

Phylogenetic Ecology of the Freshwater *Actinobacteria* acI Lineage^{∇†}

Ryan J. Newton,¹ Stuart E. Jones,² Matthew R. Helmus,³ and Katherine D. McMahon^{4*}

Microbiology Doctoral Training Program, University of Wisconsin—Madison, Madison, Wisconsin 53706¹; Limnology and Marine Sciences, University of Wisconsin—Madison, Madison, Wisconsin 53706²; Department of Zoology, University of Wisconsin—Madison, Madison, Wisconsin 53706³; and Department of Civil and Environmental Engineering, University of Wisconsin—Madison, Madison, Wisconsin 53706⁴

Received 10 April 2007/Accepted 27 August 2007

The acI lineage of freshwater *Actinobacteria* is a cosmopolitan and often numerically dominant member of lake bacterial communities. We conducted a survey of acI 16S rRNA genes and 16S-23S rRNA internal transcribed spacer regions from 18 Wisconsin lakes and used standard nonphylogenetic and phylogenetic statistical approaches to investigate the factors that determine acI community composition at the local scale (within lakes) and at the regional scale (across lakes). Phylogenetic reconstruction of 434 acI 16S rRNA genes revealed a well-defined and highly resolved phylogeny. Eleven previously unrecognized monophyletic clades, each with $\geq 97.9\%$ within-clade 16S rRNA gene sequence identity, were identified. Clade community similarity positively correlated with lake environmental similarity but not with geographic distance, implying that the lakes represent a single biotic region containing environmental filters for communities that have similar compositions. Phylogenetically disparate clades within the acI lineage were most abundant at the regional scale, and local communities were comprised of more closely related clades. Lake pH was a strong predictor of the community composition, but only when lakes with a pH below 6 were included in the data set. In the remaining lakes (pH above 6) biogeographic patterns in the landscape were instead a predictor of the observed acI community structure. The nonrandom distribution of the newly defined acI clades suggests potential ecophysiological differences between the clades, with acI clades AI, BII, and BIII preferring acidic lakes and acI clades AII, AVI, and BI preferring more alkaline lakes.

The phylum *Actinobacteria* contains several of the recognized freshwater cosmopolitan bacterial groups. Within this phylum, the acI lineage is one of the most abundant lineages in freshwater and at times accounts for more than 50% of the total 4',6'-diamidino-2-phenylindole (DAPI)-stained bacterioplankton cells in a lake's epilimnion (1, 37). It is also known that acI is an active and exclusive component of freshwater systems (36). This lineage is globally distributed and occurs in a variety of limnetic systems characterized by extremely different physical, chemical, and biological qualities (21). It has been postulated that the acI lineage's relatively small cell size, which provides protection from protistan grazing, may contribute to its global distribution (19, 30). However, it seems unlikely that small cell size is the single driving force behind acI's remarkable capacity to inhabit such a broad range of freshwater systems. Although the acI lineage is widely distributed and often very abundant, it remains an uncultured group of bacteria; therefore, identification of coherent but phylogenetically narrow clades and the distribution of these clades among lakes with different environmental characteristics are important for understanding acI's unique prevalence in freshwater ecosystems.

Recently, the acI lineage was divided into three more refined clades (36), with a small expansion by Allgaier and Grossart

(1); however, the level of 16S rRNA gene identity within these three clades is relatively low ($\sim 95\%$). The broad clade definitions prevent a robust examination of acI community composition in relation to environmental variation. Coherent responses of bacterial groups to environmental conditions have been documented for groups defined by levels of 16S rRNA gene sequence identity near 98% (35, 42). Thus, finding significant relationships between acI community composition and environmental variation may require finer-scale taxonomic resolution than has been defined previously.

Analyzing communities from a phylogenetic perspective can provide significant insight into the forces that shape community composition at regional and local scales (38). This linkage between ecology and evolutionary analysis can be termed phylogenetic ecology (39). Despite the power of a phylogenetic perspective to explain community structure, few studies have employed this approach to study microbes (15, 26).

Since closely related taxa tend to have similar traits and have similar ecology (4, 13), at local spatial scales environmental factors can select for taxa in a regional pool of dispersing taxa (i.e., the taxon pool) that are closely related, which is termed environmental filtering (13a). Environmental filtering may be an important force that affects the composition of bacterial communities (15). The implicit assumption in this line of reasoning is that closely related taxa respond similarly, in terms of prevalence or abundance, to a particular environmental factor. However, the radiation of a taxon of bacteria with a competitive advantage in one region can result in bacterial communities dominated by closely related taxa, regardless of the environmental conditions at the local-community scale (15). In contrast, physical isolation of bacterial communities can result

* Corresponding author. Mailing address: Department of Civil and Environmental Engineering, University of Wisconsin—Madison, Madison, WI 53706. Phone: (608) 263-3137. Fax: (608) 262-5199. E-mail: tcmahon@engr.wisc.edu.

