Applied and Environmental Microbiology, Aug. 2007, p. 4751–4759 Vol. 73, No. 15
0099-2240/07/$08.00+0 doi:10.1128/AEM.02945-06

Phylogeography of the Thermophilic Cyanobacterium *Mastigocladus laminosus*▽

Scott R. Miller,[1]* Richard W. Castenholz,[2,3] and Deana Pedersen[1]

Division of Biological Sciences, The University of Montana, Missoula, Montana 59812,[1] and Center for Ecology and Evolutionary Biology, Department of Biology,[2] and NASA Astrobiology Institute,[3] University of Oregon, Eugene, Oregon 97403

Received 19 December 2006/Accepted 1 June 2007

We have taken a phylogeographic approach to investigate the demographic and evolutionary processes that have shaped the geographic patterns of genetic diversity for a sample of isolates of the cosmopolitan thermophilic cyanobacterial *Mastigocladus laminosus* morphotype collected from throughout most of its range. Although *M. laminosus* is found in thermal areas throughout the world, our observation that populations are typically genetically differentiated on local geographic scales suggests the existence of dispersal barriers, a conclusion corroborated by evidence for genetic isolation by distance. Genealogies inferred using nitrogen metabolism gene sequence data suggest that a significant amount of the extant global diversity of *M. laminosus* can be traced back to a common ancestor associated with the western North American hot spot currently located below Yellowstone National Park. Estimated intragenic recombination rates are comparable to those of pathogenic bacteria known for their capacity to exchange DNA, indicating that genetic exchange has played an important role in generating novel variation during *M. laminosus* diversification. Selection has constrained protein changes at loci involved in the assimilation of both dinitrogen and nitrate, suggesting the historic use of both nitrogen sources in this heterocystous cyanobacterium. Lineage-specific differences in thermal performance were also observed.

A recent theme in the study of microbial diversity has been the issue of how genetic (and phenotypic) variation of microorganisms is distributed on a geographic scale (reviewed in reference 21). Whereas it is clear that microbial community structure varies spatially as a result of the physical and chemical heterogeneity of the environment, the discussion regarding microbial biogeography has principally centered on whether the homogenization of genetic variation by the dispersal of migrants across geographically separated but similar habitats outpaces the genetic differentiation of populations by genetic drift or in response to local selection. The results of experimental evolution experiments in the laboratory (4) suggest that spatially structured microbial populations in nature should rapidly diverge from each other, provided that migratory gene flow among them is low (42). In contrast, the idea that dispersal barriers do not exist for free-living microorganisms and, consequently, that the abundance of a microorganism at a location is determined solely by environmental factors, is longstanding (3) and has been recently championed for eukaryotic microorganisms on the basis of morphological criteria (12).

The island-like nature of hot springs makes them particularly suitable habitats for investigating microbial dispersal and its contribution to geographic distribution patterns, and several recent investigations have provided evidence for dispersal barriers for thermophilic microorganisms. Whitaker et al. (41) reported that genetic differentiation between *Sulfolobus* populations increased with increased geographic distance, a predicted outcome of dispersal limitation (43). A molecular survey of cyanobacterial diversity in several of the world's major thermal areas (28) corroborated long-standing observations of differences in their community composition (8) that are best explained by isolation. Similarly, migration was undetectable between two genetically differentiated Yellowstone National Park populations of the heterocyst-forming, moderately thermophilic cyanobacterium *Mastigocladus* (*Fischerella*) *laminosus* (order *Stigonematales*), which are separated by 50 km (24).

Here, we take a phylogeographic approach (1) to expand our analysis of the demographic and evolutionary processes that have shaped the spatial distribution of *M. laminosus* diversity. In addition to enabling the investigation of dispersal limitations, this approach also has the potential to identify the historical site(s) of origin from which extant diversity has spread. Recent dispersal is expected to leave a specific genealogical signature, in which populations from a site of origin are paraphyletic with respect to descendant populations in more recently colonized locations, as has been observed for African populations of humans, for example (6). Identifying these ancestor-descendant relationships among a sample of gene sequences is often better resolved through the inference of genealogical networks than by traditional phylogenetics (30). This is because both ancestor and descendant alleles may be extant in a sample of closely related individuals, but a bifurcating phylogenetic tree makes the limiting assumption that ancestor-descendant relationships are hierarchical (with extant descendant alleles related to each other by descent from extinct ancestors).

M. laminosus provides an excellent system for investigating the organization of microbial diversity at a variety of geographic scales. This bacterium is present virtually worldwide,

* Corresponding author. Mailing address: 32 Campus Dr. #4824, Division of Biological Sciences, The University of Montana, Missoula, MT 59812. Phone: (406) 243-5149. Fax: (406) 243-4184. E-mail: scott.miller@mso.umt.edu.

▽ Published ahead of print on 8 June 2007.

though not necessarily in great abundance, in alkaline hot springs at temperatures below approximately 57°C (8). *M. laminosus* also exhibits great tolerance of freezing and desiccation (8), which may facilitate airborne dispersal and help to explain its cosmopolitan distribution. To describe the broad phylogenetic patterns of *M. laminosus* global diversity, we first characterized 37 strains isolated from throughout most of its range at the conserved 16S rRNA gene. In light of our prior observation of population genetic differentiation of *M. laminosus* within the Yellowstone region (24), it was of interest to comprehensively evaluate whether local differentiation is the norm for this cosmopolitan bacterium over a global scale, and we therefore further characterized the strains at three more rapidly evolving genes involved in nitrogen metabolism to test for genetic isolation by distance. We also reconstructed genealogical networks for the latter loci to identify the ancestor-descendant relationships among alleles in the sample. From these networks, we inferred possible dispersal patterns during *M. laminosus* diversification. Finally, we evaluated whether it was possible to draw inferences regarding the historical nitrogen requirements of this group based on the selective histories of the different metabolic genes, and we also tested whether *M. laminosus* lineages are phenotypically differentiated from each other with respect to thermal performance.

MATERIALS AND METHODS

Strains and culture maintenance. The 37 strains of *M. laminosus* analyzed in this study were obtained from either the University of Oregon's Culture Collection of Microorganisms from Extreme Environments (CCMEE) or from The University of Montana Cyanobacterial Culture Collection (Table 1). Additional information about the CCMEE strains is available at the University of Oregon CCMEE website (http://cultures.uoregon.edu), and information for the University of Montana strains was reported by Miller et al. (24). Strains were maintained at 50°C in 25 ml of D medium under 75 μmol photons m^{-2} s^{-1} of 12-h light/12-h dark cool white fluorescent light.

DNA isolation, PCR, and sequencing. Genomic DNA was isolated from cultures, and a ca. 950-bp fragment of the small subunit rRNA gene (*Escherichia coli* positions 360 to 1326) was amplified as previously described (23). Also amplified were a roughly 800-bp fragment (*Anabaena* strain PCC 7120 nucleotide positions 88 to 884) of *nifH* (iron protein of nitrogenase), a fragment of the assimilatory nitrate reductase gene *narB* (*Anabaena* strain PCC 7120 positions 1056 to 2072), and an approximately 1.2-kbp genomic fragment including the nearly complete sequence of the heterocyst development gene *devH* (nucleotide positions 594 to 1232 in the fragment) and a partial sequence (the first 402 nucleotides of the fragment) of *argS* (arginyl tRNA synthetase). Reaction conditions and primer sets for amplification of these nitrogen metabolism loci were as reported by Miller et al. (24).

