Wesleyan University

From the SelectedWorks of Frederick M. Cohan

January 2011

Influence of molecular resolution on sequence-based discovery of ecological diversity among Synechococcus populations in an alkaline siliceous hot spring microbial mat

Contact Author

Start Your Own SelectedWorks

Notify Me of New Work

Available at: http://works.bepress.com/frederick_cohan/4

Applied and Environmental Microbiology, Feb. 2011, p. 1359–1367 Vol. 77, No. 4
0099-2240/11/$12.00 doi:10.1128/AEM.02032-10
Copyright © 2011, American Society for Microbiology. All Rights Reserved.

Influence of Molecular Resolution on Sequence-Based Discovery of Ecological Diversity among *Synechococcus* Populations in an Alkaline Siliceous Hot Spring Microbial Mat▽†

Melanie C. Melendrez,[1]* Rachel K. Lange,[1] Frederick M. Cohan,[2] and David M. Ward[1]

Land Resources and Environmental Science, Montana State University, Bozeman, Montana,[1] and Biology Department, Wesleyan University, Middletown, Connecticut[2]

Received 28 August 2010/Accepted 11 December 2010

Previous research has shown that sequences of 16S rRNA genes and 16S-23S rRNA internal transcribed spacer regions may not have enough genetic resolution to define all ecologically distinct *Synechococcus* populations (ecotypes) inhabiting alkaline, siliceous hot spring microbial mats. To achieve higher molecular resolution, we studied sequence variation in three protein-encoding loci sampled by PCR from 60°C and 65°C sites in the Mushroom Spring mat (Yellowstone National Park, WY). Sequences were analyzed using the ecotype simulation (ES) and AdaptML algorithms to identify putative ecotypes. Between 4 and 14 times more putative ecotypes were predicted from variation in protein-encoding locus sequences than from variation in 16S rRNA and 16S-23S rRNA internal transcribed spacer sequences. The number of putative ecotypes predicted depended on the number of sequences sampled and the molecular resolution of the locus. Chao estimates of diversity indicated that few rare ecotypes were missed. Many ecotypes hypothesized by sequence analyses were different in their habitat specificities, suggesting different adaptations to temperature or other parameters that vary along the flow channel.

Molecular analyses have revealed ecological diversity among the dominant *Synechococcus* populations native to microbial mats in alkaline siliceous hot springs, such as Octopus Spring and Mushroom Spring, Yellowstone National Park, Wyoming. The dominant native *Synechococcus* populations, referred to as A/B-type *Synechococcus*, are distributed uniquely along the flow path and, thus, the temperature gradient in the effluent channels (17). Specifically, *Synechococcus* 16S rRNA genotypes A″, A′, A, B′, and B are found at progressively lower temperatures from the upper temperature limit of these mats near 72°C to ~50°C (17, 53). The habitat specificity of 16S rRNA sequences has suggested the existence of populations with distinct ecologies (ecotypes), with adaptations to different temperatures, as suggested previously for *Synechococcus* isolates from an Oregon hot spring mat (30, 34). Allewalt et al. (1) succeeded in cultivating *Synechococcus* with 16S rRNA sequences corresponding to A-, B′-, and B-type *Synechococcus* genotypes, which are found at 56°C to 68°C, 53°C to 60°C, and 53°C to 55°C regions, respectively (17). Allewalt et al. (1) also confirmed that these isolates exhibited temperature adaptations corresponding to the *in situ* distributions of these genotypes. Ward (55) pointed out that these *Synechococcus* ecotypes could be considered species using an ecological species concept, wherein the individual organisms comprising a species share the same environmental niche or adaptive zone (47).

The observed correspondence between 16S rRNA genotypes and ecotypes within *Synechococcus* (and within other taxa) is consistent with expectations from population genetics theory. One model of bacterial speciation (the stable ecotype model) predicts that even when the genes surveyed are not relevant to ecological divergence, closely related ecotypes may be identifiable as sequence clusters (9, 32, 54). In the stable ecotype model, where the diversity within each ecotype is purged frequently by periodic selection (PS) at every gene in the genome and ecotype formation (EF) occurs only infrequently, each ecotype accumulates a unique set of sequence mutations (usually of no fitness effect) over the ecotype's long history as a distinct lineage. Each ecotype contains individuals that not only have a unique ecology but also are ecologically interchangeable (i.e., any individual is similar enough ecologically to extend the ecotype lineage if it is selected by PS). This dynamic potentially allows each ecotype to be recognizable as a sequence cluster for every gene in the genome. Since sequence clusters might alternatively be the result of neutral evolution (e.g., genetic drift and physical isolation), it is important to test the ecological distinctions of putative ecotypes (PEs) identified as sequence clusters.

Indeed, microbial ecologists studying a great diversity of prokaryotes have discovered closely related, ecologically distinct populations, often subsumed within a named "species" recognized by microbial systematics, through phylogenetic analysis of sequence diversity. These analyses have revealed extremely closely related ecotypes in free-living prokaryotes, such as soil *Bacillus* (12, 25), marine *Vibrio* (22), and acid mine *Leptospirillum* (13) and *Procholorococcus* (24, 29), as well as among pathogens, including *Legionella* (10), *Escherichia* (28, 50, 51), *Mycobacterium* (42), *Salmonella* (43), and *Yersinia* (58). In all these cases, clusters identified by patterns of sequence

* Corresponding author. Present address: Armed Forces Research Institute of Medical Science (AFRIMS), 315/6 Rajvithi Road, Bangkok 10400, Thailand. Phone: 66 02 696 2700, ext. 4328. Fax: 66 02 644 4760. E-mail: melendrezmc@afrims.org.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 17 December 2010.

diversity (in genes not related to ecological divergence) were confirmed to be ecologically distinct.

One significant challenge to sequence-based identification of ecotypes is to utilize a set of genes with enough resolution to discover the most newly divergent ecotypes. For example, the slowly evolving 16S rRNA molecule had too little resolution to reveal ecotypes within a clade of *Bacillus simplex*: a clade that was invariant in 16S rRNA sequence was demonstrated, through analysis of more rapidly evolving protein-coding genes, to contain nine ecotypes that were distinct in adaptations to solar exposure (25, 41). Here we explore whether previous analyses have adequately identified newly divergent ecotypes within the A/B clade of *Synechococcus*.

The need for higher molecular resolution emerged from studies of depth distribution through the ~1-mm-thick photic zone (36, 53). At different temperatures along the thermal gradient, *Synechococcus* cells inhabiting the mat surface autofluoresce less intensely than those at greater depths, where light intensity is much lower and light quality differs dramatically (53). At 60°C, 16S rRNA genotype B′ was the only *Synechococcus* genotype detected in the surface layer, whereas 16S rRNA genotype A was detected only at depths corresponding to the deeper *Synechococcus* population (36). At 68°C only the A′ *Synechococcus* 16S rRNA genotype occurs at the surface and subsurface, raising the question of whether these differently fluorescing A′ populations correspond to one differently acclimated (physiologically optimized) population or two differently adapted (evolutionarily optimized) populations that are too closely related to be differentiated using 16S rRNA sequence variation. Using the more rapidly evolving 16S-23S rRNA internal transcribed spacer (ITS) region, it was observed that the A′ surface and subsurface populations at 68°C were genetically distinct (i.e., different ecotypes) (16). At some temperatures between 60°C and 68°C, only the A genotype occurs, but no depth variation in either 16S rRNA or ITS sequences was observed (53). This suggested the need for yet higher molecular resolution to distinguish acclimated from adapted populations and thereby to ensure detection of all the ecologically distinct *Synechococcus* populations in the mat.

