Ecological Divergence with Gene Flow in a Thermophilic Cyanobacterium

Scott R. Miller & Darla Carvey

Microbial Ecology

ISSN 0095-3628

Microb Ecol
DOI 10.1007/s00248-018-1267-0

Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC, part of Springer Nature. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com”.

Springer

Author's personal copy

Microbial Ecology
https://doi.org/10.1007/s00248-018-1267-0

ENVIRONMENTAL MICROBIOLOGY

Ecological Divergence with Gene Flow in a Thermophilic Cyanobacterium

Scott R. Miller[1] · Darla Carvey[1]

Received: 22 May 2018 / Accepted: 19 September 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract
How ecological diversity is maintained and distributed within populations is a longstanding question in microbial ecology. In the thermophilic cyanobacterium *Synechococcus* B′, high observed levels of recombination are predicted to maintain ecological variation despite the simultaneous action of diverse selective pressures on different regions of the genome. To investigate ecological diversity in these bacteria, we directly isolated laboratory strains of *Synechococcus* B′ from samples collected along the thermal gradients of two geothermal environments in Yellowstone National Park. Extensive recombination was evident for a multi-locus sequence data set, and, consequently, our sample did not exhibit the sequence clustering expected for distinct ecotypes evolving by periodic clonal selection. Evidence for local selective sweeps at specific loci suggests that sweeps may be common but that recombination is effective for maintaining diversity of unlinked genomic regions. Thermal performance for strain growth was positively associated with the temperature of the environment, indicating that *Synechococcus* B′ populations consist of locally adapted ecological specialists that occupy specific thermal niches. Because this ecological differentiation is observed despite the absence of dispersal barriers among sites, we conclude that these bacteria may freely exchange much of the genome but that barriers to gene flow exist for loci under direct temperature selection.

Keywords Diversity · Recombination · Adaptation · Selective sweep · Thermophile

Introduction

Microbial populations often exhibit high levels of sequence diversity [1–4]. Understanding the factors that contribute to the origins, maintenance, and distribution of this diversity remains one of the central goals of microbial ecology and evolution. In particular, the nature of the role of recombination during ecological divergence is not clear. Moderate levels of recombination are expected to prevent divergence in the absence of selection [5]. If recombination is weak compared to natural selection, however, as proposed by the ecotype model of ecological specialization [6], selection both removes genome-wide variation within populations and drives genomic divergence between ecologically distinct populations. As a result, genomic diversity becomes packaged in distinct clonal lineages, or ecotypes. Some microbial assemblages conform to the ecotype model (e.g., [3]). By contrast, recombination has been demonstrated to maintain genomic diversity for several ecologically differentiated archaeal and bacterial populations by preventing genome-wide sweeps [1, 2, 7]. In these cases, diverging genomes may potentially become more isolated over time as sequential selective events reduce recombination at niche-defining loci, as predicted by the "fragmented speciation" model [8].

Recently, Rosen et al. [9] reported an extreme case of the recombination-mediated maintenance of genomic diversity for a population of the thermophilic cyanobacterium *Synechococcus* B′ from a microbial mat community in Yellowstone National Park. Deep sequencing of multiple loci from environmental DNA revealed high rates of homologous recombination in the population; notably, the degree of non-random associations (i.e., linkage disequilibrium, LD) between alleles at variable nucleotide sites decayed at rates typical of what is expected for sexually reproducing species. Consequently, these bacteria exhibit mosaic genomes of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00248-018-1267-0) contains supplementary material, which is available to authorized users.

✉ Scott R. Miller
scott.miller@umontana.edu

[1] Division of Biological Sciences, The University of Montana, 32 Campus Dr. #4824, Missoula, MT 59812, USA

Published online: 28 September 2018

Author's personal copy

randomly assorted genetic variation rather than ecologically distinct genomic backbones. Extensive recombination maintains diversity by restricting the purging of variation that occurs during a sweep to the local regions of the genome that are directly under selection.

Although phenotypic diversity was not directly examined by Rosen et al. [9], an implication of their results is that ecological variation should persist in the population despite diverse selective pressures simultaneously acting on individual regions of the genome. However, it is not clear whether ecological diversity is indeed maintained or how it is distributed along the environmental gradients characteristic of these systems. Because *Synechococcus* B′ cells can be distributed broadly along the thermal gradients of these environments [10], spatial variation may potentially facilitate the maintenance of locally adaptive variation as a consequence of a balance between the opposing forces of migration along the gradient, which introduces diversity, and selection, which removes it [11–13]. In the absence of barriers to dispersal, this could occur without evidence for genetic structure throughout much of the genome, since the extensive recombination exhibited by these bacteria is expected to homogenize neutral variation. Alternatively, with dispersal barriers, these bacteria may exhibit ecotypic differentiation at the landscape scale, for the reason that genetic divergence is expected to increase between populations with physical distance when gene flow is restricted [14].