Phylogenetic analyses and genealogical networks. *M. laminosus* phylogenies were reconstructed from 839 nucleotides of the 16S rRNA gene with maximum likelihood, maximum parsimony, and neighbor-joining methods using PAUP* version 4.0b (37) following sequence alignment as described previously (23). The data set included outgroups *Chlorogloeopsis* strain PCC 6718 (AF132777) and *Chroococcidiopsis* strain PCC 7203 (from the RDPII website [http://rdp.cme.msu.edu]), and trees were rooted with the latter sequence. For the likelihood analysis, the model of DNA sequence evolution was chosen by hierarchical likelihood ratio tests and Akaike information criterion as implemented in Modeltest (31). The model selected (HKY plus G plus I) estimates the transition/transversion ratio and incorporates among nucleotide site rate heterogeneities by estimating both the proportion of invariant sites and the shape of the discrete approximation ($n = 4$ categories) of a gamma distribution for variable sites. The starting tree for the likelihood analysis was obtained by random stepwise sequence addition, and the subsequent heuristic search was performed using the tree-bisection reconnection branch-swapping algorithm. An exhaustive search of all trees was performed for the parsimony analysis. The likelihood analysis was bootstrap replicated 1,000 times, whereas both the neighbor-joining tree and the maximum parsimony tree were bootstrapped 10,000 times.

An approach analogous to that described above was used to reconstruct phylogenies for the nitrogen metabolism genes *nifH* (738 nucleotides), *narB* (918 nucleotides), and the *devH* region (1,232 nucleotides). Phylogenies were rooted with the outgroup *Anabaena* strain PCC 7120. The models of DNA evolution chosen by Modeltest for use in likelihood analyses were as follows: the K80 (Kimura two-parameter model [19]) plus the G model for *devH*; the K80 plus G plus I model for *narB*; and the GTR plus G model for *nifH*. All trees were bootstrap replicated 1,000 times.

Minimum spanning trees (20) were computed for each nitrogen metabolism gene from a matrix of pair-wise nucleotide differences, and a minimum spanning network was subsequently inferred by embedding all minimum spanning trees for a gene (11), as implemented in the software package Arlequin version 2.0 (36).

Population genetic analyses. Aligned nucleotide sequences for the nitrogen metabolism loci were analyzed with respect to the following with DnaSP (34): the number of alleles observed at each locus, synonymous and nonsynonymous (replacement) polymorphic sites, and nucleotide diversity at synonymous and

TABLE 1. Multilocus haplotypes for 16S rRNA gene groups of *M. laminosus*

Group and strain	Origin	Haplotype (alleles[a])
Group 1		
B6A	Boiling River, Yellowstone National Park (NP)	1 (1/1/1)
CCMEE 5207	Chocolate Pots, Yellowstone NP	2 (2/1/1)
CCMEE 5208	Obsidian Pool, Yellowstone NP	3 (2/2/1)
W16C	White Creek, Yellowstone NP	4 (3/2/1)
W17B	White Creek, Yellowstone NP	5 (3/3/1)
W1C	White Creek, Yellowstone NP	6 (3/3/2)
W25B	White Creek, Yellowstone NP	7 (2/3/2)
CCMEE 5327	Spring near Cava Scura, Ischia, Italy	8 (5/2/3)
CCMEE 5328	Spring near Cava Scura, Ischia, Italy	8 (5/2/3)
CCMEE 5337	McCredie Hot Springs, OR	9 (6/3/7)
CCMEE 5320	Hot spring, Krisuvik Hot Springs, Iceland	10 (6/3/8)
CCMEE 5273	Harrison Hot Springs, British Columbia	10 (6/3/8)
CCMEE 5198	Ohinemutu, New Zealand	11 (-/5/4)
CCMEE 5186	Ngaratuatara, Whakarewawera, New Zealand	12 (-/10/4)
CCMEE 5335	Ngaratuatara, Whakarewawera, New Zealand	12 (-/10/4)
CCMEE 5281	Sklene Teplice, Slovakia	13 (5/2/6)
CCMEE 5282	Sklene Teplice, Slovakia	13 (5/2/6)
CCMEE 5193	Puyeohe Springs, Chile	14 (7/2/5)
Group 2		
CCMEE 5201	Waitangi Springs, Rotoma, New Zealand	15 (12/13/13)
CCMEE 5202	Waitangi Springs, Rotoma, New Zealand	15 (12/13/13)
CCMEE 5203	Waitangi Springs, Rotoma, New Zealand	15 (12/13/13)
Group 3		
CCMEE 5323	Geysir Springs, Iceland	16 (13/11/14)
CCMEE 5321	Hveragerdi Springs, Iceland	17 (13/12/14)
CCMEE 5324	Hveragerdi Springs, Iceland	17 (13/12/14)
CCMEE 5326	Surtsey steam vents, Iceland	17 (13/12/14)
CCMEE 5272	Furnas, São Miguel, Azores, Portugal	17 (13/12/14)
Group 4		
CCMEE 5332	Bozeman Hot Springs, MT	18 (11/8/9)
CCMEE 5192	Puyeohe Springs, Chile	18 (11/8/9)
CCMEE 5204	Oman	18 (11/8/9)
CCMEE 5205	Oman	18 (11/8/9)
CCMEE 5379	Oman	18 (11/8/9)
CCMEE 5380	Oman	18 (11/8/9)
Group 5		
CCMEE 5268	Chena Hot Springs, AK	19 (4/6/9)
Group 6		
CCMEE 5319	Hot spring near Lake Amatitlan, Guatemala	20 (8/9/10)
CCMEE 5267	Chena Hot Springs, AK	21 (9/7/10)
CCMEE 5318	El Salvador	22 (10/9/11)
Group 7		
CCMEE 5329	Sokokura Hot Spring, Hakone, Japan	23 (14/4/12)
CCMEE 5331	Sokokura Hot Spring, Hakone, Japan	23 (14/4/12)

[a] Each multilocus haplotype is based on the allele combination for *narB*/*nifH*/*devH*. Haplotypes are grouped according to 16S rRNA lineage membership.

nonsynonymous sites (the average number of synonymous and nonsynonymous nucleotide differences between a pair of sequences, respectively).

A Mantel test was used to test whether there was a statistically significant relationship between genetic differentiation (nucleotide diversity) and geographic distance of strain pairs. This test is more appropriate than conventional correlation analysis, because physical distances between population pairs are not independent, and was based on 10,000 randomizations as implemented by the program IBD version 1.52 (5).