Another significant challenge to sequence-based discovery of ecotypes is in identifying the size of phylogenetic clusters that most closely represent the ecologically homogeneous ecotypes. To this end, we have utilized two newly developed computer algorithms for discovering the clades representing ecotypes. We note that one of these methods, ecotype simulation (ES), bases its predictions of ecotype demarcation solely on sequence diversity as input; the ecotypes so demarcated must then be independently confirmed to be ecologically distinct. In our approach, we aim to confirm the ecotype demarcations of ES through comparing PEs for their habitat specificities. The other method, AdaptML, incorporates both sequence diversity and habitat data and simultaneously hypothesizes ecotypes and confirms their ecological distinctness on the basis of habitat associations. Each approach holds unique advantages (8).

Here we have hypothesized that there are more *Synechococcus* ecotypes in the mat than the number revealed by 16S rRNA/ITS variation alone. Previous single- and multilocus studies suggested that protein-encoding loci are less conserved than the 16S rRNA/ITS region and may yield more highly

TABLE 1. Characteristics and primer sequences for protein-encoding loci

Locus	Primer	Sequence (5′–3′)	Size of sequence analyzed (bp)		% divergence of *Synechococcus* A and B′ homologs	No. of A/B′ metagenomic sequences analyzed	% divergence (variance) of metagenomic homologs:		d_N/d_S for metagenomic homologs:	
			Synechococcus A	*Synechococcus* B′			A-like *Synechococcus*	B′-like *Synechococcus*	*Synechococcus* A	*Synechococcus* B′
apcAB	apcABF[a,b]	TTAACTTCTCCATGGCCGAGCTGT	423[c]	425[d]	8.5	7/11	0.04 (0.1)	1.84 (0.6)	0.01	0.1
	apcABR[a,b]	AAGCTGCCGATCAACTGTTCCAGA								
aroA	aroA68F[a]	TGAGCATTCCAGGCGATAAGTCCA	218	696	15.1	6/10	0.3 (0.05)	3.47 (0.1)	0.08	0.9
	aroA809R[a]	AGACCCACATCCTGTAGCAGCAAT								
	aroA77F[b]	CCGGCGACAAATCCATTTCCCACC								
	aroA809R[b]	AGGCCCACATCCCTTAGCAGCAAT								
rbsK	rbsK34F[a]	AATCTGGACTTGGTGGTGCA GGTA	597	627	18.2	3/14	—[e]	3.06 (3.98)	0.82[f]	0.11
	rbsK645R[a]	GATGATCACTGTCGGGATCCCCTG								
	rbsK34F[b]	AACATGGATTTGGTGGTGCA GGTG								
	rbsK730R[b]	TGGTATCCACCACTTCCACAGGAT								

[a] Primers used for amplification of A-like *Synechococcus*.
[b] Primers used for amplification of B′-like *Synechococcus*.
[c] Includes 60 bp of noncoding sequence between *apcA* and *apcB* genes.
[d] Inlcudes 63 bp of noncoding sequence between *apcA* and *apcB* genes.
[e] —, not enough metagenomic sequence data to calculate average divergence and variance.
[f] On the basis of analysis of three metagenomic *rbsK* sequences.

TABLE 2. PCR clone library compositions

Gene	Sample	No. of clones in library	%	
			A-like *Synechococcus*	B′-like *Synechococcus*
apcAB	M60	309	34	66
	M65	117	100	0
aroA	M60	423	60	40
	M65	107	100	0
rbsK	M60	257	35	65
	M65	67	100	0

resolved phylogenies for discerning closely related populations of organisms (21, 25, 33, 57). In this study, we use ES and AdaptML to analyze diversity at three protein-encoding loci of various molecular resolutions to investigate the influence of molecular resolution on prediction of ecotypes inhabiting two temperature-defined sites in this mat. The loci were *rbsK* (encoding ribokinase), *aroA* (encoding 3-phosphoskimate 1-carboxyvinyltransferase), and *apcAB* (encoding the allophycocyanin alpha and beta subunits).

MATERIALS AND METHODS

Study sites and sample collection. We studied two temperature-defined sites, 60°C and 65°C, in the effluent channel of Mushroom Spring (44.5386°N, 110.7979°W) (termed M60 and M65, respectively), an alkaline siliceous hot spring in the Lower Geyser Basin, Yellowstone National Park, Wyoming (36). This allowed us to exploit genomes for *Synechococcus* sp. strains A (strain JA-3-3Ab, referred to as the A genome) and B′ [strain JA-2-3B a(2-13), referred to as the B′ genome], both isolated from Octopus Spring (1), and metagenomes that were prepared from the same 60°C and 65°C environments within Mushroom Spring (3; C. G. Klatt, J. M. Wood, D. B. Rusch, M. M. Bateson, N. Hamamura, J. F. Heidelberg, A. R. Grossman, D. Bhaya, F. M. Cohan, M. Kühl, D. A. Bryant, and D. M. Ward, unpublished data). Mat core samples collected with a no. 4 cork borer on 2 October 2003 were transported to the laboratory on dry ice and kept frozen at −80°C until further analysis, as previously described (36).

DNA extraction and purification. The top green layers (~1 mm thick) of three samples were removed using a clean razor blade, the samples were combined, and cells were lysed using the bead-beating and phenol-chloroform extraction protocol previously used to obtain DNA from *Synechococcus* strain A and B′ genomes, mat metagenomes, and 16S rRNA/ITS sequences (3, 17, 18, 53; Klatt et al., unpublished). DNA concentration was estimated using agarose gel electrophoresis analysis with high-molecular-weight DNA ladders (Invitrogen).

Locus selection. This study was linked to a cultivation-independent, multilocus sequence analysis (MLSA) study based on bacterial artificial chromosome cloning of 100- to 130-kb genome segments containing 16S rRNA genes and surrounding protein-encoding loci from A/B-lineage *Synechococcus* spp. (M. C. Melendrez, J. M. Wood, D. B. Rusch, and D. M. Ward, unpublished data; M. C. Melendrez, F. M. Cohan, and D. M. Ward, unpublished data). Over a range of 100 genes in both directions from the two 16S rRNA loci in the *Synechococcus* sp. strain A and B′ genomes and metagenomic homologs of these genes, we searched for protein-encoding genes with the following characteristics: (i) presence in both genomes, (ii) proximity to but not immediately adjacent to the 16S rRNA locus, (iii) high degree of nucleotide divergence between *Synechococcus* strain A and B′ homologs, (iv) high average nucleotide divergence and variance of metagenomic homologs from *Synechococcus* strain A and B′ homologs, (v) not under positive evolutionary selection, (vi) not flanked by highly mobile elements (to decrease the probability that the locus had recently undergone transfer from some other organism and might exhibit a different phylogeny than other genes in the organisms containing them), and (vii) functionally useful (e.g., may be highly expressed under different light conditions). Nucleotide divergence between *Synechococcus* strain A and B′ homologs or between metagenomic sequences and the homolog in the more closely related of the two reference genomes was determined by finding the reciprocal best BLAST matches in the National Center for Biotechnology Information nucleotide collection (NCBI, BLAST-nr/nt) sequence database (using default values) (2). The degree of evolutionary selection pressure was assayed by computing the ratios of nonsynonymous to synonymous substitutions (dN/dS values) using the method of Nei and Gojobori (31) implemented in the DnaSP program package (27). Genes with dN/dS values greater than 1, indicative of positive selection, were not considered.