To address these issues, we directly isolated a sample of *Synechococcus* B′ laboratory strains from environmental samples collected along the thermal gradients of two Yellowstone geothermal environments for which these bacteria comprise a significant component of the microbial communities based on an rRNA-barcoding analysis of environmental DNA [10]. We first used multi-locus sequence data to test whether strains exhibit a mosaic genetic structure comparable to the Rosen et al. [9] environmental sample, as well as whether there is evidence for barriers to dispersal among sites. We next assessed whether there is evidence for local selective sweeps, which are predicted to occur but which were not directly observed by Rosen et al. [9]. We then assayed the temperature dependence of growth for a subset of strains to investigate the diversity in thermal performance of the sample. Finally, we evaluated whether the observed ecological diversity of the strains is tightly associated with position along the environmental gradients of these communities, as expected if there has been local adaptation.

Materials and Methods

Strain Isolation

Strains of *Synechococcus* B′ were directly isolated from samples collected in June 2008 from White Creek and Rabbit Creek in Yellowstone National Park. Samples were taken from three sites at White Creek (WC5, 54 °C; WC6, 57 °C; WC7, 61 °C) and from five sites at Rabbit Creek (RC3, 51 °C; RC5, 55 °C; RC6, 61 °C; RC7, 63.5 °C; RC8, 63.5 °C) that had been previously determined by rRNA-barcoding to contain these bacteria in high abundance [10]. Environmental samples were sterilely collected with a 3-mL syringe, transferred to a sterile microfuge tube and stored in the dark at ambient temperature until return to the laboratory at the University of Montana. Isolations were performed at either 50 °C (RC3) or 55 °C (all other samples). Direct isolation (i.e., without enrichment) was accomplished by the application of serially diluted microbial mat sample homogenates on glass fiber filters. Filters were incubated in flasks containing mineral salts medium D [15] under 100 μmol photons m^{-2} s^{-1} of cool-white fluorescent light in a temperature controlled growth chamber. Blue-green colonies developed within 1 week of filtering and were subsequently transferred to a fresh sterile tube of medium D with sterilized forceps. To confirm that strains belonged to the *Synechococcus* B′ lineage, we amplified and sequenced partial fragments of the 16S rRNA gene of each strain as in Miller and Castenholz [16]. Amplified fragments were cleaned by QIAquick PCR purification (QIAGEN) and then Sanger sequenced with an ABI 3130 genetic analyzer at the University of Montana Murdock Lab DNA Sequencing Facility. Strain sequences were ~99.5% identical to the *Synechococcus* strain OS-B′. Sequence data were deposited in the GenBank database (accession numbers MH709082-MH709095).

Multi-locus Sequence Analysis and Population Genetics

Strains were genotyped by sequencing at 11 loci distributed around the *Synechococcus* OS-B′ genome [17]. Primers for PCR amplification of these loci (Supplementary Table 1) were designed using shotgun-sequenced metagenome data obtained from White Creek as part of the Yellowstone Metagenome Working Group [18, 19]. By using this approach, we did not introduce potential bias regarding the history of selection on loci. Paired-end sequence reads from the metagenome indicated strong conservation of local synteny with the published *Synechococcus* OS-B′ genome. All fragments were amplified as 50-μL reactions with an MJ Research PTC-100 thermal cycler (MJ Research, Waltham, MA, USA) for 40 cycles of 94 °C for 1 min, either 52 °C (4719, 3496), 55 °C (1837, 2490, 1124, C311, 2807, 4733), or 58 °C (C88, 4043) for 1 min, and 72 °C for 1 min. Amplified fragments were cleaned and sequenced as above. Sequence data were deposited in the GenBank database (accession numbers MH688302-MH688452, MH703909-MH704089, MH709272-MH709307 and MH746486-MH746523). To infer the evolutionary relationships for individual loci and for a concatenated

Author's personal copy

data set, neighbor net genealogical networks were reconstructed with SplitsTree version 4.14.4 [20] using the default setting of uncorrected P-distance (appropriate for these data, since SNPs were two-variant). Similar results were obtained with the Jukes-Cantor model of estimating evolutionary distance between sequences, as well as for other network-building approaches such as median joining. Population genetic analyses were performed with DnaSP version 5 [21]. To test for genetic isolation by distance, we performed Mantel tests with IBD version 1.52 [22].