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

[∇] Published ahead of print on 7 September 2007.

TABLE 1. Lake characteristics

Lake	Location ^a	Landscape ^b	Area (km ²)	Depth (m)	pH ^c	Dissolved organic carbon concn (mg/liter)	Total phosphorus concn (μg/liter)	NH ₄ concn (μg/liter)	NO ₂ /NO ₃ concn (μg/liter)
Crystal Bog (CB)	N	Seepage	0.01	3	5.1	11	33	67	23
Hook (HO)	S	Seepage	0.51	1	5.3	12	17	152	45
Jimi Hendrix Bog (JB)	N	Seepage	0.01	4	6.4	8	9	150	20
Little Trout (LT)	N	Drainage	4.07	20	6.5	6	5	134	21
Firefly (FF)	N	Seepage	0.11	14	6.6	3	4	17	0
Ike Walton (IW)	N	Seepage	5.76	64	6.6	4	4	110	21
Brandy (BR)	N	Drainage	0.45	13	7.0	4	4	16	1
Red Cedar (RC)	S	Seepage	1.45	2	7.1	13	11	57	8
Trout (TR)	N	Drainage	16.08	33	7.3	4	5	45	8
Salmo Pond (SA)	S	Seepage	0.01	5	7.4	2	4	22	426
Little Arbor Vitae (LA)	N	Drainage	2.16	8	8.4	5	11	52	15
Mendota (ME)	S	Drainage	39.38	25	8.6	6	83	0	156
Mirror (MI)	S	Drainage	0.55	4	8.6	3	21	29	542
Green (GR)	S	Drainage	29.73	72	8.8	6	7	9	21
Delton (DE)	S	Drainage	1.08	3	8.8	5	27	36	30
Clear (CL)	S	Seepage	0.33	4	8.9	6	11	0	0
Cox Hollow (CH)	S	Drainage	0.39	6	8.9	4	10	38	254
Twin Valley (TW)	S	Drainage	0.62	3	8.9	4	9	8	2

^a Lakes were placed into two categories: north (N) and south (S). Lakes categorized as north were above 45°N latitude, and lakes categorized as south were below 44°N latitude.

^b Categorized by watershed input.

^c The lakes are in order from the lake with the lowest pH to the lake with the highest pH.

in relatively unrelated communities in a region even though the environmental conditions of systems in the region are very similar (29, 40). Thus, the phylogeographic patterns of taxa must be taken into account in order to assess whether environmental filtering occurs in a set of communities.

At local spatial scales, a community may instead be composed of more distantly related taxa. In local communities, this pattern is consistent with the hypothesis of competitive exclusion (22) or environmental filtering for unrelated taxa with convergently evolved tolerances to particular environmental factors (6). The taxon pool may also be dominated by more distantly related taxa. This pattern is consistent with the hypothesis that the abundant taxa inhabit separate niches (i.e., they are separate ecotypes) and are the survivors of the most recent selective sweep (11).

In this study we sought to significantly improve the current acI phylogeny and to use this phylogeny to identify lake environmental factors that are related to acI community composition. We applied a novel phylogenetic ecology approach to our data set and illustrated the usefulness of taking a phylogenetic perspective when forces that influence bacterial community composition are examined.

MATERIALS AND METHODS

Lake selection and environmental variables. In July 2002 (16 to 31 July), 18 lakes distributed throughout Wisconsin were sampled (20, 34, 43). The lakes were chosen because they varied in inorganic and organic nutrient concentrations, pH, surface area, landscape position (drainage or seepage), and geographic position (latitude and longitude). The water temperature, secchi depth, depth to the bottom, and pH were measured onsite. Triplicate integrated epilimnetic samples were taken from each lake as described by Yannarell and Triplett (43). The water samples were used for measurement of dissolved organic carbon, total phosphorus, ammonia, and nitrite/nitrate contents and for determination of the bacterial community composition (43). Data for the lake characteristics measured in this study are shown in Table 1.