PCR amplification, cloning, and sequencing. Primers for the separate amplification of *Synechococcus* A-like and B′-like population target genes (Table 1) were obtained from Integrated DNA Technologies. The primers were designed using the primer design tool in SciTools on the Integrated DNA Technologies website and analyzed by BLAST-nr/nt to ensure specificity to *Synechococcus* A-like or B′-like genomic homologs. All primers were dissolved in water to a final concentration of 50 μM for use in PCR amplification, under the following conditions: initial denaturing step of 94°C for 2 min (10 min for *apcAB*), followed by 30 cycles (33 cycles for *apcAB*) of 94°C (1 min), 55°C (1 min) (60°C [1 min, −0.5°C/cycle] for *apcAB*), and 72°C (1 min), with a final extension at 72°C for 10 min, followed by storage at 4°C. After sizes were verified using agarose gel electrophoresis, PCR products were purified using MiniElute 96 UF PCR or QIAquick PCR purification kits (Qiagen), according to the manufacturer's instructions. Purified PCR products were cloned using a TOPO-TA cloning kit for sequencing (Invitrogen), according to the manufacturer's instructions. Ligation products were used to transform OneShot Mach-T1 chemically competent *Esch-*

TABLE 3. Comparison of Chao estimates of OTU diversity and ecotype diversity predicted from ecotype simulation analysis of ~70 variants and 3 loci of *Synechococcus* A-like and B′-like populations

Sample	*Synechococcus* type	Gene	No. of variants	AED (%)	Chao richness estimates			Ecotype simulation			
					Average observed no. of taxa[a]	$S_{c\text{-pe}}$[b]	$S_{c\text{-nt}}$ (95% CI)	Prediction of *n* parameter (95% CI)	No. of demarcated PEs[c]	Omega (95% CI)	Sigma (95% CI)
M60 + M65	A	16S rRNA	70	0.002	1	1	1 (1–1)	58 (2–71)	1	5.42 (5–18.8)	3.39 (0.05–>100)
		ITS	70	0.002	3	2	3 (0–3)	3 (2–10)	2	0.136 (0.01–0.45)	4.04 (0.58–20.1)
		apcAB	73	0.016	10	5.5	7.8 (7–30)	11 (3–11)	5	0.035 (0.01–0.12)	1.82 (0.01–>100)
		aroA	73	0.013	3	4	3.5 (0–3)	5 (2–30)	4	0.373 (0.17–1.2)	143 (0.64–>100)
		rbsK	72	0.095	13	14.3	13 (12–23)	15 (8–30)	14	0.043 (0.02–0.09)	3.76 (0.25–>100)
M60	B′	16S rRNA	73	0.002	2	1	2 (2–2)	7 (2–75)	1	0.519 (0.01–13.2)	5.48 (0.25–>100)
		ITS	70	0.01	6	2	7 (0–7)	51 (3–70)	2	1.00 (0.68–3)	1.60 (0.03–>100)
		apcAB	67	0.01	5	11	7 (10–35)	32 (9–70)	8	0.314 (0.1–0.68)	1.09 (0.02–>100)
		aroA	70	0.03	13	15	15 (11–23)	32 (3–70)	13	0.242 (0.11–0.81)	23.8 (0.01–>100)
		rbsK	68	0.05	15	14.5	15 (14–25)	20 (11–46)	13	0.092 (0.04–0.14)	1.8 (0.12–>100)

[a] At 99% cutoff.
[b] Data do not include 95% confidence intervals; discrete values only are provided.
[c] No 95% confidence interval for demarcated PE value; see *n* parameter confidence interval values.

erichia coli cells (Invitrogen), which were spread on LB-Amp (LB medium and 0.1 μg/μl ampicillin) plates (Fisher Scientific) and incubated overnight. Transformant colonies were picked, and the sizes of the inserts were analyzed using PCR amplification with the M13F/R primers (primer M13F, 5′-GTAAAACGACGGCCAG-3′; primer M13R, 5′-CAGGAAACAGCTATGAC-3′) or T3/T7 primers (primer T3, 5′-AATTAACCCTCACTAAAGGG-3′; primer T7, 5′-TAATACGACTCACTATAGGG-3′) specific to the cloning vector and agarose gel electrophoresis. Clones with correctly sized inserts were sequenced in both directions using the M13F/R or T3/T7 primer and the BigDye (version 3.1) cycle sequencing kit (Applied Biosystems) at the University of Nevada—Reno Sequence Center (Reno, NV).

Sequence alignment and phylogenetic analysis. Sequence data were analyzed using the Sequencher (version 4.8) program and alignments were made using ClustalX software with *Synechococcus* strain A and B′ genomic homologs as references. Sequences were specified to belong to *Synechococcus* sp. A-like or B′-like populations on the basis of a reciprocal best BLAST match to one of the reference genomes using the BLAST-nr/nt database and default parameters. Two suspected *Synechococcus* sp. A′-like sequences identified in the 65°C sample (see the information in the supplemental material) were removed. Alignments were manually verified and neighbor-joining trees were constructed using MEGA (version 4.0) software (46) and viewed using NJPlot (35). Estimates of average evolutionary divergence (AED) were computed using the overall mean distance function over all sequence pairs in a clade using the maximum-composite-likelihood method in MEGA. All nucleotide sites were included in the analysis, but positions containing gaps (≤11) and missing data were eliminated from the data set (complete deletion option) for consistency of sequence data input across all analyses.

Ecotype simulation and demarcation. Sequence alignments were analyzed using ES to predict the number of PEs (*n* parameter), rates of PS (sigma) and EF (omega), and 95% confidence intervals (CIs) for all parameters (http://fcohan.web.wesleyan.edu/ecosim/), using the option of 1.5× precision match between observed and simulated data. Neighbor-joining trees were constructed and uploaded into ES in Newick format for PE demarcation using the ES manual demarcation program. Sequences from the largest clades were analyzed, and if the 95% CI included one PE, the predicted PE was considered a single PE even if the most likely number of PEs was greater than one. If ES predicted the clade to contain more than one PE (i.e., 1 was not included in the 95% CI), sequences were removed systematically starting from the root of the entire phylogeny until clades in which the 95% CI included one PE were identified.

AdaptML. AdaptML was used as directed with default parameter values. Inputs included the habitat of isolation as well as the phylogeny of organisms, based on sequence data. The algorithm yielded ecotype demarcations and the habitat specificity of each (http://almlab.mit.edu/ALM/Research/Entries/2008/5/10_Adaptation_in_ Metagenomes.html) (22).

Chao estimation of diversity. Collector's curves and Chao estimates (S_c) of the lower bound of operational taxonomic unit (OTU) richness were generated for a normalized number of sequence variants (see Results for rationale) (5, 6, 11), using the software package Mothur (version 1.4.0; http://schloss.micro.umass.edu/mothur/) (38–40). Two S_c estimates were obtained. The first, $S_{c\text{-nt}}$, was based on aligned nucleotide sequence data, which were grouped into OTUs on the basis of ≥99% sequence identity. The second, $S_{c\text{-pe}}$, was based on the PEs identified from ES analysis. This provides an estimate of PE richness rather than OTU richness.