Growth Experiments

Growth rates of strains were assayed for duplicate independent cultures over the range of 37–66 °C under 100 μmol photons m^{-2} s^{-1} of cool-white fluorescent light on a 12/12-h light/dark cycle. Cell homogenates were inoculated into Erlenmeyer flasks containing 75 mL of medium D to a starting optical density (OD_{750}) of 0.005, as determined with a Beckman Coulter DU 530 spectrophotometer (Brea, CA, USA). Flask position was randomized every 24 h, and growth was monitored as the change in OD_{750} over time. The generation time during exponential growth phase was estimated from the regression of logarithmically transformed OD_{750} data on time, and results were reported as population doublings per day.

Results and Discussion

A Diverse Sample of *Synechococcus* B′ Strains Exhibits Extensive Recombination and No Dispersal Barriers

Synechococcus B′ populations are major components of the microbial mat communities that develop below 65 °C at Rabbit Creek and White Creek in Yellowstone National Park [10]. *Synechococcus* B′ sequences can account for up to ~ 40% of the rRNA barcodes recovered from samples collected from these sites [10]. We collected samples from established *Synechococcus* B′-containing sites at Rabbit Creek (a 140-m stretch spanning 51–63.5 °C) and White Creek (a 120-m stretch spanning 54–61 °C), respectively. To obtain a representative sample of *Synechococcus* B′ diversity without enrichment, we applied serially diluted microbial mat sample homogenates to sterilized glass fiber filters and picked individual blue-green colonies that developed following incubation of the filters in liquid medium (see the "Materials and Methods" section). By this approach, we obtained 29 strains from Rabbit Creek and 11 strains from White Creek (the letter in the strain name indicates creek of origin, and the first digit indicates the collection site as in Miller et al. [10], with increasing number more upstream). We confirmed through sequencing of an approximately 1-kbp fragment of the 16S rRNA gene that each strain belonged to the *Synechococcus* B′ lineage.

We next obtained sequence data for the strains at 11 loci distributed around the *Synechococcus* strain OS-B′ genome (Fig. 1). For each locus, the maximum sequence identity between one of our sampled alleles and the strain OS-B′ genome ranged from 99% (3496) to 100% (865, 4733, 2807, C311). Based on the observation of extensive recombination within a deeply sequenced *Synechococcus* B′ population [9], we expected that the relationships among strains would be better described by a genealogical network rather than by a traditional bifurcating phylogeny, which does not accommodate recombination. This was indeed the case. A neighbor net network [23] inferred for a concatenated alignment of the sequence data for the 11 markers (approximately 6000 nucleotides) showed evidence of substantial recombination, i.e., the loops in the network (Fig. 2).

Strains did not generally cluster by site in the network. The broad distribution of genetic variation among locations suggests that *Synechococcus* is not dispersal-limited over the spatial scale of sampling (White Creek and Rabbit Creek are separated by ~2 km). This conclusion was corroborated by the lack of a significant positive correlation between genetic distance and geographic distance ($P = 0.13$ by a Mantel test). The sequenced loci also appear to be neutral with respect to environmental temperature: the partial correlation of genetic distance and temperature difference between sites (controlled for geographic distance) was also not significant ($P = 0.11$).

Rosen et al. [9] reported that linkage disequilibrium (LD, the degree of nonrandom association between variable nucleotide positions) decayed approximately 10-fold within several hundred nucleotides in their environmental sample as a consequence of high amounts of recombination. Based on these results, we should expect to observe similar levels of LD decay within loci for our sample. We measured LD as the square

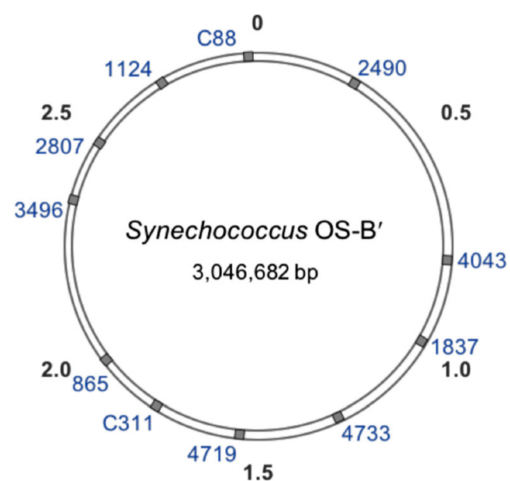

Fig. 1 Locations of the 11 sequenced loci on the *Synechococcus* strain OS-B′ genome