Recombination rates for each locus were estimated by a modification of the composite likelihood method of Hudson (15), which accommodates a finite sites model of sequence evolution (22), as implemented by the software package LDhat (www.stats.ox.ac.uk/~mcvean/LDhat.html). The latter model allows for recurrent mutations to have occurred at a nucleotide position, a possibility that becomes more probable as sequences diverge. The method assumes that the gene genealogies of the loci can be modeled as a coalescent process according to the neutral Wright-Fisher model (14), but the estimates and significance tests appear to be robust to minor deviations from this model (22). Population-scaled recombination rates were expressed relative to a finite sites approximation of Watterson's population-scaled mutation rate (39). A likelihood permutation test (22) was used to test whether each estimated recombination rate was significantly greater than zero. In the absence of recombination, the likelihood estimate is expected to be independent of the order in which nucleotide sites are arranged along a sequence alignment (i.e., locations of nucleotide sites in the alignment are exchangeable). In the presence of recombination, however, the likelihood of observing the data does depend on data order, because close sites are more likely to be in linkage disequilibrium. Likelihood estimates for random rearrangements of the order of the nucleotide site data therefore can test the null hypothesis that data are exchangeable (i.e., no recombination). We rejected the null hypothesis of no recombination if fewer than 5% of 1,000 permutated data sets had a composite likelihood score equal to or higher than the estimate for the original data.

Growth experiment. Temperature dependence of growth rate was estimated for two randomly selected strains belonging to each of the seven 16S rRNA gene groups identified in the phylogenetic survey (with the exception of the group consisting solely of strain CCMEE 5268). Maintenance cultures (75 ml of D medium [7] in 125-ml Erlenmeyer flasks incubated at 50°C under 100 μmol photons m^{-2} s^{-1} of CWF light for a 12-h light/12-h dark cycle) of each strain were split into duplicate subclones (in 25 ml of D medium in test tubes) and propagated for three generations, at which time each subclone was divided into duplicate 25-ml cultures to an initial optical density (measured at 750 nm) of 0.005. Thus, a total of $13 \times 2 \times 2 = 52$ culture tubes were propagated during the initial growth experiment at 50°C. The exponential growth rate was estimated for each culture from semilogarithmic plots of the increase in optical density over time. Provided that a strain was capable of growth, one replicate of each subclone was randomly selected to serve as inoculum for the next round of the experiment. Subsequent assays were performed at 55, 57, and 58°C, the temperature at which no strain survived.

Nucleotide sequence accession numbers. The sequences reported in this paper have been deposited in the GenBank database and assigned accession numbers as follows: *rrn-16S* (EF566829 to EF566864); *argS* (EF571857 to EF571894); *devH* (EF581319 to EF581356); *narB* (EF570505 to EF570539); *nifH* (EF570540 to EF570572, DQ385912, DQ385913, DQ385916, DQ385917, and DQ385924).

RESULTS AND DISCUSSION

16S rRNA gene lineages of *M. laminosus*. To investigate the global genetic diversity of *M. laminosus*, we characterized 37 laboratory isolates of this cyanobacterium from thermal areas around the globe (Table 1). Each of the isolates belonged to one of seven closely related groups based on the complete identity of a partial sequence (835 bp spanning *E. coli* positions 390 to 1249) of the 16S rRNA gene (Table 1). Nearly half of the sampled strains belonged to group 1. Only 27 (3.2%) of the nucleotide sites were variable, and each group could be assigned to one of three clusters based on their relative amount of sequence divergence: group 2, which shares no greater than 98.1% similarity with any other group; group 3, which shares no greater than 99.0% similarity with any other group; and the remaining five groups, each of which differ from group 1 by at most 0.5% (group 5 by a single nucleotide, or 0.2%; group 4 and group 7 by 0.4%; group 6 by 0.5%). The level of dissimilarity for the two most divergent lineages (group 2 and group 3) was under 3%, similar to that observed for ecologically divergent A and B clades of the thermophilic cyanobacterium *Synechococcus* sp. (23, 38). That is, extant global *M. laminosus* diversity is comparable to the degree of genetic differentiation between *Synechococcus* lineages that have diverged in thermal ecology and which in some cases may be separated by only millimeters along a hot spring outflow channel.

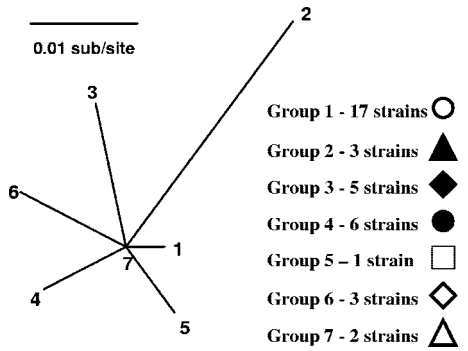

FIG. 1. Maximum likelihood phylogeny of 37 isolates of *M. laminosus* reconstructed from 839 nucleotides of the 16S rRNA gene, rooted with outgroup *Chroococcidiopsis* strain PCC7203 (not shown). Bootstrap support at the root node for the monophyly of *M. laminosus* was 96% for the likelihood analysis and 100% for neighbor-joining and parsimony analyses, based on 1,000, 10,000, and 10,000 pseudoreplicates, respectively. Lineage symbols are retained in Fig. 4. sub, substitution.

Comparison with the limited sequence data for *M. laminosus* deposited in GenBank suggests that our sample adequately captures the 16S rRNA level diversity of this bacterium. Cultured strains from other collections (New Zealand [AB075986], Japan [AB093487], Jordan [DQ471442 and DQ471444], and Yellowstone [AB075987]) belong to group 3, group 7, group 4, and group 1, respectively. Similarly, environmental clones from Australia (AF407731 and AF407696), the Philippines (DQ131173), and Greenland (AF132788) belong to group 1, group 7, and group 3, respectively.

Trees reconstructed from these highly similar sequence data were unresolved, resembling a star phylogeny in which the seven 16S rRNA groups radiate from a central root node (Fig. 1). Likelihood, parsimony, and neighbor-joining methods yielded qualitatively similar results, with the monophyly of *M. laminosus* supported by strong bootstrap analysis in all three phylogenies. The only other node that received even moderate bootstrap support (and only for parsimony and neighbor-joining analyses, at 52% and 57%, respectively) joined group 4 and group 6 as sister taxa.

Small subunit rRNA gene sequences of the different groups were also distinguished by the presence or absence of an unusual quasi-palindromic insert within the V1 region of the gene (*E. coli* positions 61 to 106), which encodes the spur of the ribosomal small subunit. Group 1 and group 7 strains have this feature, resulting in a longer spur RNA hairpin (Fig. 2). These groups differ at five nucleotide positions within the insert (not shown). Because this region was not shared by all strains, it was not included in the phylogenetic analyses above. We believe

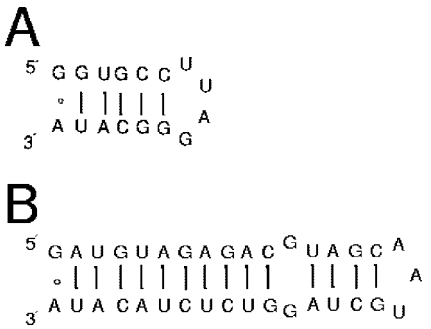

FIG. 2. MFold-inferred secondary structural models for the distal end of the V1-encoded ribosomal spur (*E. coli* positions 68 to 101) for group 2 (A) and group 7 (B) *M. laminosus*.

