Mira, A., Ochman, H. and Moran, N.A. 2001. Deletional bias and the evolution of bacterial genomes. *TRENDS in Genetics*, 17(10): 589–596. https://doi.org/10.1016/s0168-9525(01)02447-7

Missoury, S., Plancqueel, S., de La Sierra-Gallay, I.L., Zhang, W., Liger, D., Durand, D., Dammak, R., Collinet, B. and van Tilbeurgh, H. 2018. The structure of the TsaB/TsaD/TsaE complex reveals an unexpected mechanism for the bacterial t6A tRNA-modification. *Nucleic Acids Research*, 46(11): 5850–5860. https://doi.org/10.1093/nar/gky323

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D. and Bateman, A. 2021. Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1): D412–D419. https://doi.org/10.1093/nar/gkaa913

Moore, K.R., Magnabosco, C., Momper, L., Gold, D.A., Bosak, T. and Fournier, G.P. 2019. An Expanded Ribosomal Phylogeny of Cyanobacteria Supports a Deep Placement of Plastids. *Frontiers in Microbiology*, 10(7): 1–14. https://doi.org/10.3389/fmicb.2019.01612

Moore, L.R., Goericke, R. and Chisholm, S.W. 1995. Comparative physiology of *Synechococcus* and *Prochlorococcus*: influence of light and temperature on growth, pigments, fluorescence and absorptive properties. *Marine Ecology Progress Series*, 116(1/3): 259–275. http://www.jstor.org/stable/44635011

Moran, N.A. and Wernegreen, J.J. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology & Evolution*, 15(8): 321–326. https://doi.org/10.1016/S0169-5347(00)01902-9

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valin, F. and Bernardi, G. 2005. The correlation between genomic G+C and optimal growth temperature of prokaryotes is robust: A reply to Marashi and Ghalanbor. *Biochemical and Biophysical Research Communications*, 330(20): 357–360. https://doi.org/10.1016/j.bbrc.2005.02.133

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F. and Bernardi, G. 2006. Genomic GC level, optimal growth temperature, and genome size in prokaryotes. *Biochemical and Biophysical Research Communications*, 347(1):1–3. https://doi.org/10.1016/j.bbrc.2006.06.054

Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A., McDonald, L., Utterback, T.R., Malek, J.A., Linher, K.D., Garrett, M.M., Stewart, A.M., Cotton, M.D., Pratt, M.S., Phillips, C.A., Richardson, D., Heidelberg, J., Sutton, G.G., Fleischmann, R.D., Eisen, J.A., White, O., Salzberg, S.L., Smith, H.O., Venter, J.C. and Fraser, C.M. 1999. Evidence for lateral gene transfer between archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, 399(6734): 323–329. https://doi.org/10.1038/20601

Nguyen, L.T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1): 268–274. https://doi.org/10.1093/molbev/msu300

Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O. and Thiele, I. 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences*, 109(7): 2678–2683. https://doi.org/10.1073/pnas.1117907109

Nordström, K.M. and Laakso, S. v. 1992. Effect of growth temperature on the fatty acid composition of ten *Thermus* strains. *Applied and Environmental Microbiology*, 58(5): 1656–1660. https://doi.org/10.1128/aem.58.5.1656-1660.1992

Nowack, S., Olsen, M.T., Schaible, G.A., Becraft, E.D., Shen, G., Klapper, I., Bryant, D.A. and Ward, D.M. 2015. The molecular dimension of microbial species: 2. *Synechococcus* strains representative of putative ecotypes inhabiting different depths in the Mushroom Spring microbial mat exhibit different adaptive and acclimative responses to light. *Frontiers in Microbiology*, 6(6): 1–13. https://doi.org/10.3389/fmicb.2015.00626

Nübel, U., Garcia-Pichel, F. and Muyzer, G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Applied and Environmental Microbiology*, 63(8): 3327–3332. https://doi.org/10.1128/aem.63.8.3327-3332.1997

Ochi, A., Makabe, K., Yamagami, R., Hirata, A., Sakaguchi, R., Hou, Y.M., Watanabe, K., Nureki, O., Kuwajima, K. and Hori, H. 2013. The catalytic domain of topological knot tRNA methyltransferase (TrmH) discriminates between substrate tRNA and nonsubstrate tRNA via an induced-fit process. *Journal of Biological Chemistry*, 288(35): 25562–25574. https://doi.org/10.1074/jbc.M113.485128