Author's personal copy

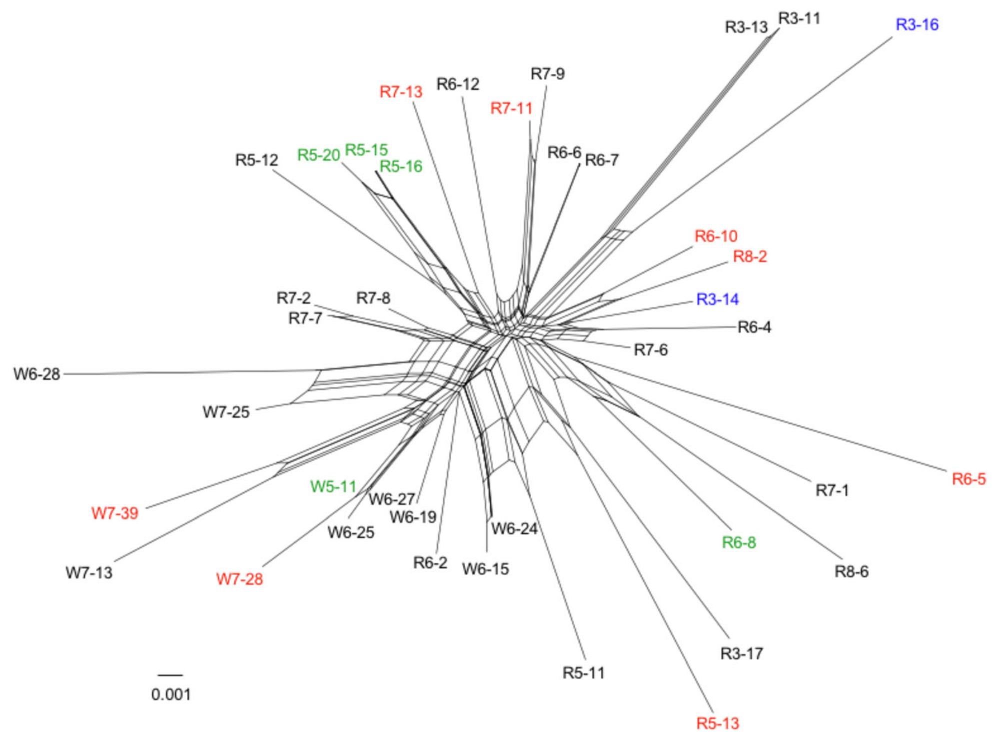

Fig. 2 Neighbor net network of the relationships among 40 *Synechococcus* B′ strains from Rabbit Creek and White Creek inferred for a concatenated alignment of 6309 nucleotides from 11 markers from throughout the genome. Scale bar is in units of nucleotide substitutions per site. Strains for which growth rate data were collected are color-coded with respect to whether no growth was observed at 55 °C (blue), no growth was observed at 60 °C (green), or growth was observed at ≥ 60 °C (red), respectively

of the correlation (r^2) between haplotypes at two polymorphic nucleotide positions (i.e., SNPs), which takes on values between 0 (at equilibrium) and 1 (when the loci are completely linked). We observed a similar general pattern for individual loci from our sample of strains (Fig. 3; Supplementary Fig. 1). Seven of the loci exhibited a rapid decay of LD (Fig. 3a, b; Supplementary Fig. 1a–f). However, for four loci, strong LD was observed between sites separated by hundreds of nucleotides (Fig. 3c, d; Supplementary Fig. 1c, g). For the latter loci, LD persists due to the limited number of recombination events between polymorphic sites. These differences are evident in the genealogical networks for individual genes. For example, the C88 network is highly reticulated due to historic recombination events that break up linkage between SNPs (Fig. 4a). By contrast, the loci with high levels of LD (e.g., 4733 in Fig. 4b) exhibited a star phylogeny structure with reduced levels of variation radiating from a central node, few recombination events (and therefore few cycles in the network), and the persistence of ancestral alleles (i.e., sampled alleles at internal nodes in the network).