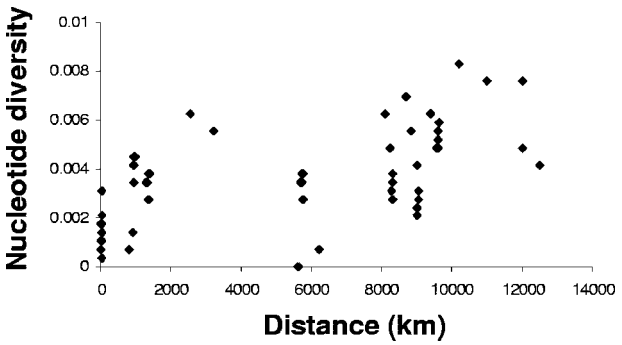

FIG. 3. Relationship between genetic differentiation and geographic distance for group 1 *M. laminosus* strains.

that this segment is the result of a single insertion event (with inferred recombination breakpoints at *E. coli* positions 74 and 94), based on its unusual length (47 nucleotides) for a cyanobacterium: the average length based on a survey of V1 DNA for 531 cyanobacterial laboratory isolates is 28 nucleotides (25). The V1 DNA of another member of the *Stigonematales*, the genus *Nostochopsis*, also shares this feature (13a). The only other known example of an unusually long V1 region in the cyanobacteria is for chlorophyll *d*-producing cyanobacteria, which appear to have acquired DNA encoding a longer spur by lateral gene transfer from a betaproteobacterium (25). The origins of this feature in *M. laminosus*, and its possible functional significance, are not known.

Local genetic differentiation of a cosmopolitan bacterium. Whereas a few of the identified 16S rRNA gene groups were endemic to a single location (e.g., group 2 in Waitangi Springs, New Zealand), many were found at multiple, geographically distant sites (Table 1), highlighting this bacterium's dispersal capabilities. For example, representatives of group 1 have been isolated from hot springs in North and South America, Europe (continental and Iceland), and New Zealand (Table 1).

A closer look at the genetic diversity of these isolates, however, revealed a general pattern of genetic differentiation on local geographic scales that is consistent with the presence of dispersal barriers. We obtained additional sequence data for three nitrogen metabolism genes that we had previously found to be polymorphic within and between Yellowstone populations of *M. laminosus* (24): *nifH*, encoding the iron protein of nitrogenase; *narB*, the gene encoding assimilatory nitrate reductase; and the genomic region including *devH*, which codes for a DNA-binding protein required for the development of a functional heterocyst (32), and *argS*, arginyl tRNA synthetase. These additional data revealed 23 distinct multilocus haplotypes (Table 1), resulting from different combinations of 13 observed alleles for *nifH* and 14 alleles each for *narB* and the *devH* region. With one exception (the sharing of *devH* allele 9 by group 4 and group 5), each allele at a locus was unique to a 16S rRNA gene group. Fourteen of the haplotypes belonged to group 1. Although a few of these haplotypes were observed in geographically distant populations (e.g., haplotype 18 in Oman, Montana, and Chile), most haplotypes were only observed at a single location. For example, within Yellowstone National Park, haplotype 1 was unique to Boiling River, haplotype 2 to Chocolate Pots, haplotype 3 to Obsidian Pool, and haplotypes 4 to 7 to White Creek in the Lower Geyser Basin.

Genetic isolation by distance in group 1 *M. laminosus* strains. In addition to this observed local population differentiation, another expectation for organisms with restricted migration is that there should be a positive correlation between the amount of genetic divergence between pairs of strains and the geographic distances separating their populations of origin (43). Despite the growing evidence for the geographic structuring of microbial diversity (reviewed in reference 21), there has been little investigation of this specific issue of isolation by distance. Whitaker et al. (41) found a positive association for *Sulfolobus* strains from a total of seven populations from Yellowstone National Park, Lassen National Park, and the Kamchatka peninsula. Our *M. laminosus* haplotype data enable a more comprehensive analysis of the relationship between genetic differentiation and physical distance than previously investigated for microorganisms. We expect that the existence and strength of such a pattern will be dependent on the evolutionary time scale examined. Comparisons at the macroevolutionary scale (e.g., between 16S rRNA lineages) may be expected to obscure such a pattern, since more-ancient groups have simply had more time and, therefore, opportunity, to disperse over large geographic distances than have less-divergent lineages. Therefore, we first investigated this relationship for group 1 *M. laminosus* strains, the most abundant, genetically diverse (in terms of number of haplotypes) and geographically widespread group in the sample. For a combined data set including all three loci (a total of 2,888 bp), there was a statistically significant positive correlation ($R = 0.57$; $P = 0.0015$) between genetic differentiation (measured as the number of nucleotide differences per nucleotide site between a pair of sequences) and geographic distance, providing evidence for isolation by distance (Fig. 3). Though *M. laminosus* has a cosmopolitan distribution, this correlation indicates that migration barriers do exist for this organism. Still, rare long distance dispersal events do occur, and have occurred recently, as evidenced, for example, by the lack of genetic differentiation between group 1 strains CCMEE 5320 and CCMEE 5273 from Iceland and British Columbia, respectively. Over longer time scales, it is also the case that long distance dispersal, even if relatively rare compared with the rate of population differentiation, could potentially bring more divergent lineages into geographic proximity, which would weaken the relationship