Ohkubo, S. and Miyashita, H. 2017. A niche for cyanobacteria producing chlorophyll *f* within a microbial mat. *ISME Journal*, 11(10): 2368–2378. https://doi.org/10.1038/ismej.2017.98

Olsen, M.T., Nowack, S., Wood, J.M., Becraft, E.D., LaButti, K., Lipzen, A., Martin, J., Schackwitz, W.S., Rusch, D.B., Cohan, F.M., Bryant, D.A. and Ward, D.M. 2015. The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* strains with different light responses and in situ diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. *Frontiers in Microbiology*, 6(6): 1–13. https://doi.org/10.3389/fmicb.2015.00604

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F. and Stevens, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1). https://doi.org/10.1093/nar/gkt1226

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A. and Parkhill, J. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, 31(22): 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Papke, R.T., Ramsing, N.B., Bateson, M.M. and Ward, D.M. 2003. Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology*, 5(8): 650–659. https://doi.org/10.1046/j.1462-2920.2003.00460.x

Paradis, E. and Schliep, K. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3): 526–528. https://doi.org/10.1093/bioinformatics/bty633

Partridge, S.M. and Davis, H.F. 1950. Preferential release of aspartic acid during the hydrolysis of proteins. *Nature*, 165(4185): 62–63. https://doi.org/10.1038/165062a0

Paz, A., Mester, D., Baca, I., Nevo, E. and Korol, A. 2004. Adaptive role of increased frequency of polypurine tracts in mRNA sequences of thermophilic prokaryotes. *Proceedings of the National Academy of Sciences*, 101(9): 2951–2956. https://doi.org/10.1073/pnas.0308594100

Pedersen, D. and Miller, S.R. 2017. Photosynthetic temperature adaptation during niche diversification of the thermophilic cyanobacterium *Synechococcus* A/B clade. *ISME Journal*, 11(4): 1053–1057. https://doi.org/10.1038/ismej.2016.173

Pedone, E., Ren, B., Ladenstein, R., Rossi, M. and Bartolucci, S. 2004. Functional properties of the protein disulfide oxidoreductase from the archaeon *Pyrococcus furiosus*. *European Journal of Biochemistry*, 271(16): 3437–3448. https://doi.org/10.1111/j.0014-2956.2004.04282.x

Pena, R.T., Blasco, L., Ambroa, A., González-Pedrajo, B., Fernández-García, L., López, M., Bleriot, I., Bou, G., García-Contreras, R., Wood, T.K. and Tomás, M. 2019. Relationship between quorum sensing and secretion systems. *Frontiers in Microbiology*, 10: 1100. https://doi.org/10.3389/fmicb.2019.01100

Perugino, G., Valenti, A., D'Amaro, A., Rossi, M. and Ciaramella, M. 2009. Reverse gyrase and genome stability in hyperthermophilic organisms. *Biochemical Society Transactions*, 37(1): 69–73. https://doi.org/10.1042/BST0370069

Perutz, M.F. and Raidt, H. 1975. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature*, 255(5505): 256–259. https://doi.org/10.1038/255256a0

Pierpont, C.L., Ohkubo, S., Miyashita, H. and Miller, S.R. 2022. Draft genome sequence of the cyanobacterium *Synechococcus* sp. strain Nb3U1. *Microbiology Resource Announcements*, 11(5): 9–10. https://doi.org/10.1128/mra.00251-22

Pinheiro, J. and Bates, D. 2022. *nlme: Linear and Nonlinear Mixed Effects Models*. 3.1-159 https://cran.r-project.org/web/packages/nlme/nlme.pdf

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. and Korobeynikov, A. 2020. Using SPAdes *de novo* assembler. *Current Protocols in Bioinformatics*, 70(1). https://doi.org/10.1002/cpbi.102

Rampelotto, P.H. 2013. Extremophiles and extreme environments. *Life*, 3: 482–485. https://doi.org/10.3390/life3030482

Rees, D.C., Johnson, E. and Lewinson, O. 2009. ABC transporters: The power to change. *Nature Reviews Molecular Cell Biology*, 10: 218–227. https://doi.org/10.1038/nrm2646

Revsbech, N.P. and Ward, D.M. 1984. Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. *Applied and Environmental Microbiology*, 48(2): 270–275. https://doi.org/10.1128/aem.48.2.270-275.1984

Rivera, M.C., Jain, R., Moore, J.E. and Lake, J.A. 1998. Genomic evidence for two functionally distinct gene classes. *Proceedings of the National Academy of Sciences*, 95(11): 6239–6244. https://doi.org/10.1073/pnas.95.11.6239