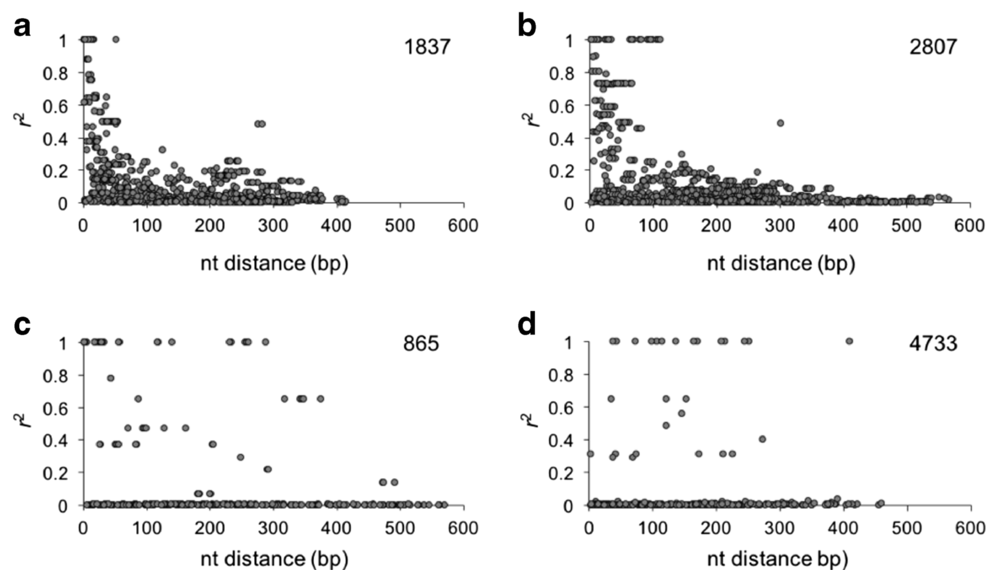

Fig. 3 Linkage disequilibrium (r^2) between polymorphic nucleotide sites as a function of nucleotide distance for the following loci: 1837 (**a**); 2807 (**b**); 865 (**c**); and 4733 (**d**)

Author's personal copy

Fig. 4 Neighbor net networks for *Synechococcus* B′ strains for C88 (**a**) and 4733 (**b**). Scale bars are in units of nucleotide substitutions per site

Author's personal copy

Evidence for Local Selective Sweeps

Rosen et al. [9] predicted but did not observe the existence of local selective sweeps in the *Synechococcus* B′ population. Our sample of different regions of the genome offered the potential to detect local sweeps. A gene-specific star phylogeny (Fig. 4b) can potentially be caused by a local selective sweep in the evolutionary past, as new diversity arises from a common ancestral allele following the purging of genetic variation from a population. Because a region of the genome that has recently experienced a selective sweep should exhibit a number of sequence signatures, we estimated several population genetic parameters from the sequence data to further evaluate the evidence for past sweeps (Table 1). For example, sequence diversity of a sample should be low following a sweep. In addition, we expect to observe an excess of neutral SNPs at low frequency in the sample, indicated by a significantly negatively distorted value of Tajima's *D* [24]. In genealogical terms, this corresponds to the case for which the terminal branches of a genealogy are abnormally long compared with neutral expectations (i.e., observed SNPs represent relatively new variation). Finally, we also expect a comparatively low number of recombination events since the sweep, because there has been insufficient time for them to accumulate.

In our sample, mean silent nucleotide diversity (π_{silent}; i.e., synonymous codon sites and non-coding sequence) for the sample was 0.04 (Table 1), comparable to the value of 0.03 in the Rosen et al. [9] sample. Two of the sampled loci (865, 4733) exhibited evidence of all three of the above signatures of a selective sweep (Table 1). Both loci have comparatively low π_{silent} levels. Furthermore, the amount of population polymorphism relative to sequence divergence (π_{silent}/K_{silent}) from a more thermotolerant congener (*Synechococcus* strain OS-A; [17]) is also low for 865 (Table 1; the OS-A genome lacks an ortholog of 4733). Because silent sites are not necessarily selectively neutral, normalization by divergence controls for potential differences among loci in the constraints on evolutionary rate. We can conclude that it is unlikely that the low diversity at this locus is due to an intrinsically low evolutionary rate. In addition, 865 and 4733 exhibit significantly negative values of Tajima's *D* as well as a low number of estimated recombination events (Table 1). Two additional loci (1124, 4043) show more modest support for a past sweep based on more than one of these signatures (Table 1).

The above results indicate that selective sweeps have shaped *Synechococcus* B′ genome evolution. However, these sweeps are local, since most loci in our data set have maintained high levels of variation. This variation is maintained by the extensive recombination that we detected, which breaks up linkage between the target(s) of selection and other regions of the genome.