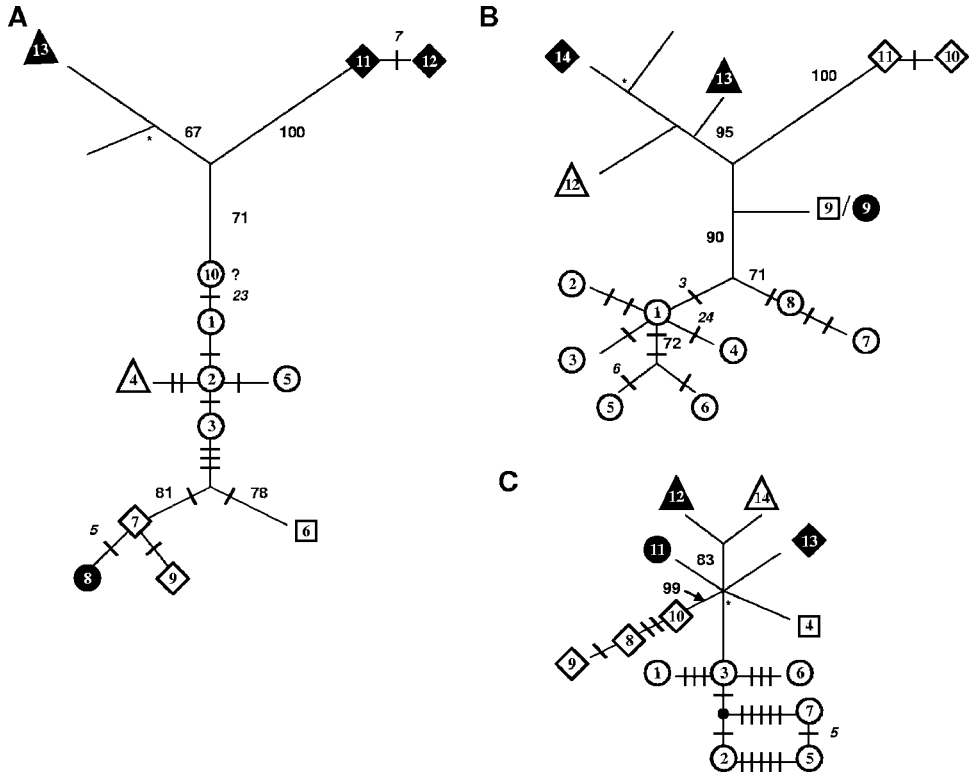

FIG. 4. Minimum spanning gene networks for *nifH* (A), *devH* (B), and *narB* (C). Note that branch lengths are not to scale in units of nucleotide substitutions per site, but that, where noted, hash marks indicate the number of nucleotide differences among alleles. Bootstrap values greater than 50% (based on 1,000 pseudoreplicates) for a maximum likelihood phylogeny are indicated at bifurcating nodes. The root of each network was also inferred with maximum likelihood and is indicated by an asterisk. Haplotype symbols refer to 16S rRNA groups identified in Fig. 1.

between genetic and geographic distance. This was observed for the *M. laminosus* data set as a whole, since there was not a positive relationship between genetic differentiation and distance when data for all seven 16S rRNA groups were analyzed (data not shown).

Genealogical networks of nitrogen metabolism loci. To attempt to better resolve the genealogical relationships among the groups of *M. laminosus*, we reconstructed the evolutionary histories of the three nitrogen metabolism loci. A traditional bifurcating phylogeny frequently fails to accurately describe the genealogical relationships of closely related individuals for a variety of reasons (30), including that the relationship between ancestor and descendant may not be hierarchical (because the ancestor is extant rather than extinct) and that some nodes may be multifurcating rather than bifurcating (because older alleles in a sample are expected to leave more, and possibly more than two, descendants). We therefore took a two-step approach to infer the evolutionary relationships among alleles in our sample for each nitrogen metabolism locus. We first used traditional phylogenetics to identify boot-strap-supported bifurcating nodes. Next, for unresolved polytomies in each phylogenetic analysis, we inferred minimum spanning networks (11), which accommodate ancestor-descendant relationships among extant alleles (with ancestors at internal nodes in the network), multifurcations, and intragenic recombination. The final product for each locus was a diagram of the genealogical relationships among alleles that combines the hierarchical structure of a phylogeny with nested networks to describe unresolved clades (Fig. 4).

Although the tree lengths for the *M. laminosus* clade were similar (maximum likelihood estimates of total length ranged between 0.136 and 0.148 substitutions per site), there were substantial differences among these nitrogen metabolism loci with respect to both genealogy shape and topology (Fig. 4). Three major bootstrap-supported clades were present in the *nifH* genealogy. One consisted of the group 2 allele (Fig. 4A, allele 13), and group 3 alleles 11 and 12. The largest clade included the 10 alleles from the remaining groups. In this clade, group 1 alleles (Fig. 4A) were paraphyletic, with alleles for four other 16S rRNA groups (accounting for 38% of the total observed allelic diversity) descended from extant ancestral group 1 alleles: allele 4 (group 7) and alleles 6 to 9 from groups 4, 5, and 6, respectively (Fig. 4A). Group 6 was also paraphyletic with respect to group 4. All of the observed nucleotide differences among these 10 alleles were synonymous substitutions, indicating that there are no differences in the NifH protein sequence within this clade. Older, ancestral alleles are predicted by population genetic theory to be more abundant and more widespread than younger alleles (10, 40). This expectation was borne out for group 1 alleles 2 and 3, which together account for 70% of the group 1 alleles in the sample, and 32% of the total sample. These alleles are likewise geographically widespread, with allele 2 found in populations from Yellowstone, Italy, Slovakia, and Chile and allele 3 found

in strains isolated from Yellowstone, Oregon, British Columbia, and Iceland.

The inferred position of allele 10 should be regarded with caution. Though grouped by phylogeny reconstruction with other group 1 alleles with moderate bootstrap support, a closer analysis of the distribution pattern of the molecular variation indicates that this allele is likely a recombinant between a group 1 and a group 2 allele. Over the entire sequence, the genetic distances between allele 10 and allele 13 (group 2) and between allele 10 and allele 5 (group 1) are comparable (23 versus 25 nucleotide differences, respectively). However, allele 10 more closely resembles allele 13 over the 5′ end of the sequence (nucleotides 1 to 551), sharing nucleotide identity with allele 13 at 24 of 28 sites at nucleotide positions at which allele 13 and allele 5 differ. In contrast, allele 10 shares identity with allele 5 at all 19 variable sites between sites 552 and 738. Like allele 10, both allele 13 and allele 5 were also derived from New Zealand strains.

In contrast to *nifH*, group 1 *M. laminosus* alleles clustered as a bootstrap-supported, monophyletic clade at the *devH* region and did not exhibit any ancestor-descendant relationships with alleles from other groups (Fig. 4B). Many of the group 1 alleles have descended from allele 1, which was observed in strains isolated from four Yellowstone hot springs. These descendant alleles were found in populations from Yellowstone, continental Europe, Chile, and New Zealand. Other bootstrap-supported clades grouped alleles from groups 2, 3, and 7, as well as the two group 6 alleles, respectively.

The *narB* genealogy is the least resolved of the three, with alleles for different groups radiating as bursts from the inferred root (Fig. 4C). Only the three group 6 alleles and group 2 and group 7 alleles, respectively, clustered with significant bootstrap support. This star-like topology could be the result of early and rapid diversification of different *narB* allele lineages during *M. laminosus* evolution (i.e., it indicates a hard polytomy). Another possibility is that it may reflect the effect of historic recombination events, since intragenic exchange among sequences makes genetic distances among them more similar and therefore more like a star phylogeny (35). The loop connecting group 1 alleles 2, 5, and 7 (along with an inferred allele not observed in the sample) (Fig. 4C) provides clear evidence for intragenic recombination (30). To more explicitly evaluate the historic role of recombination during *M. laminosus* diversification, we estimated recombination rates for each locus below. The observed genealogical discordance among the three loci could be due to differences in the extent of intragenic recombination (in which case different regions within a gene have different evolutionary histories), different patterns of the retention and sorting of ancestral polymorphisms among loci during diversification, or both (2).