Nucleotide sequence accession numbers. Sequences of protein-encoding loci have been submitted to GenBank (accession numbers HQ187783 to HQ188198). Sequence data for 16S rRNA and ITS previously obtained from the same samples and additional samples by PCR amplification (53) are available from GenBank under accession numbers DQ979027 to DQ979319 (collected 2 October 2003 from the Mushroom Spring 60°C and 65°C sites).

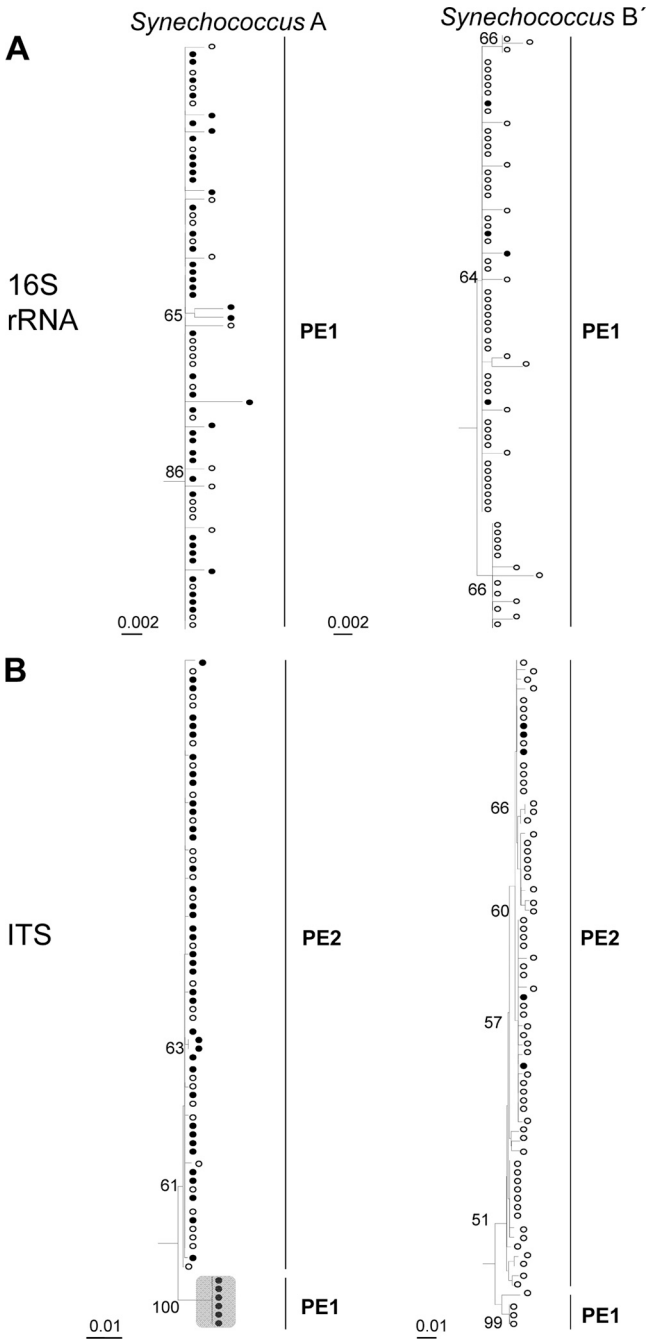

FIG. 1. Neighbor-joining phylogenetic trees of A-like and B′-like *Synechococcus* diversity in Mushroom Spring at 60°C (open circles) and 65°C (closed circles) for the 16S rRNA gene (A) and the ITS region (B), with putative ecotypes (PE*n*) demarcated by ecotype simulation. Replicate sequences of the dominant variant within a demarcated putative ecotype are highlighted in the shaded box for *Synechococcus* sp. A-like ITS. Bootstrap values of >50% are shown. Bars represent the number of fixed point mutations per sequence position.

RESULTS

Locus selection and characteristics. Two housekeeping genes, *rbsK* and *aroA*, and a pair of adjacent genes important in photosynthesis, *apcAB* (separated by a noncoding spacer), were selected as loci satisfying the criteria outlined above. All exist as one copy in the genomes of *Synechococcus* strains A and B′; they occur within 70 kb of the same 16S rRNA gene locus, and they were not directly adjacent to open reading frames annotated as transposons, resolvases, recombinases, and mobile elements (see Fig. S1 in the supplemental material). As reported in Table 1, all loci are under purifying selection. The nucleotide divergence levels between homologs in the *Synechococcus* sp. A and B′ genomes were 8.5%, 15.1%, and 18.2% for *apcAB*, *aroA*, and *rbsK*, respectively. Across the three loci, average divergences of metagenomic sequences

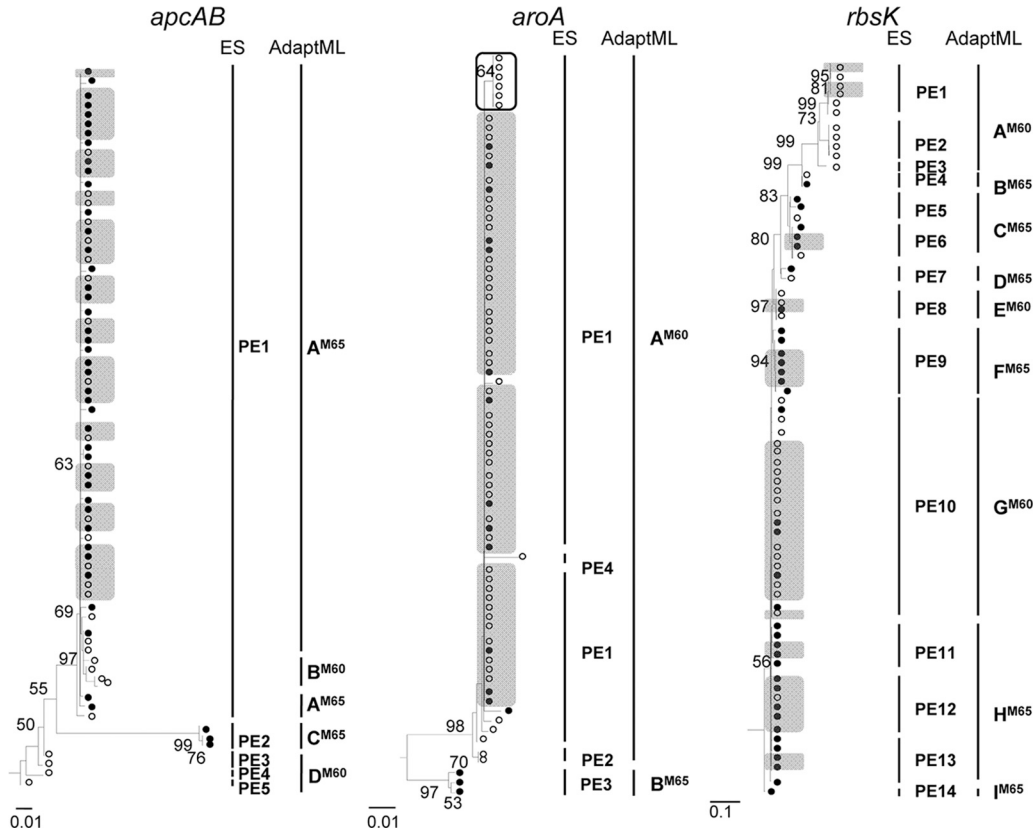

FIG. 2. Neighbor-joining phylogenies for *Synechococcus* A-like population sequences for *apcAB*, *aroA*, and *rbsK* genes, with putative ecotypes demarcated by ecotype simulation and habitat-specific clusters demarcated by AdaptML. Open and closed circles demarcate sequences from 60°C and 65°C samples, respectively. Replicate sequences of the dominant variant within a demarcated putative ecotype are highlighted in shaded boxes. Putative subdominant variants in *aroA* PE1 are highlighted in the unshaded box. Superscripts M60 and M65 next to AdaptML cluster demarcation indicate habitat specificity. Bootstrap values of >50% are shown. Bars represent the number of fixed point mutations per sequence position.