These evolutionary dynamics are consistent with the results of Rosen et al. [9], are analogous to patterns observed for sexually reproducing eukaryotes, and have been observed for other bacterial populations as well [2]. They are fundamentally different from the ecotype model [6], which posits that

Table 1 Diversity among loci for *Synechococcus* B′ strains

Locus	N[a]	Length (nt)[b]	No. of SNPs	π_{total}[c]	π_{silent}[d]	K_{silent}[e]	π_{silent}/K_{silent}	D[f]	R_M[g]
2807	38	606	44	0.023	0.038	0.50	0.11	1.15	12
C88	39	537	45	0.022	0.081	0.62	0.13	0.30	16
865	40	612	26	0.004	0.010	0.46	0.02	− 2.07*	2
1837	36	433	42	0.035	0.079	0.49	0.16	1.63	15
1124	40	562	27	0.007	0.013	0.49	0.03	− 1.44	4
3496	39	649	46	0.022	0.073	0.64	0.11	1.02	14
2490	38	614	79	0.017	0.023	0.43	0.10	− 1.75	5
4043	33	622	24	0.008	0.011	0.68	0.07	− 0.77	2
4719	37	533	85	0.026	0.036	0.20	0.20	− 1.23	5
4733	34	553	25	0.005	0.005	–[h]	–	− 1.87*	1
C311	39	588	79	0.023	0.070	0.62	0.11	− 1.09	16

[a] Sample size

[b] nt, nucleotide

[c] Total nucleotide diversity

[d] Nucleotide diversity at silent nucleotide sites

[e] Jukes-Cantor corrected silent nucleotide site divergence from the *Synechococcus* strain OS-A genome