Rosen, M.J., Davison, M., Bhaya, D. and Fisher, D.S. 2015. Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. *Science*, 348(6238): 1019–1023. https://doi.org/10.1126/science.aaa4456

Sabath, N., Ferrada, E., Barve, A. and Wagner, A. 2013. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biology and Evolution*, 5(5): 966–977. https://doi.org/10.1093/gbe/evt050

Salichos, L. and Rokas, A. 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, 497(7449): 327–331. https://doi.org/10.1038/nature12130

Saunders, N.F.W., Thomas, T., Curmi, P.M.G., Mattick, J.S., Kuczek, E., Slade, R., Davis, J., Franzmann, P.D., Boone, D., Rusterholtz, K., Feldman, R., Gates, C., Bench, S., Sowers, K., Kadner, K., Aerts, A., Dehal, P., Detter, C., Glavina, T., Lucas, S., Richardson, P., Larimer, F., Hauser, L., Land, M. and Cavicchioli, R. 2003. Mechanisms of thermal adaptatation revealed from genomes of the anatarctic Archaea *Methanogenium frigidum* and *Methanacoccoides burtonii*. *Genome Research*, 13(7): 1580–1588. https://doi.org/10.1101/gr.1180903

Savage, V.M., Gillooly, J.F., Brown, J.H., West, G.B. and Charnov, E.L. 2004. Effects of body size and temperature on population growth. *American Naturalist*, 163(3): 429–441. https://doi.org/10.1086/381872

Schumann, J., Böhm, G., Jaenicke, R., Schumacher, G. and Rudolph, R. 1993. Stabilization of creatinase from *Pseudomonas putida* by random mutagenesis. *Protein Science*, 2(10): 1612–1620. https://doi.org/10.1002/pro.5560021007

Shih, P.M., Wu, D., Latifi, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., Tandeau De Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K.W., Han, C.S., Rubin, E.M., Eisen, J.A., Woyke, T., Gugger, M. and Kerfeld, C.A. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 110(3): 1053–1058. https://doi.org/10.1073/pnas.1217107110

Shuter, B.J., Thomas, J.E., Taylor, W.D. and Zimmerman, A.M. 1983. Phenotypic correlates of genomic DNA content in unicellular eukaryotes and other cells. *The American Naturalist*, 122(1): 26–45. https://www.jstor.org/stable/2461004

Singer, G.A.C. and Hickey, D.A. 2003. Thermophilic prokaryotes have characteristic patterns of codon usage, amino acid composition and nucleotide content. *Gene*, 317: 39–47. https://doi.org/10.1016/S0378-1119(03)00660-7

Smith, R.B. and Braile, L.W. 1994. The Yellowstone hotspot. *Journal of Volcanology and Geothermal Research*, 61(3–4). https://doi.org/10.1016/0377-0273(94)90002-7

Somero, G.N., Lockwood, B.L. and Tomanek, L. 2017. *Biochemical Adaptation: Response to Environmental Challenges from Life's Origins to the Antrhopocene*. First edition. Sunderland, MA, Sinauer Associates, Inc.

Staisch, L.M., O'Connor, J.E., Cannon, C.M., Holm-Denoma, C., Link, P.K., Lasher, J. and Alexander, J.A. 2021. Major reorganization of the Snake River modulated by passage of the Yellowstone Hotspot. *GSA Bulletin*(November): 1–11. https://doi.org/10.1130/b36174.1

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steinberg, R. and Koch, H.G. 2021. The largely unexplored biology of small proteins in pro- and eukaryotes. *FEBS*, 288: 7002–7024. https://doi.org/10.1111/febs.15845

Stewart, P.S. and Franklin, M.J. 2008. Physiological heterogeneity in biofilms. *Nature Reviews Microbiology*, 6: 199–210. https://doi.org/10.1038/nrmicro1838

Stouthamer, A.H. and Bettenhaussen, C.W. 1980. Growth and physiology of potassium-limited chemostat cultures of *Paracoccus denitrificans*. *Archives of Microbiology*, 125: 239–244. https://doi.org/10.1007/BF00446883

Swinehart, W.E. and Jackman, J.E. 2015. Diversity in mechanism and function of tRNA methyltransferases. *RNA Biology*, 12(4): 398–411. https://doi.org/10.1080/15476286.2015.1008358