from homologs in the more closely related reference genome ranged from 0.04% to 0.3% (variance, 0.05 to 0.1) for A-like homologs and 1.8% to 3.5% (variance, 0.1 to 3.98) for B′-like homologs, and the degree of divergence among loci followed the same order observed with *Synechococcus* sp. strain A and B′ homologs. Given that this analysis was based on only 3 to 14 metagenomic sequences, these were considered rough estimates.

Clone library composition. PCR clone libraries of the various genes contained 67 to 423 sequenced clones (Table 2). *Synechococcus* strain B′-like sequences were recovered only from the M60 habitat, whereas *Synechococcus* strain A-like sequences were found in both M60 and M65 habitats for all loci, as expected from previous analyses of 16S rRNA and ITS sequence variation (53). The relative percentages of A-like and B′-like *apcAB* and *rbsK* sequences were also consistent with previous results (17, 53); A-like *aroA* sequences were unexpectedly dominant in the M60 sample, probably due to PCR primer bias.

Effect of sampling and molecular resolution on detection of diversity. Collector's curve analyses (see Fig. S3 in the supplemental material) revealed how sampling affected measurements of genetic diversity (OTUs). Normalization for number of sequences was necessary for comparative analyses of different loci in order to exclude the possibility that large differences in sampling might influence the outcome. As a result, we restricted analyses of each lineage to 67 to 73 sequences, which allowed comparison of nearly equal numbers of B′-like and A-like variants. As shown in Fig. S3 in the supplemental material, this should not have had a major impact on our estimation of diversity in the sample. Richness estimates ($S_{c\text{-nt}}$ values) across loci ranged from 1 to 13 OTUs for A-like populations and from 2 to 15 OTUs for B′-like populations and were roughly proportional to the depth of divergence, as estimated by AED (Table 3).

Differences between A-like and B′-like lineages. The loci that we studied appear to have undergone different evolutionary histories in the two lineages (Table 3). For instance, the AED for *aroA* was higher than that for *apcAB* in the B′-like population, consistent with the degree of divergence in these genes between *Synechococcus* sp. strains A and B′ (Table 1), but the reverse was true in the A-like population. Also, a greater divergence was found for ITS and *aroA* in the B′-like lineage, which shows a greater depth of divergence across the genome (3), whereas greater depths of divergence were found for *apcAB* and *rbsK* in the A-like population. Since we could not discern A′-like from A-like *rbsK* sequences (see the information in the supplemental material), any A′-like sequences present in the M65 sample may have artificially increased the AED value for this locus.

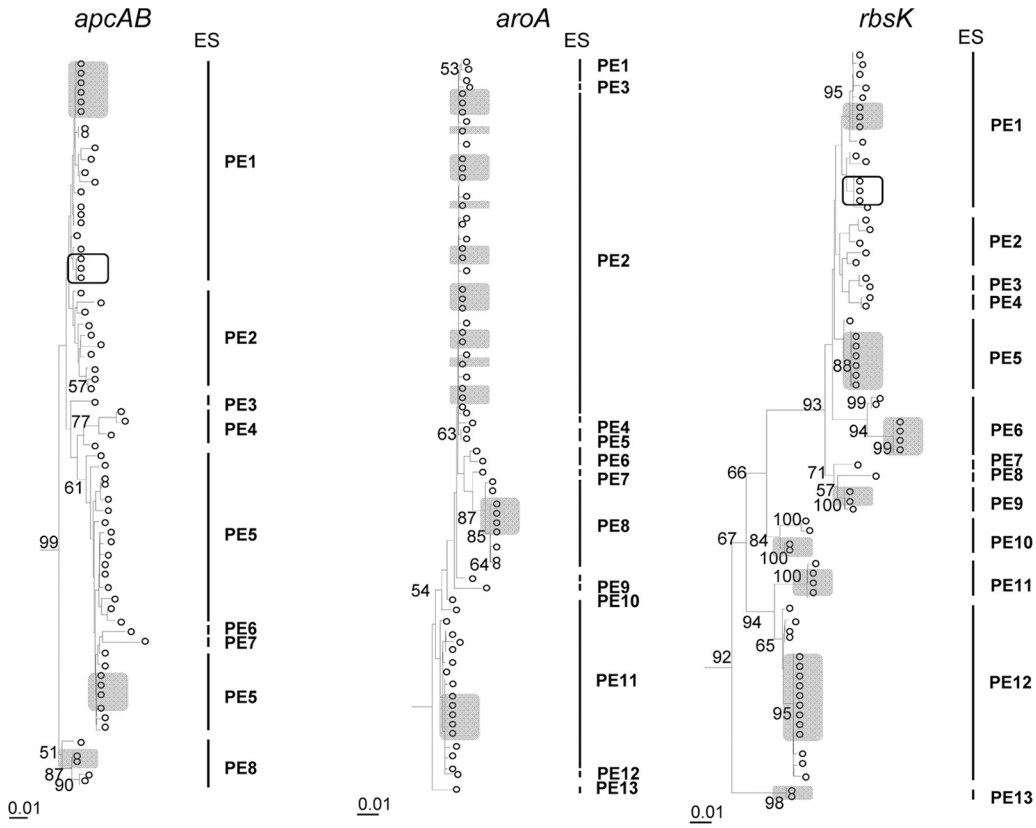

FIG. 3. Neighbor-joining phylogenies for *Synechococcus* B′-like population sequences for *apcAB*, *aroA*, and *rbsK* genes, with putative ecotypes demarcated by ecotype simulation. Replicate sequences of the dominant variant are highlighted in shaded boxes for each demarcated ecotype. Putative subdominant variants of *apcAB* PE1 and *rbsK* PE1 are highlighted in unshaded boxes. Bootstrap values of >50% are shown. Bars represent the number of fixed point mutations per sequence position.