DNA extraction and clone library construction. Total DNA was extracted from stored filters using a FastPrep Spin DNA purification kit (QBiogene) and

the methods described for it. We amplified bacterial 16S rRNA genes and the corresponding 16S-23S internal transcribed spacer (ITS) regions from the resulting DNA pool using a PCR (Eppendorf Mastercycler) with the following conditions: 2 min at 94°C, followed by 20 cycles of 35 s at 94°C, 45 s at 55°C, and 2 min at 72°C and then 2 min at 72°C; and primers 8F (5'-AGAGTTTGATCM TGGCTCAG-3') (bacterium specific, 16S rRNA gene) and 23SR (5'-GGGTT BCCCATTCRG-3') (bacterium specific, 23S rRNA gene). Pooled DNA from each of the three replicates was used in the PCRs. We gel purified the PCR products using standard methods specified in the Novagen SpinPrep gel DNA purification protocol (EMD Biosciences Inc.).

We constructed one clone library for each of the 18 lakes using standard protocols of the Invitrogen Topo TA cloning system with TOP10 chemically competent cells. After cloning, the inserts were amplified directly from cells using the following PCR conditions: 2 min at 94°C, followed by 35 s at 94°C, 45 s at 55°C, and 2 min at 72°C for 20 cycles and then 2 min at 72°C; and vector primers M13 Forward (5'-GTAAAACGACGGCCAG-3') and M13 Reverse (5'-CAGG AACAGCTATGAC-3'). The PCR products were purified using an Agencourt AMPure kit as recommended by the manufacturer's protocol.

Clone library screening. We screened a total of 1,850 clones with 16S rRNA gene-ITS inserts for the presence of lineage acI in the 18 lakes; this corresponded to 93 clones per lake for all lakes except Jimi Hendrix Bog, for which 269 clones were screened in order to obtain enough acI sequences for analysis. We screened the clone libraries for the presence of acI phylotypes with primers HGC664F (5'-GGGGAGACTGGAATTCCT-3') (*Actinobacteria* specific) and ACI856R (5'-TCGCASAAACCGTGGAAG-3') (acI specific) (37). The 194-bp amplicon was labeled with SYBR green 1 and was detected using the software iCycler iQ v3.0A linked to a Bio-Rad iCycler. Each PCR included an internal fluorescein standard, and all 93 reactions for a single lake were run simultaneously in a 96-well plate. A positive control containing a cloned acI 16S rRNA gene, a negative control containing a cloned non-acI 16S rRNA gene with acI-specific primer amplification, and a no-DNA control were run on each plate. The PCR cycling conditions were 5 min at 95°C, followed by 25 cycles of 35 s at 94°C, 40 s at 68°C, and 90 s at 72°C and then 4 min at 72°C. All clones that showed amplification a cycle earlier than the negative control were considered acI positive and were used for further analysis by sequencing.

Sequence analysis and phylogenetic reconstruction. We sequenced candidate acI clones using an ABI Prism BigDye terminator sequencing kit (PE Applied Biosystems) and standard PCR sequencing conditions. Following PCR sequencing, we purified the products with an Agencourt CleanSeq kit. Three sequencing passes using the M13 Forward, M13 Reverse, and 515 Forward (5'-GTGCCAG CMGCCGCGTAA-3') primers were required to complete a nearly full-length 16S rRNA and ITS sequence for each cloned amplicon. We assembled and