Timing of *M. laminosus* diversification. Despite their discordance, the networks did provide better resolution of the deeper splits among *M. laminosus* lineages than the 16S rRNA gene phylogenies. Specifically, group 2 and group 3 were the only lineages that were monophyletic in all three networks (Fig. 4). Relationships among alleles of the other groups included examples of paraphyly (e.g., group 1 at *nifH*) and the lack of divergence between the *devH* allele of group 4 and group 5. By chance and/or due to differences in their selective histories, gene genealogies for a pair of lineages may become reciprocally monophyletic at different times at different loci. But, in general, more closely related lineages are more likely to be paraphyletic at a locus than are more-divergent lineages, since the probability of attaining reciprocal monophyly via lineage sorting of ancestral allele polymorphisms increases with time following divergence (33). These patterns therefore suggest that group 2 and group 3 are older lineages and that the others diverged more recently from each other. This conclusion is corroborated by the trend for group 2 and group 3 to be positioned closer to the root than group 1, for example, in the nitrogen metabolism networks (Fig. 4), and by their greater inferred divergence from the root of the 16S rRNA gene tree (Fig. 1).

We can obtain provisional estimates of the divergence times of *M. laminosus* lineages by assuming the 16S rRNA divergence rate calibration of 1 to 2% per 50 million years (Ma) inferred for both the aphid symbiont *Buchnera* (26) and the domain *Bacteria* in general (27). An upper bound on the age of the most recent common ancestor of extant *M. laminosus* is provided by the amount of divergence of group 2 and group 3 (an average divergence of 0.19 nucleotide substitutions per site from the root node in Fig. 1) and is estimated to be approximately 47 to 95 Ma. Although we can only speculate whether the timing of *M. laminosus* diversification was related to a specific event in Earth's geologic history, it is interesting that this estimate circumscribes the Paleocene-Eocene thermal maximum of 55 Ma. This brief period of extreme, greenhouse gas-induced climate change (9, 18) was marked by terrestrial warming of 5 to 10°C (13). *M. laminosus* is a moderate thermophile and has been observed to dominate biomass in hot spring outflow channels at temperatures below 37°C (unpublished observation). This warming could therefore potentially have created additional suitable habitat of nongeothermal origin and, thereby, opportunities for *M. laminosus* diversification. Similarly, the remaining groups, which are all greater than 99.5% similar to group 1 at the 16S rRNA gene, appear to have diverged from each other within the past approximately 6 to 12.5 Ma.

Evidence for a geologically recent range expansion of group 1 *M. laminosus*. One of the advantages of identifying ancestor-descendant relationships in a sample of alleles is the potential to make inferences regarding the demographic processes that have generated the observed biogeographic pattern of diversity. In particular, it is of interest to try to infer the site(s) of origin from which extant diversity has spread. Although the genealogies for the nitrogen metabolism loci differ, some general conclusions may be drawn. A feature shared by all three networks is for ancestral group 1 alleles to be derived from Yellowstone populations (Fig. 4; Table 1). These include *nifH* alleles 1, 2, and 3, *devH* allele 1, and *narB* allele 3. Given that many alleles in the total sample at each locus either belonged to group 1 or, in the case of *nifH*, descended from an ancestral group 1 allele (8 of 14 alleles for *devH*, 6 of 14 for *narB*, and 9 of 12 for *nifH*, excluding the likely hybrid allele 10), these observations suggest that a significant fraction of the extant global diversity (i.e., haplotype richness) of *M. laminosus*, which appears to have originated within the past 12.5 Ma (see above), may be traced back to a common ancestor associated with the northeast-propagating, western North American hot spot that originated approximately 16 Ma ago and is currently

TABLE 2. Summary of molecular population genetic variation in *Mastigocladus* isolates

Locus	No. of strains sampled	Length (nt)[a]	No. of substitutions		Diversity[b]		
			Silent	Replacement	No. of alleles	$\pi_S \pm$ SD	$\pi_A \pm$ SD
nifH	37	738	64	8	13	0.113 ± 0.0003	0.005 ± 0.0001
narB	34	918	58	15	14	0.074 ± 0.0002	0.005 ± 0.0000
devH region	37	1,232			14		
devH		587	29	1		0.051 ± 0.0002	0.001 ± 0.0000
argS		391	21	36		0.065 ± 0.0004	0.041 ± 0.0003

[a] Sequence length in nucleotides (with indels removed).
[b] π_S and π_A are the number of nucleotide differences between two randomly chosen sequences per silent and replacement site, respectively.

located below Yellowstone National Park (29). Moreover, because descendant (and in some cases extant ancestor) group 1 alleles are geographically widespread (Table 1; Fig. 4), we conclude that there has likely been a geologically recent range expansion of group 1 *M. laminosus* from the Yellowstone hotspot region to thermal areas in Iceland, New Zealand, South America, and continental Europe.

In some cases this proposed range expansion appears to have resulted in secondary contact with 16S rRNA groups that are apparently more restricted in their present geographic distributions (e.g., New Zealand group 2). Secondary contact between divergent groups creates the opportunity for genetic admixture and thereby the generation of novel variation: as noted above, the mosaic structure of *nifH* allele 10 is a possible example of admixture between group 1 and group 2 *M. laminosus*.

Functional constraints on the evolution of nitrogen assimilation loci. By investigating the patterns of molecular evolution of functional genes, it is possible to draw inferences regarding the historical ecology of a group over evolutionary time. In the case of *M. laminosus*, it is germane to ask whether different nitrogen assimilation pathways have experienced different selective histories. Our intuition may lead us to expect that *M. laminosus*, a "heterocystous" cyanobacterium, has specialized in nitrogen-limited habitats and, consequently, that genes involved in nitrate assimilation, such as *narB*, have been under a relaxed selective constraint. On the other hand, examples of both nitrate-assimilating and dinitrogen-assimilating contemporary populations of *M. laminosus* have been reported (24), which suggests that both nutrient sources have been historically important. We can evaluate these alternatives by comparing patterns of mutation among the nitrogen assimilation loci with that of *argS* upstream of *devH*. The loci exhibit an approximately twofold difference in the normalized silent mutation rate, π_S. In contrast, there was a 40-fold range in the rate of protein-changing, replacement substitutions, π_A (Table 2). π_A/π_S estimates the average amount of functional constraint on protein evolution at each locus during *M. laminosus* diversification, with a value for π_A/π_S of 1 indicating neutral evolution. Although the evolution of *argS* has been approximately neutral ($\pi_A/\pi_S = 0.63$), all of the nitrogen assimilation loci have been under severe constraints, with π_A/π_S values of 0.02 for *devH*, 0.04 for *nifH*, and 0.07 for *narB*. For comparison, the low observed π_A for the nitrogen assimilation loci are comparable to the substitution rate for the slowly evolving 16S rRNA gene ($\pi = 0.006$). These results indicate historically comparable levels of functional constraint among genes involved in the assimilatory metabolism of both dinitrogen and nitrate and suggest that both nitrogen sources have been key nutrient sources during the evolutionary diversification of *M. laminosus*.