Putative ecotype demarcations across loci and lineages. ES analyses were conducted on the same subsets of sequence data used to make AED calculations (Table 3 and Fig. 1 to 3). In general, well-sampled PEs contained a dominant sequence variant (shaded boxes in Fig. 1 to 3), which we interpret to be the survivor of the most recent periodic selection event in the PE clade; we interpreted minor sequence variants to be recent mutants of the survivor. The clade that we have labeled PE1 in A-like *Synechococcus aroA* analysis (Fig. 2), as well as B′-like *Synechococcus apcAB* PE1 and *rbsK* PE1 (Fig. 3), each contained two dominant allelic variants (the second denoted by an unshaded box in Fig. 2 and 3). The number of manually demarcated PEs clustered toward the lower end of the 95% CIs for the *n* parameter predicted by ES (Table 3). In general, predicted PS rates (sigma) were higher than predicted EF rates (omega), suggesting that EF events are relatively rare and that PEs are recurrently constrained by PS, consistent with the stable ecotype model. The average sequence divergence among variants within each PE varied from 0% (all identical) to 1.9%, and the maximum nucleotide variation within a PE clade ranged from 0.2% to 7.8% (Table 4).

ES predicted 1 to 2 demarcated PEs from 16S rRNA and ITS sequence variation (Fig. 1) for *Synechococcus* A-like and B′-like sequences, similar to the results of Ward et al. (53). For the *Synechococcus* A-like population, ES demarcated 5, 4, and 14 PEs from *apcAB*, *aroA*, and *rbsK* sequence variations, respectively (Fig. 2). For the *Synechococcus* B′-like population, ES demarcated 8, 13, and 13 PEs from *apcAB*, *aroA*, and *rbsK* sequence variations, respectively (Fig. 3). PE predictions were a function of the AED, which was taken as a measure of the molecular resolution of the locus (Fig. 4).

Habitat associations of A-like putative ecotypes. Many A-like PEs contained members that were found in only one of the two temperature-defined habitats, suggesting that they constitute ecologically distinct populations (Fig. 2). A Fisher's exact test indicated significant heterogeneity across ecotypes with respect to their habitat associations (Table 5). The two most highly resolving loci, *aroA* and *rbsK*, showed significant habitat associations, while *apcAB* showed a marginally significant association. In analysis of each gene, AdaptML identified two classes of ecotypes that were significantly different in their habitat specialization: one class of ecotypes specialized on M60 and one specialized on M65 (Table 5). On average (across the three genes), the M60-specialized ecotypes were predicted to occur in the M60 habitat 85% of the time and in the M65 habitat 15% of the time; the M65-specialized ecotypes were predicted to occur in the M60 habitat 23% of the time and in the M65 habitat 77% of the time. In most cases, AdaptML identified the same ecotypes as ES (Table 5). In all but one case of noncongruence between algorithms, ES split an AdaptML ecotype into multiple PEs. In one case, AdaptML

TABLE 4. Percent sequence variation within PE clades that contain two or more sequences

Organism	Locus	PE	No. of sequences in PE	% sequence variation	
				Avg	Maximum
A-like *Synechococcus*	*apcAB*	1	66	0.4	2.9
		2	3	0.2	1.0
		3	2	1.0	1.0
	aroA	1	68	0.5	7.8
		2	2	0	0
		3	3	0.7	1.0
	rbsK	1	6	0.4	0.9
		2	4	0	0
		4	2	0.2	0.2
		5	3	1.1	1.6
		6	4	0.2	0.4
		7	2	0.7	2.7
		8	4	0.2	0.4
		9	7	0.7	1.7
		10	24	0.1	0.4
		11	4	0.2	0.4
		12	7	0.1	0.2
		13	5	0.5	0.7
B′-like *Synechococcus*	*apcAB*	1	22	1.1	2.9
		2	9	1.9	4.0
		4	3	1.0	3.1
		5	25	1.7	3.8
		8	5	1.3	2.6
	aroA	1	3	1.1	1.2
		2	11	0.5	1.2
		5	2	0.2	0.2
		6	2	0.8	0.8
		8	9	0.5	1.2
		11	17	0.8	2.2
	rbsK	1	15	0.9	1.8
		2	5	1.3	1.6
		3	2	0.2	0.2
		4	2	0.2	0.2
		5	7	0.3	0.5
		6	6	0.9	1.8
		9	3	0.2	0.2
		10	4	1.1	1.6
		11	4	0.1	0.2
		12	16	0.4	1.0
		13	2	0	0

split a PE demarcated by ES (*apcAB* PE1) into two ecotypes (Fig. 2 and Table 5).

DISCUSSION

The protein-encoding loci that we studied offered higher molecular resolution than 16S rRNA and ITS sequences for prediction of the genetic and ecological diversity of native *Synechococcus* populations inhabiting these mats. Between 4 and 14 times more PEs were predicted from the *rbsK* locus, which offers the greatest molecular resolution. Compared to predictions based on 16S rRNA and ITS sequence variation, ES-demarcated PEs for A-like *Synechococcus* populations are significantly heterogeneous in their associations with the 65°C and 60°C collection sites (Fig. 2 and Table 5), suggesting that they are, in fact, ecologically distinct populations. AdaptML results were consistent, with AdaptML mainly identifying the same PEs as ES. We previously found a tendency of ES to split ecotypes identified by AdaptML (12), possibly because AdaptML is able to identify only those ecotypes that are distinguishable in the environmental parameters provided by the investigator, whereas the vision of ES is not restricted by the environments hypothesized by the investigators (8, 12).

Since we studied samples from only two temperature-defined sites along the effluent channel, further work will be required to understand the adaptations that might explain why PEs differ in their temperature distributions and why more than one PE is present at the same temperature site. More detailed studies of the distribution of PEs predicted from protein-encoding sequence variation provide evidence of fine-scale distribution differences of PEs along flow path and vertical dimensions, as well as PE-specific gene expression patterns and PE-specific population dynamic responses to environmental change (E. D. Becraft, F. M. Cohan, M. Kühl, S. Jensen, D. Rusch, and D. M. Ward, unpublished data). It is important to note that to observe the habitat associations of PE clades, it is necessary to have sufficient molecular resolution to prevent the collapse of ecologically distinct clades into larger clades that appear to contain members that are ecologically heterogeneous. Overly conservative demarcation could also cause this effect. The existence of two dominant variants within some PE clades could foretell additional ecological diversity.

There have been many suggestions of molecular cutoffs for the demarcation of species-like clusters. Originally, 70% DNA-DNA hybridization based on phenotypic clustering of bacterial strains within named species was suggested (23, 56), which was roughly correlated with 97 to 98% 16S rRNA gene sequence identity (20, 45) (this was later revised to 98.7 to 99% [44]). With the advent of genomic analysis, ~94% average nucleotide identity (ANI) of genomes was suggested (26) and used to demarcate species (48); Konstantinidis and Tiedje (26) suggested that organisms that share a particular ecology and >99% ANI may be the biologically meaningful units in nature.

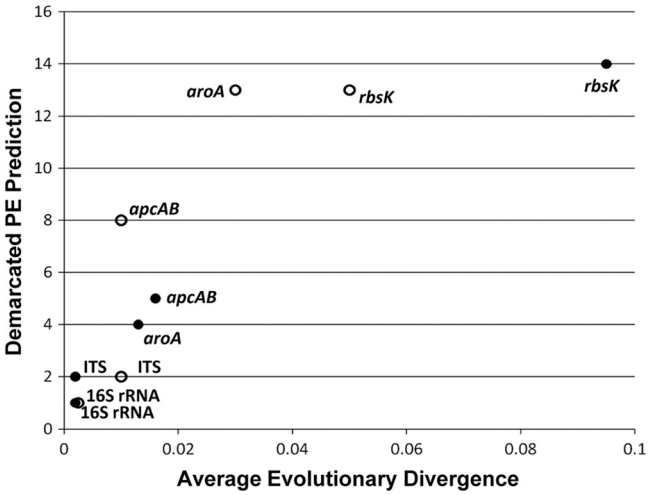

FIG. 4. Demarcated putative ecotype prediction of ecotype simulation as a function of molecular resolution (i.e., average evolutionary divergence) for *Synechococcus* A-like (closed circles) and B′-like (open circles) variants used in ecotype simulation analyses, as shown in Fig. 1 to 3.