Sydow, J.F., Lipsmeier, F., Larraillet, V., Hilger, M., Mautz, B., Mølhøj, M., Kuentzer, J., Klostermann, S., Schoch, J., Voelger, H.R., Regula, J.T., Cramer, P., Papadimitriou, A. and Kettenberger, H. 2014. Structure-based prediction of asparagine and aspartate degradation sites in antibody variable regions. *PLoS ONE*, 9(6). https://doi.org/10.1371/journal.pone.0100736

Tatusova, T., Dicuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M. and Ostell, J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44(14): 6614–6624. https://doi.org/10.1093/nar/gkw569

Tekaia, F., Yeramian, E. and Dujon, B. 2002. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. *Gene*, 297(1–2): 51–60. https://doi.org/10.1016/S0378-1119(02)00871-5

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S. v, Crabtree, J., Jones, A.L., Scott Durkin, A., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., B O, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R. and Fraser, C.M. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences*, 102(39): 13950–13955. https://doi.org/10.1073/pnas.0506758102

Thiel, V., Costas, A.M.G., Fortney, N.W., Martinez, J.N., Tank, M., Roden, E.E., Boyd, E.S., Ward, D.M., Hanada, S. and Bryant, D.A. 2019. "*Candidatus thermonerobacter thiotrophicus*," a non-phototrophic member of the Bacteroidetes/Chlorobi with dissimilatory sulfur metabolism in hot spring mat communities. *Frontiers in Microbiology*, 10(1). https://doi.org/10.3389/fmicb.2018.03159

Thomas, T. and Thomas, T.J. 2001. Polyamines in cell growth and cell death: Molecular mechanisms and therapeutic applications. *Cellular and Molecular Life Sciences*, 58(2): 244–258. https://doi.org/10.1007/PL00000852

Ueno, Y. 2003. *GUPPY*. 4.5.4. https://staff.aist.go.jp/yutaka.ueno/guppy/

Villain, E., Fort, P. and Kajava, A. v. 2022. Aspartate-phobia of thermophiles as a reaction to deleterious chemical transformations. *BioEssays*, 44(1). https://doi.org/10.1002/bies.202100213

Vogt, G., Woell, S. and Argos, P. 1997. Protein thermal stability, hydrogen bonds, and ion pairs. *Journal of Molecular Biology*, 269(4): 631–643. https://doi.org/10.1006/jmbi.1997.1042

Wang, Q., Cen, Z. and Zhao, J. 2015. The survival mechanisms of thermophiles at high temperatures: An angle of omics. *Physiology*, 30(2): 97–106. https://doi.org/10.1152/physiol.00066.2013

Ward, D.M. 1978. Thermophilic methanogenesis in a hot-spring algal-bacterial mat (71 to 30 degrees C). *Applied and Environmental Microbiology*, 35(6): 1019–1026. https://doi.org/10.1128/aem.35.6.1019-1026.1978

Ward, D.M., Castenholz, R.W. and Miller, S.R. 2012. Cyanobacteria in Geothermal Habitats. In: *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. pp. 39–63. Dordrecht, Netherlands, Springer. https://doi.org/10.1007/978-94-007-3855-3

West-Eberhard, M.J. 2003. *Developmental Plasticity and Evolution*. New York, NY, Oxford University Press, Inc.

Wick, R.R., Schultz, M.B., Zobel, J. and Holt, K.E. 2015. Bandage: Interactive visualization of de novo genome assemblies. *Bioinformatics*, 31(20): 3350–3352. https://doi.org/10.1093/bioinformatics/btv383

van Wolferen, M., Ajon, M., Driessen, A.J.M. and Albers, S.V. 2013. How hyperthermophiles adapt to change their lives: DNA exchange in extreme conditions. *Extremophiles*, 17(4): 545–563. https://doi.org/10.1007/s00792-013-0552-6

Wood, D.E., Lu, J. and Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20(1). https://doi.org/10.1186/s13059-019-1891-0

Wright, T.H. 1991. Nonenzymatic deamidation of asparaginyl and glutaminyl residues in protein. *Critical Reviews in Biochemistry and Molecular Biology*, 26(1): 1–52. https://doi.org/10.3109/10409239109081719

Zale, S.E. and Klibanov, A.M. 1986. Why does ribonuclease irreversibly inactivate at high temperatures? *Biochemistry*, 25(19): 5432–5444. https://doi.org/10.1021/bi00367a014

Zeldovich, K.B., Berezovsky, I.N. and Shakhnovich, E.I. 2007. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Computational Biology*, 3(1): 0062–0072. https://doi.org/10.1371/journal.pcbi.0030005