Substantial recombination during *Mastigocladus* diversification. The genealogical networks, particularly the *narB* network (Fig. 4C), suggest that recombination has been an important process for generating novel variation during *M. laminosus* diversification. To investigate this issue further, it was necessary to consider the possibility that recurrent mutation has occurred during sequence evolution. Multiple mutations at the same nucleotide site during the genealogical history of a sample can create patterns that resemble the effects of recombination and mislead tests for recombination that assume an infinite sites model (i.e., only one mutation per nucleotide site), such as the four-gamete test (16). The sequence data for each of the three nitrogen metabolism loci exhibited evidence for limited recurrent mutation, manifested by the presence of a minor proportion of three-variant nucleotide sites in the sequence alignments (e.g., six three-variant sites among the 112 variant sites for the *devH* data set). Therefore, to estimate the recombination rate at each locus, we implemented a method that assumes a finite sites model to allow for recurrent mutation. LDhat is a maximum likelihood estimator of the population-scaled recombination rate, $\rho = 2N_e r$, where N_e is the effective population size and r is the product of the physical distance and the recombination rate per nucleotide site of the region analyzed (22). A nonparametric permutation test (of 1,000 permutations) tests the hypothesis that the maximum likelihood estimate is not different from zero.

In each case we could reject the null hypothesis that no recombination has occurred, with the intragenic recombination rate estimated to be 0.003 per nucleotide site for the *devH* region ($P = 0.006$), 0.004 per site for *nifH* ($P = 0.011$), and 0.009 per site for *narB* ($P = 0.014$). The higher estimated recombination rate for the latter is consistent with the evidence for greater impact of recombination on the shape of the network for this locus (Fig. 4C), as discussed above. Estimated ratios of the recombination and mutation rates (r/μ) in units of events per nucleotide site were 0.14 for the *devH* region, 0.17 for *nifH*, and 0.47 for *narB*. These are comparable to those of human pathogenic bacteria known for their capacity for genetic exchange. For example, values of r/μ estimated for a data set consisting of 107 disease-associated isolates of *Neisseria meningitidis* collected from around the world ranged between 0.08 and 1.41 (17).

Genetic exchange has clearly been a relevant mechanism for generating intragenic diversity in *M. laminosus*. One could

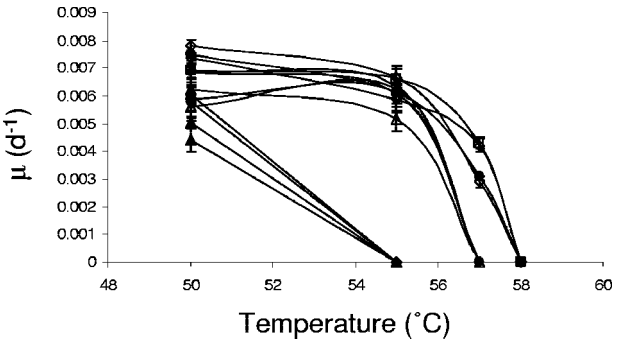

FIG. 5. Temperature dependence of the exponential growth rate (mean ± standard error) for strains of *M. laminosus* representing different 16S rRNA groups: group 1 strains CCMEE 5186 and W25B (open circles); group 2 strains CCMEE 5201 and CCMEE 5202 (closed triangles); group 3 strains CCMEE 5321 and CCMEE 5323 (closed diamonds); group 4 strains CCMEE 5192 and CCMEE 5204 (closed circles); group 5 strain CCMEE 5268 (open squares); group 6 strains CCMEE 5318 and CCMEE 5267 (open diamonds); and group 7 strains CCMEE 5329 and CCMEE 5331 (open triangles).

further estimate the relative role of recombination and mutation for generating novel intergenic variation by $\pi t(r/\mu)/2$ (17), where t is the average recombinational fragment size. An estimate for t requires knowledge of the order of these loci and their physical distances from each other along the chromosome. Though we do not presently have this information for *M. laminosus*, if t were ~1,000 nucleotides, as found for *N. meningitidis* (17), then, based on the above estimates of r/μ, recombination would be about two to five times more important than mutation for generating novel genome combinations in *M. laminosus*.

Differences in thermotolerance among 16S rRNA gene groups. To evaluate whether these genetically differentiated lineages of *M. laminosus* have diversified phenotypically, we tested whether 16S rRNA groups differ in thermal performance. Two representative strains from each 16S rRNA group were randomly selected (with the exception of strain CCMEE 5268, for which a second representative was not available) and subsequently divided into two subclone lines for the duration of the experiment. Exponential growth rates were estimated for duplicate cultures of each subclone at a series of temperatures to test whether strains exhibit differences in the thermal maximum for growth. Strains exhibited substantial differences in thermal performance, particularly with respect to the maximum temperature limit for growth, which can largely be explained by 16S rRNA lineage membership (Fig. 5). At 50°C (strain maintenance temperature), differences among 16S groups were very highly significant ($P < 0.0001$) and explained 58% of the observed variation in growth rate, whereas differences between subclones of a strain or between strains within a group were not significant ($P = 0.17$ and $P = 0.21$, respectively). Results were similar for 55°C, for which differences among groups explained nearly all (95%) of the variation ($P < 0.0001$) and effects of subclone and strain within a group were again not significant ($P = 0.76$ and $P = 0.30$, respectively). Members of group 2 and group 3 (including strains CCMEE 5201/CCMEE 5202 and CCMEE 5321/CCMEE 5323, respectively) tended to have a lower growth rate at 50°C and were unable to grow at 55°C (Fig. 5). Few strains were capable of growth at 57°C (but not at 58°C). These included strains CCMEE 5318 and CCMEE 5267, which belong to group 6, group 5 strain CCMEE 5268, the sole representative of its group in our collection, and group 4 strain CCMEE 5204. While both growth rate and yield were poor for strain CCMEE 5204, the other strain from this group, strain CCMEE 5192, did not grow at this temperature. These results emphasize that, although there are group-specific differences in thermotolerance, at least in one case there appears to be subtle within-group variation in the thermal limit. Whether these differences are indicative of realized ecological differences among lineages remains to be investigated. Given that at least some of the genetic material (e.g., *nifH*) of group 4, group 5, and group 6 has descended from a less thermotolerant group 1 ancestor (Fig. 4A), the genealogical data, while not conclusive, do suggest that the strains with enhanced thermotolerance may have been derived from less thermotolerant ancestors, as in the case of thermophilic *Synechococcus* (23).

Concluding remarks. A phylogeographic approach, in particular the identification of the ancestor-descendant relationships among closely related gene lineages, offers the potential to trace the dynamic spatial histories of microorganisms that are recorded in contemporary sequence variation. In the present study, we have revealed evidence of a geologically recent diversification of the *M. laminosus* clade, a more recent range expansion of group 1 *M. laminosus*, with subsequent local differentiation at most sites, and several apparent examples of very recent, long distance dispersal events, including the recovery of identical multilocus haplotypes from sites in Montana, Chile, and Oman.

Because genetic drift and local selective sweeps remove variation within, but not between, populations in the absence of significant migration, the general pattern of genetic differentiation among populations that we have observed is to be expected. We therefore conclude that the examples of haplotypes with a broad geographic distribution are the product of recent dispersal events and that, given sufficient time, they too will diverge.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for their comments on an earlier version of the manuscript. We also thank Carin Kistler for her technical support during the project.

This research was supported by an NSF research starter grant (MCB-0354738) to S.R.M. We also thank Christie Hendrix and the National Park Service for supporting our work in Yellowstone National Park (NPS permit number YELL-5482 to S.R.M.).