TABLE 5. Habitat distributions of putative ecotypes demarcated by ecotype simulation and AdaptML

Gene	PE		No. of sequences from:		Habitat specificity[e]	*P* value[f]
	Ecotype simulation	AdaptML	M60	M65		
apcAB[b]	1α[g]	A	20	35	M65	0.098
	1β[g]	B[a]	4	0	M60	
	2	C	0	3	M65	
	3	D	2	0	M60	
	4		1	0	M60	
	5		1	0	M60	
aroA[c]	1	A	52	12	M60	0.015
	2		2	0	M60	
	4		1	0	M60	
	3	B	0	3	M65	
rbsK[d]	1	A	6	0	M60	<0.001
	2		4	0	M60	
	3		1	0	M60	
	4	B	1	1	M65	
	5	C	1	2	M65	
	6		1	3	M65	
	7	D	1	1	M65	
	8	E	3	1	M60	
	9	F	0	7	M65	
	10	G	17	5	M60	
	11	H	0	5	M65	
	12		1	5	M65	
	13		0	5	M65	
	14	I	0	1	M65	

[a] AdaptML ecotype B of *apcAB* contained the four sequences M60c639B6, M60c504C8, M60c096H4, and M60c494H7 from ecotype simulation PE1.

[b] For *apcAB*, ecotypes with an M60 habitat spectrum were estimated by AdaptML to occur 93% of the time at 60°C and 7% of the time at 65°C; ecotypes with an M65 habitat spectrum were estimated to occur 33% of the time at 60°C and 67% of the time at 65°C.

[c] For *aroA*, ecotypes with an M60 habitat spectrum were estimated to occur 79% of the time at 60°C and 21% of the time at 65°C; ecotypes with an M65 habitat spectrum were estimated to occur 22% of the time at 60°C and 78% of the time at 65°C.

[d] For *rbsK*, ecotypes with an M60 habitat spectrum were estimated to occur 83% of the time at 60°C and 17% of the time at 65°C; ecotypes with an M65 habitat spectrum were estimated to occur 13% of the time at 60°C and 87% of the time at 65°C.

[e] Indicated by AdaptML.

[f] Fisher's exact test was employed as a contingency test for differences in habitat associations among ecotype simulation's putative ecotypes (only PE clades with >1 sequence were included).

[g] Part of 1 PE.

Although our Chao estimates of diversity using a 99% sequence identity cutoff to define OTUs were similar to our demarcated PE estimates (Table 3), we do not favor universal molecular cutoffs, as different microorganisms evolve at different rates, species may be younger or older, and sequence divergence alone does not equate to a biological notion of species (52). Indeed, PEs predicted from the same locus varied in terms of average percent nucleotide divergence (Table 4). Twenty-five percent of the PEs detected had an average within-ecotype nucleotide divergence of ≥1%, and 55% of the same PEs had a maximum nucleotide variation of ≥1%. In one instance, the maximum nucleotide identity within a PE clade was 7.8% (Table 4). The specific percentages are for the loci that we studied, which may not be representative of the entire genome, complicating direct comparison of our results with genomic ANI values. Although clusters of closely related sequences can result from causes other than adaptation (52), our observation of *Synechococcus* ecotypes is one of a growing number of examples of taxa that exhibit numerous extremely closely related sequence clusters corresponding to ecologically distinct populations (10, 12, 13, 22, 25, 42, 43, 50, 51, 58). These clusters occur at a scale of sequence difference well within that formerly used to classify named species. A new tradition of understanding small-scale genetic diversity is needed in microbiology, one that takes full advantage of insights into the mechanisms driving diversification and is open-minded with respect to the amount of genetic difference that correlates with evolved population structure. As pointed out by Fuhrman and Campbell (19), realization that the divergence of extremely small sequence clusters may involve the evolution of ecologically specialized populations "could multiply our perception of natural diversity by several times."

One limitation of single-locus analysis of diversity is variation in the molecular resolution of the locus being studied, as some loci evolve more slowly than others and it is difficult to discern which locus may best represent the evolutionary trajectory of the organism. The relationship between molecular resolution and PE prediction shows clearly that PE prediction is a function of the resolving power of the locus. Even the most rapidly evolving locus that we examined may underestimate ecological diversity, on the basis of the occurrence of multiple dominant sequences in single demarcated PEs. Another major limitation of single-locus analysis is that it does not account for the extent of recombination or linkage between loci, which could lead to incorrect assignment of loci to PEs, leading to a skewed impression of the evolutionary history of the organismal populations whose relationships are being modeled by genetic data. This may partially explain the low bootstrap values that were sometimes seen in trees (4, 7). Misclassification of organisms from the sequence of a single gene would also obfuscate the habitat associations of ecotypes. To account for these limitations, multilocus analyses are increasingly implemented to provide a more realistic evolutionary history of cultivated microorganisms (14, 15, 37, 49). In separate work, we have developed a cultivation-independent, multilocus sequence analysis (MLSA) approach (Melendrez et al., unpublished), and the single-locus data from this study provide a reference for those studies.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation Frontiers in Integrative Biology Research Program (EF-0328698), the National Aeronautics and Space Administration Exobiology Program (NAG5-8807), and the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Genomic Science Program (GSP). This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL), contract no. 112443.

We appreciate the assistance from National Park Service personnel at Yellowstone National Park. We also thank John Heidelberg for his assistance with metagenomic analyses that became important in locus selection, Alex Koeppel for his assistance with answering questions and troubleshooting protocols for ES, Juan Carlos Francisco for analyzing ecological diversity with AdaptML, and Jason Wood for metagenomic analysis assistance and optimization of ES.