[f] Tajima's *D*

[g] Minimum number of recombination events

[h] Locus is not present in the *Synechococcus* strain OS-A genome

*$P < 0.05$

Author's personal copy

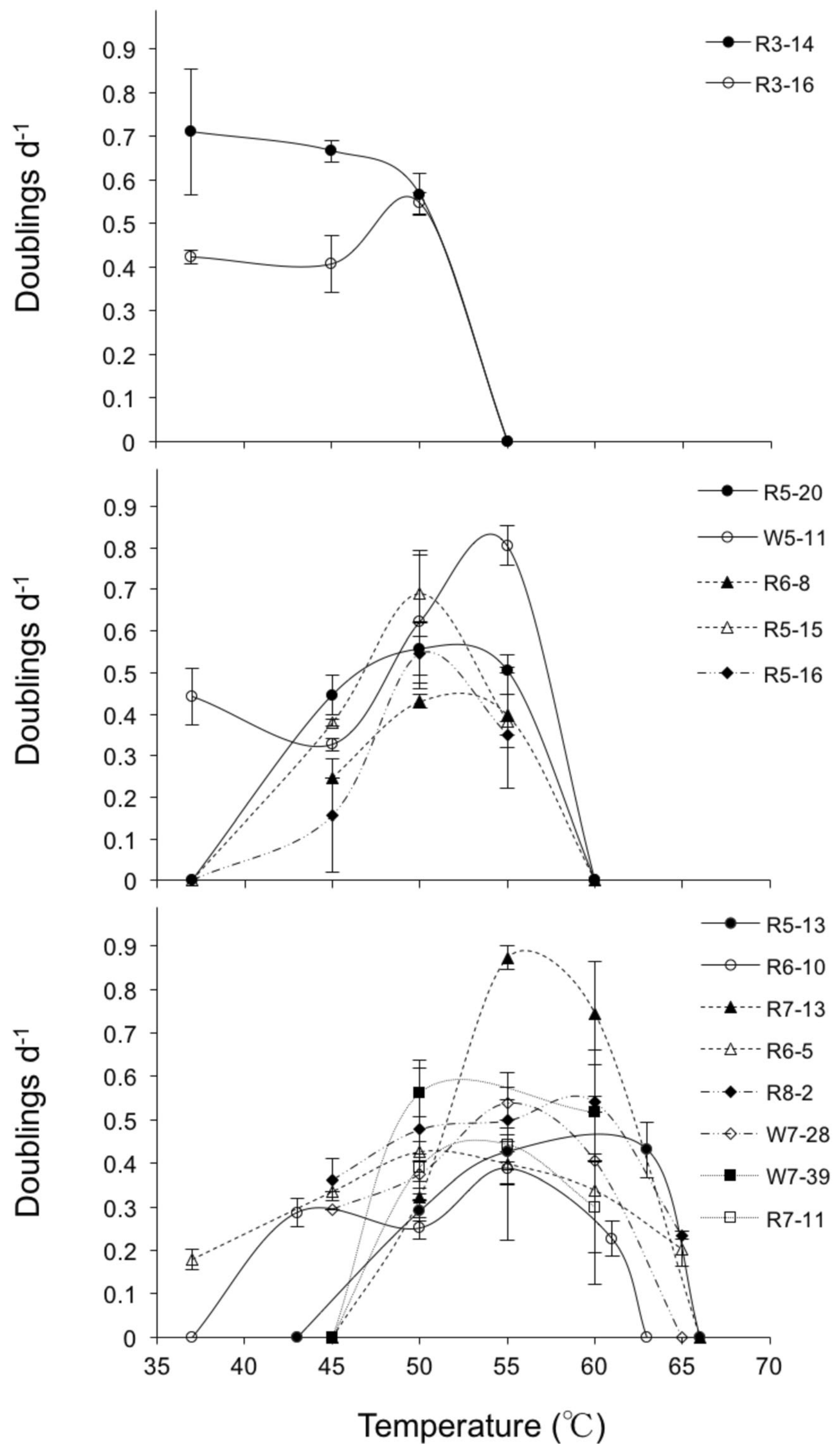

Fig. 5 Temperature dependence of population growth rate in doublings per day for a sample of 15 *Synechococcus* B′ strains. Error bars are two standard errors. Strains are distinguished by whether no growth was observed at 55 °C (top panel), no growth was observed at 60 °C (middle panel), or growth was observed at ≥ 60 °C (bottom panel), respectively

the fixation of beneficial mutations results in the genome-wide removal of variation from the population. At these loci, all Rabbit Creek and White Creek alleles have descended from a beneficial allele that was fixed in an ancestral population (e.g., Fig. 4b). This illustrates that beneficial mutations may rapidly spread across the landscape without a loss of variation throughout much of the *Synechococcus* genome.

Local Temperature Adaptation Despite High Gene Flow

The study of Rosen et al. [9] does not investigate phenotypic diversity but proposes that a recombination-mediated "cloud" of ecological variation may persist at the population level despite diverse selective pressures simultaneously acting on

Author's personal copy

different regions of the genome. Temperature is a major driver of ecological diversity among anciently divergent, 16S rRNA-defined *Synechococcus* lineages, which specialize on different thermal niches in these ecosystems [16, 25, 26]. We might also expect that ecological variation in thermal performance exists within *Synechococcus* B′ populations, given the breadth of thermal habitat that they occupy [10]. If this is the case, this raises the question of whether locally adapted *Synechococcus* B′ genotypes are sorted and maintained along these gradients, notwithstanding the lack of both dispersal limitations and apparent genetic structure for these organisms (Fig. 1).

According to this scenario, we should observe both that variation in thermal performance exists among strains and that these phenotypic differences are associated with variation in the environmental temperature. To investigate this possibility, we assayed the temperature dependence of growth for 15 randomly selected strains representing the population genetic diversity of our sample (Fig. 1). We observed considerable variation in thermal performance among strains (Fig. 5). This included differences among strains in the minimum and maximum temperatures that support growth, as well as a negative correlation between growth rates at 45 °C and 60 °C ($R = -0.61$; $F_{[1,\ 12]} = 7.1$; $P = 0.02$). Both results are consistent with apparent trade-offs in performance at physiological temperature extremes that have resulted in the ecological specialization of *Synechococcus* B′ strains. There was also a strong match between phenotype and the environment along the geothermal gradients of the two streams: the observed maximum temperature of growth in culture was positively associated with the environmental temperature of the sample from which a strain was isolated ($R = 0.76$; $F_{[1,\ 13]} = 17.6$; $P = 0.001$). This is evidence for local adaptation despite the absence of genetic structure at sequenced loci (Fig. 1). For example, strains R3–14, R6–10, and R8–2 have diverged in thermal niche (Fig. 5) and are locally adapted to their position along the Rabbit Creek thermal gradient, yet they share highly similar genetic backgrounds (Fig. 1).

This local adaptation is similar to that observed for anciently divergent *Prochlorococcus* ecotypes [3]. However, unlike *Prochlorococcus* ecotypes, which rarely recombine, local adaptation of *Synechococcus* B′ persists despite high levels of recombination. We propose that the extensive recombination that we observe could potentially facilitate the spatial structuring of ecological variation by breaking up linkage and creating fine-scale differences in migration-selection balance among genomic regions. As a result, gene flow may become restricted specifically for regions of the genome that are under temperature selection (as well as other environmental variables that potentially co-vary with temperature).