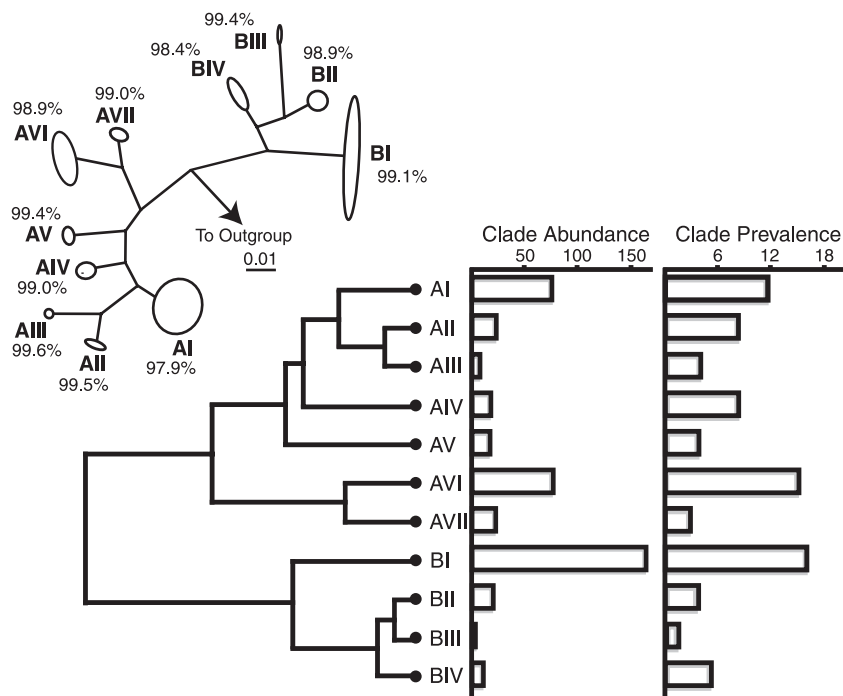


FIG. 1. (Upper left panel) Unrooted consensus radial phylogram showing 11 newly defined *Actinobacteria* *aci* clades based on nearly full-length (>1,500-bp) 16S rRNA gene sequences. The clade designation and minimum within-clade 16S rRNA gene sequence identity are indicated next to each clade oval. Scale bar = 1% sequence divergence. (Bottom left panel) Unrooted consensus contemporaneous tipped phylogram used in PSV and PSE analyses. Relationships in both phylograms were determined by Bayesian analysis (MrBayes 3.1.2 software) with a general time-reversible gamma-distributed rate variation model. (Right panel) Representative bar graphs indicating clade clone abundance and clade prevalence in 18 lakes.

heuristically adjusted the 16S rRNA gene amplicons using the Staden v1.6.0 package (5). All sequences were analyzed in the context of the complete data set using Bellerophon (16) and Mallard (3) to identify putative chimeras. Suspected chimeras were additionally checked with the program Pintail (2), which allows pairwise comparisons of sequences and site-specific identification of anomalies. After removal of seven putative chimeras, 434 nearly full-length 16S rRNA gene sequences were included in all subsequent analyses.

We aligned the 16S rRNA gene sequences using the ARB software package (24) containing a publicly available 16S rRNA gene ARB database (accessed January 2002) (17) supplemented with freshwater *Actinobacteria* 16S rRNA gene sequences (28, 36). Sequences were initially aligned using the FAST_ALIGNER ARB tool before the alignment was heuristically adjusted using primary and secondary rRNA structure as a guide.

All unique 16S rRNA gene sequences (350 sequences) were used for phylogenetic tree reconstruction with the MrBayes v. 3.0 software program (31). A general time-reversible gamma-distributed rate variation model was specified. Three independent Markov chain Monte Carlo analyses, each starting with random trees for each of four simultaneous chains, were performed for 150,000 to 350,000 generations with sampling every 100 generations to create a posterior probability distribution between 1,500 and 3,500 trees. Trees recovered before chain stabilization were discarded with appropriate burn-in values, and a 50% majority rule tree was calculated. Clades were defined both to minimize within-group sequence divergence based on Dotur furthest-neighbor comparisons and to adhere to consistently identified monophyletic branches with strong (>75%) posterior probability support from multiple tree reconstructions run in MrBayes. The resultant collapsed tree was constructed with representative branch lengths and separately with contemporaneous tips (Fig. 1). Similarly, two other sets of sequences were included in Bayesian phylogenetic tree reconstruction; one set was from 143 clones that represented all lakes and the majority of ITS lengths present in each clade (see the supplemental material), and the other set was from 89 clones where each clone represented a unique lake-clade combination (used in principal coordinate analysis [PCoA] analysis).

All *aci* 16S rRNA gene sequences obtained in this study were grouped into defined clades based on phylogenetic reconstruction as described above (Fig. 1). All 18 libraries were sampled to near saturation for the *aci* lineage based on the

average of the Ace, Chao, and Boot statistics (33) at 99% gene sequence identity but not at more phylogenetically refined levels (see the supplemental material).

PCR and cloning bias. We acknowledge that biases are associated with the use of relativized clone occurrences, which is a method dependent upon PCR- and cloning-based techniques that may cause artificial overrepresentation of some DNA templates. However, the amplified 16S rRNA gene and 16S-23S ITS region of all members of the *aci* lineage were very similar in terms of length, G+C content, and overall sequence conservation, which suggests that biases involved in the molecular methodology should have been consistent for all members. To test this hypothesis, we constructed a DNA template mixture containing equal numbers of copies of clone amplicons from 4 of the 11 identified clades. Two of the clades used in this constructed template mixture, AII and AIII, were among the least frequently recovered clades in our 18 clone libraries, while the other two clades, AVI and BI, were the most frequently recovered clades in our libraries (Fig. 1). This DNA mixture was used as a template for PCR amplification with the 8F and 23SR primers (see above for the conditions used). The amplified products were then cloned and sequenced using the M13 Reverse primer described above and were identified based on the sequence. If there was no bias in our PCR amplification and cloning process, then each of the four clades would have been recovered as 25% of the total clones sequenced.