REFERENCES

1. **Avise, J. C.** 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA.
2. **Avise, J. C., and J. R. M. Ball.** 1990. Principles of genealogical concordance in species concepts and biological taxonomy. Oxford Surv. Evol. Biol. **7:** 45–67.
3. **Baas-Becking, L. G. M.** 1934. Geologie of Inleiding Tot de Mileau-Kunde. W. P. van Stokum, The Hague, The Netherlands.
4. **Bennett, A. F., and R. E. Lenski.** 1993. Evolutionary adaptation to temperature II. Thermal niches of experimental lines of *Escherichia coli*. Evolution **47:**1–12.
5. **Bohanak, A. J.** 2002. IBD (Isolation By Distance): a program for analyses of isolation by distance. J. Hered. **93:**153–154.
6. **Cann, R. L., M. Stoneking, and A. C. Wilson.** 1987. Mitochondrial DNA and human evolution. Nature **325:**31–36.
7. **Castenholz, R. W.** 1988. Culturing methods for cyanobacteria. Methods Enzymol. **167:**68–93.

8. **Castenholz, R. W.** 1996. Endemism and biodiversity of thermophilic cyanobacteria. Nova Hedwigia Beih. **112:**33–47.
9. **Dickens, G. R., C. K. Paull, and P. Wallace.** 1997. Direct measurement of in situ methane quantities in a large gas-hydrate reservoir. Nature **385:**426–428.
10. **Donnelly, P., and S. Tavaré.** 1986. The ages of alleles and a coalescent. Adv. Appl. Prob. **18:**1–19.
11. **Excoffier, L., and P. Smouse.** 1994. Using allele frequencies and geographic subdivision to reconstruct gene genealogies within a species. Molecular variance parsimony. Genetics **136:**343–359.
12. **Finlay, B.** 2002. Global dispersal of free-living microbial species. Science **296:**1061–1063.
13. **Fricke, H. C., and S. L. Wing.** 2004. Oxygen isotope and paleobotanical estimates of temperature and δ^{18}O-latitude gradients over North America during the early Eocene. Am. J. Sci. **304:**612–635.

13a. **Gugger, M. F., and L. Hoffmann.** 2004. Polyphyly of true branching cyanobacteria (Stigonematales). Int. J. Syst. Evol. Microbiol. **54:**349–357.

14. **Hudson, R. R.** 1990. Gene genealogies and the coalescent process. Oxford Surv. Evol. Biol. **7:**1–44.
15. **Hudson, R. R.** 2001. Two-locus sampling distributions and their application. Genetics **159:**1805–1817.
16. **Hudson, R. R., and N. L. Kaplan.** 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:**147–164.
17. **Jolley, K. A., D. J. Wilson, P. Kriz, G. McVean, and M. C. J. Maiden.** 2005. The Influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. Mol. Biol. Evol. **22:**562–569.
18. **Kennett, J. P., and L. D. Stott.** 1991. Abrupt deep-sea warming, palaeoceanographic changes and benthic extinctions at the end of the Palaeocene. Nature **353:**225–229.
19. **Kimura, M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16:**111–120.
20. **Kruskal, J. B.** 1956. On the shortest spanning subtree of a graph and the travelling salesman problem. Proc. Am. Math. Soc. **7:**48–50.
21. **Martiny, J. B. H., B. J. M. Bohannan, J. H. Brown, R. K. Colwell, J. A. Fuhrman, J. L. Green, M. C. Horner-Devine, M. Kane, J. A. Krumins, C. R. Kuske, P. J. Morin, S. Naeem, L. Øvreås, A.-L. Reysenbach, V. H. Smith, and J. T. Staley.** 2006. Microbial biogeography: putting microorganisms on the map. Nat. Rev. Microbiol. **4:**102–112.
22. **McVean, G., P. Awadalla, and P. Fearnhead.** 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics **160:**1231–1241.
23. **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.
24. **Miller, S. R., M. D. Purugganan, and S. E. Curtis.** 2006. Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium *Mastigocladus laminosus*. Appl. Environ. Microbiol. **72:**2793–2800.
25. **Miller, S. R., S. Augustine, R. E. Blankenship, J. Selker, T. Le Olson, and A. M. Wood.** 2005. Discovery of a free-living, chlorophyll *d*-producing cyanobacterium with a hybrid proteobacterial/cyanobacterial SSU rRNA gene. Proc. Natl. Acad. Sci. USA **102:**850–855.
26. **Moran, N. A., M. A. Munson, P. Baumann, and H. Ishikawa.** 1993. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. Proc. R. Soc. Lond. B **253:**167–171.
27. **Ochman, H., and A. C. Wilson.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:**74–86.
28. **Papke, R. T., N. B. Ramsing, M. M. Bateson, and D. M. Ward.** 2003. Geographic isolation in thermophilic cyanobacteria. Environ. Microbiol. **5:**650–659.
29. **Pierce, K. L., and L. A. Morgan.** 1992. The track of the Yellowstone hotspot: volcanism, faulting and uplift, p. 1–53. *In* P. K. Link, M. A. Kuntz, and L. B. Platt (ed.), Regional geology of eastern Idaho and western Wyoming: Geological Society of America Memoir 179. Geological Society of America, Boulder, CO.
30. **Posada, D., and K. A. Crandall.** 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. **16:**37–45.
31. **Posada, D., and K. A. Crandall.** 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14:**817–818.
32. **Ramírez, M. E., P. B. Hebbar, R. Zhou, C. P. Wolk, and S. E. Curtis.** 2005. *Anabaena* sp. strain PCC 7120 gene *devH* is required for synthesis of the heterocyst glycolipid layer. J. Bacteriol. **187:**2326–2331.
33. **Rosenberg, N. A.** 2003. The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. Evolution **57:**1465–1477.
34. **Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496–2497.
35. **Schierup, M. H., and J. Hein.** 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics **156:**879–891.
36. **Schneider, S., D. Roessli, and L. Excoffier.** 2000. Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland.
37. **Swofford, D. L.** 1998. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4. Sinauer Associates, Sunderland, MA.
38. **Ward, D. M., M. J. Ferris, S. C. Nold, and M. M. Bateson.** 1998. A natural view of microbial diversity within hot spring cyanobacterial mat communities. Microbiol. Mol. Biol. Rev. **62:**1353–1370.
39. **Watterson, G. A.** 1975. On the number of segregating sites in genetical models without recombination. Theor. Pop. Biol. **7:**256–276.
40. **Watterson, G. A., and H. A. Guess.** 1977. Is the most frequent allele the oldest? Theor. Popul. Biol. **11:**141–160.
41. **Whitaker, R. J., D. W. Grogan, and J. W. Taylor.** 2003. Geographic barriers isolate endemic populations of hyperthermophilic Archaea. Science **301:** 976–978.
42. **Wright, S.** 1931. Evolution in Mendelian populations. Genetics **16:**97–159.
43. **Wright, S.** 1943. Isolation by distance. Genetics **28:**114–138.