REFERENCES

1. **Allewalt, J. P., M. M. Bateson, N. P. Revsbech, K. Slack, and D. M. Ward.** 2006. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in Octopus Spring microbial mat community of Yellowstone National Park. Appl. Environ. Microbiol. **72:**544–550.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://blast.ncbi.nlm.nih.gov.
3. **Bhaya, D., et al.** 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analysis. ISME J. **1:**703–713.
4. **Boni, M. F., D. Posada, and M. W. Feldman.** 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. Genetics **176:** 1035–1047.
5. **Chao, A. and T.-J. Shen.** 2003. Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. Environ. Ecol. Stat. **10:**429–443.
6. **Chao, A.** 1984. Nonparametric estimation of the number of classes in a population. Scand. Stat. Theory Appl. **11:**265–270.
7. **Chare, E. R., E. A. Gould, and E. C. Holmes.** 2003. Phylogenetic analysis reveals a low rate of homologous recombination in negative-sense RNA viruses. J. Gen. Virol. **84:**2691–2703.
8. **Cohan, F. M., and A. F. Koeppel.** 2008. The origins of ecological diversity in prokaryotes. Curr. Biol. **18:**R1024–R1034.
9. **Cohan, F. M., and E. B. Perry.** 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr. Biol. **17:**R373–R386.
10. **Cohan, F. M., A. Koeppel, and D. Krizanc.** 2006. Sequence-based discovery of ecological diversity within *Legionella*, p. 367–376. *In* N. P. Cianciotto et al. (ed.), *Legionella*: state of the art 30 years after its recognition. ASM Press, Washington, DC.
11. **Colwell, R. K., and J. A. Coddington.** 1994. Estimating terrestrial biodiversity through extrapolation. Philos. Trans. R. Soc. Lond. B Biol. Sci. **345:**101–118.
12. **Connor, N., et al.** 2010. The ecology of speciation in *Bacillus*. Appl. Environ. Microbiol. **76:**1349–1358.
13. **Denef, V. J., et al.** 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proc. Natl. Acad. Sci. U. S. A. **107:**2383–2390.
14. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186:**1518–1530.
15. **Feil, E. J., J. Maynard-Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombination parameters in *Streptococcus pneumoniae* from multi-locus sequence typing data. Genetics **154:**1439–1450.
16. **Ferris, M. J., M. Kühl, A. Wieland, and D. M. Ward.** 2003. Cyanobacterial ecotypes in different optical microenvironments of a 68°C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. Appl. Environ. Microbiol. **69:**2893–2898.
17. **Ferris, M. J., and D. M. Ward.** 1997. Season distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. **63:**1375–1381.
18. **Ferris, M. J., G. Muyzer, and D. M. Ward.** 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. **62:**340–346.
19. **Fuhrman, J. A., and L. Campbell.** 1998. Microbial microdiversity. Nature **393:**410–411.
20. **Goodfellow, M., G. P. Manfio, and J. Chun.** 1997. Towards a practical species concept for cultivable bacteria, p. 25–59. *In* M. F. Claridge, H. A. Dawah, and M. R. Wilson (ed.), Species: the units of biodiversity. Chapman & Hall, London, United Kingdom.
21. **Hanage, W. P., W. Fraser, and B. G. Spratt.** 2006. Sequences, sequence clusters and bacterial species. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361:**1917–1927.
22. **Hunt, D. E., et al.** 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science **320:**1081–1085.
23. **Johnson, J. L.** 1973. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. Int. J. Syst. Bacteriol. **23:**308–315.
24. **Kettle, G. C., et al.** 2007. Patterns and implications of gene gain and loss in the evolution of *Procholorococcus*. PLoS Genet. **3:**2515–2528.
25. **Koeppel, A., et al.** 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc. Natl. Acad. Sci. U. S. A. **105:**2504–2509.
26. **Konstantinidis, K. T., and J. M. Tiedje.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. **102:** 2567–2572.
27. **Librado, P., and J. Rozas.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451–1452.
28. **Manning, S. D., et al.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U. S. A. **105:**4868–4873.
29. **Martiny, A. C., A. P. K. Tai, D. Veneziano, F. Primeau, and S. W. Chisholm.** 2009. Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. Environ. Microbiol. **11:**823–832.
30. **Miller, S. R., and R. W. Castenholz.** 2000. Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. Appl. Environ. Microbiol. **66:**4222–4229.
31. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the number of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
32. **Palys, T., L. K. Nakamura, and F. M. Cohan.** 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int. J. Syst. Bacteriol. **47:**1145–1156.
33. **Papke, R. T., et al.** 2007. Searching for species in *Haloarchaea*. Proc. Natl. Acad. Sci. U. S. A. **104:**14092–14097.
34. **Peary, J. A., and R. W. Castenholz.** 1964. Temperature strains of a thermophilic blue-green alga. Nature **202:**720–721.
35. **Perriere, G., and M. Gouy.** 1996. WWW-Query: an online retrieval system for biological sequence banks. Biochimie **78:**364–369.
36. **Ramsing, N. B., M. J. Ferris, and D. M. Ward.** 2000. Highly ordered vertical structure of *Synechococcus* population within the one-millimeter-thick photic zone of a hot spring cyanobacterial mat. Appl. Environ. Microbiol. **66:**1038–1049.
37. **Salerno, A., et al.** 2007. Recombining population structure of *Plesiomonas shigelloides* (*Enterobacteriaceae*) revealed by multilocus sequence typing. J. Bacteriol. **189:**7808–7818.
38. **Schloss, P. D., and J. Handelsman.** 2006. Introducing SONS, a tool for OTU-based comparisons of membership and structure between microbial communities. Appl. Environ. Microbiol. **72:**6773–6779.
39. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501–1506.
40. **Schloss, P. D., B. R. Larget, and J. Handelsman.** 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. Appl. Environ. Microbiol. **70:**5485–5492.
41. **Sikorski, J., and E. Nevo.** 2007. Patterns of thermal adaptation of *Bacillus simplex* to the microclimatically contrasting slopes of 'Evolution Canyons' I and II, Israel. Environ. Microbiol. **9:**716–726.
42. **Smith, N. H., S. V. Gordon, R. de la Rua-Domenech, R. S. Clifton-Hadley, and R. G. Hewinson.** 2006. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat. Rev. Microbiol. **4:**670–681.
43. **Soyer, Y., R. H. Orsi, L. D. Rodriguez-Rivera, Q. Sun, and M. Wiedmann.** 2009. Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes. BMC Evol. Biol. **9:**264.
44. **Stackenbrandt, E., and J. Ebers.** 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today **November:**153–155.
45. **Stackenbrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846–849.
46. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
47. **Van Valen, L.** 1976. Ecological species, multispecies, and oaks. Taxon **25:** 233–239.
48. **Venter, J. C., et al.** 2004. Environmental shotgun sequencing of the Sargasso Sea. Science **304:**66–74.
49. **Vitorino, L. R., et al.** 2008. Fine-scale phylogeographic structure of *Borrelia lusitaniae* revealed by multilocus sequence typing. PLoS One **3:**1–13.
50. **Walk, S. T., et al.** 2009. Cryptic lineages of the genus *Escherichia*. Appl. Environ. Microbiol. **75:**6534–6544.
51. **Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky, and T. S. Whittam.** 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. Environ. Microbiol. **9:**2274–2288.
52. **Ward, D. M., et al.** 2008. Genomes, environmental genomics and the issue of microbial species. Heredity **100:**207–219.
53. **Ward, D. M., et al.** 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. Philos. Trans. R. Soc. Lond. B Biol. Sci. **361:**1997–2008.
54. **Ward, D. M., and F. M. Cohan.** 2005. Microbial diversity in hot spring cyanobacterial mats: pattern and prediction, p. 185–201. *In* B. Inskeep and T. McDermott (ed.), Geothermal biology and geochemistry in Yellowstone National Park. Thermal Biology Institute, Bozeman, MT.
55. **Ward, D. M.** 1998. A natural species concept for prokaryotes Curr. Opin. Microbiol. **1:**271–277.
56. **Wayne, L. G., et al.** 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. **37:**463–464.
57. **Whitaker, R. J., and J. F. Banfield.** 2005. Population dynamics through the lens of extreme environments. Rev. Mineral. Geochem. **59:**259–277.
58. **Zhou, D. S., et al.** 2004. DNA microarray analysis of genome dynamics in *Yersinia pestis*: insights into bacterial genome microevolution and niche adaptation. J. Bacteriol. **186:**5138–5146.