One implication of our results is that the mosaic genomes of *Synechococcus* offer the potential to identify the genetic basis of locally adaptive ecological variation. Because historical recombination events break up linkage between the genes that underlie ecological variation and surrounding regions of the genome, the sequences of ecologically similar genotypes are only expected to cluster near niche-defining targets of selection. Future investigations will take a genome-wide approach to identify these loci as extreme outliers in an analysis of genetic differentiation between samples of phenotypically divergent strains [27]. The *Synechococcus* system therefore promises insights on the longstanding challenge to explicitly map genotype to phenotype at fine-scale (i.e., nucleotide) resolution. The approach will also enable us to determine whether outlier loci are linked to regions of novel, locally adaptive gene content. This is of particular interest in light of genome evolution of the freshwater cyanobacterium *Microcystis aeruginosa*. Strains of *M. aeruginosa* also exhibit mosaic genomes as a consequence of extensive genomic rearrangements and horizontal gene transfer [28], and differences in *Microcystis* gene content among strains appear to be ecologically important [29]. Together, this will enable joint evaluation of the contributions of both allelic variation within the core genome and strain-specific differences in gene content, respectively, to *Synechococcus* local adaptation.

Acknowledgments Field work was conducted under National Park Service research permit YELL-5482. We thank two anonymous reviewers for comments on an earlier version of the manuscript.

Funding Information This work was supported by National Science Foundation award EF-0801999 to SRM.

References

1. Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, Krause DJ, Whitaker RJ (2012) Patterns of gene flow define species of thermophilic Archaea. PLoS Biol. 10:e1001265
2. Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ (2012) Population genomics of early events in the ecological differentiation of bacteria. Science 336:48–51
3. Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, Ding H, Marttinen P, Malmstrom RR, Stocker R, Follows MJ, Stepanauskas R, Chisholm SW (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344:416–420
4. Kraemer SA, Boynton PJ (2017) Evidence for microbial local adaptation in nature. Mol. Ecol. 26:1860–1876
5. Fraser C, Hanage WP, Spratt BG (2007) Recombination and the nature of bacterial speciation. Science 315:476–480
6. Cohan FM, Perry EB (2007) A systematics for discovering the fundamental units of bacterial diversity. Curr Biol 17:R373–R386
7. Wall CA, Koniges GJ, Miller SR (2014) Divergence with gene flow in a population of thermophilic bacteria: a potential role for spatially varying selection. Mol Ecol 23:3371–3383
8. Retchless AC, Lawrence JG (2007) Temporal fragmentation of speciation in bacteria. Science 317:1093–1096
9. Rosen MJ, Davison M, Bhaya D, Fisher DS (2015) Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche. Science 348:1019–1023
10. Miller SR, Strong AL, Jones KL, Ungerer MC (2009) Bar-coded pyrosequencing reveals shared bacterial community properties

Author's personal copy

along the temperature gradients of two alkaline hot springs in Yellowstone National Park. Appl Environ Microbiol 75:4565–4572
11. Haldane JBS (1948) The theory of a cline. J Genet 48:277–284
12. Endler J (1973) Differentiation of populations. Science 165:1228–1232
13. Wu C-I (2001) The genic view of the process of speciation. J Evol Biol 14:851–865
14. Wright S (1943) Isolation by distance. Genetics 28:114–138
15. Castenholz RW (1988) Culturing methods for cyanobacteria. Methods Enzymol 167:68–93
16. Miller SR, Castenholz RW (2000) Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. Appl Environ Microbiol 66:4222–4229
17. Bhaya D, Grossman AR, Steunou A-S, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, Heidelberg JF (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J 1:703–713
18. Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB, YNP Metagenome Project Steering Committee and Working Group Members (2013) The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol 4:67
19. Klatt CG, Inskeep WP, Herrgard MJ, Jay ZJ, Rusch DB, Tringe SG, Parenteau MN, Ward DM, Boomer SM, Bryant DA, Miller SR (2013) Community structure and function of high-temperature chlorophototrophic microbial mats inhabiting diverse geothermal environments. Front Microbiol 4:106
20. Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14(1):68–73
21. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452
22. Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. J Hered 93:153–154
23. Huson DH, Bryant D (2005) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267
24. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
25. Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM (2006) Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. Appl Environ Microbiol 72:544–550
26. Ward DM, Castenholz RW, Miller SR (2012) Cyanobacteria in geothermal habitats. In: Whitton BE (ed) Ecology of cyanobacteria II. Springer, Dordrecht, pp 39–63
27. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet 4:981–994
28. Humbert J-F, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T, Lajus A, Castelli V, Oztas S, Samson G, Longin C, Medigue C, de Marsac NT (2013) A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. PLoS One 8:e70747
29. Meyer KA, Davis TW, Watson SB, Denef VJ, Berry MA, Dick GJ (2017) Genome sequences of lower Great Lakes *Microcystis* sp. reveal strain-specific genes that are present and expressed in western Lake Erie blooms. PLoS One 12:e0183859