Nonphylogenetic approaches to explain community composition. We tested if community similarity (11 defined clades) and environmental similarity (18 lakes) were correlated and if there was spatial autocorrelation for lake community similarity. If there was no spatial pattern for community similarity, then the lakes may be considered part of a single biotic region, while a correlation between community similarity and environmental similarity would suggest that the lake environment affects bacterial community composition. We used partial Mantel tests (25) to correlate the Bray-Curtis community dissimilarity matrix to a matrix of pairwise lake environmental Euclidian distances (calculated from z-scored environmental data) and to a matrix of spatial proximity (distance in kilometers between pairs of lakes). Correlation significance using an α value of 0.05, Bonferroni corrected for multiple comparisons, was determined through permutation of the rows and columns in one of the matrices included in the partial Mantel tests (1,000 permutations). Partial Mantel tests were implemented in the

R statistics environment using the vegan packages (R Development Core Team, 2005).

Phylogenetic approaches to explain community composition. We examined the possibility that individual environmental variables may be related to the phylogeny and community composition within lakes by performing PCoA with the UniFrac web interface (23), using a reconstructed tree containing sequences representing all unique lake-clade combinations, weighted by the occurrences of each sequence group. Using the sample scores of the first and second axes from the PCoA for each of our lakes, we calculated the Spearman's rank correlation coefficients (ρ) between scores and the measured environmental variables as an indication of a relationship between the phylogenetic community compositions of the lakes and their environmental parameters.

We summarized the relatedness of clades in lake acI communities with two phylogenetic biodiversity metrics and then tested for phylogenetic community patterns in community composition with permutation tests (13a). We first calculated the phylogenetic species variability (PSV) (13a) of each lake community. When all clades of a community are phylogenetically unrelated (i.e., the community phylogeny is a star), the PSV is 1; as phylogenetic relatedness increases among clades, PSV approaches 0. Thus, PSV is basically a bounded summary of the average degree to which individual clades are related in a community. The phylogenetic species evenness (PSE) (13a) of each community was also calculated. PSE is PSV modified to incorporate clade abundances. PSE is 1 when the community phylogeny is a star and the abundances of all clades are equal, and PSE approaches 0 as clades become more related to each other and/or less even in terms of species abundance. The PSE of a community is equal to the community's PSV if all clades have the same abundance. The percentage of each clade represented in each clone library (constructed from a single lake) is used as a proxy for clade abundance. A full description of the statistical properties of these metrics has been published previously (13a).

We statistically compared the mean observed PSV (\overline{PSV}_{obs}) and observed PSE (\overline{PSE}_{obs}) values across lake communities to distributions of mean null values produced using two different permutation procedures run 2,000 times each for the observed data. This allowed us to test whether communities are composed of clades that are more or less related to each other than expected. For simplicity, here we describe the two null models only in the context of PSV. For a presence/absence matrix with lakes as rows and clades as columns, null model 1 shuffled cells along rows, while null model 2 shuffled cells along columns. Null model 1 tested the hypothesis that \overline{PSV}_{obs} equaled \overline{PSV}_{pool} (i.e., the PSV value of a community containing all clades) and was rejected if there was a phylogenetic pattern in the prevalence of clades in the clade pool (e.g., there was a group of related clades that was very prevalent in all communities regardless of environmental conditions). Null model 2 maintained clade prevalence during permutation (i.e., the number of lakes in which each clade was found) and tested for phylogenetic structure within local communities independent of any structure caused by species prevalence (for more discussion, see reference 13a). For example, if environmental filtering for closely related species occurred, \overline{PSV}_{obs} fell significantly below the permutation distribution of null model 2. Correlations between individual lake PSV and PSE values and lake characteristics were examined using linear, second-order polynomial, and exponential regressions with a Bonferroni correction for multiple comparisons. We used PSV and PSE, as opposed to other metrics (e.g., those described by Webb et al. [38]), because of the advantageous statistical properties that these metrics have when they are used in particular null model tests for phylogenetic structure (13a). Basically, the null models that we used tested specific hypotheses about the causes of phylogenetic structure when they were used with PSV and PSE.

Communities are aggregations of taxa, and if there is a phylogenetic signal in community composition, it can be hypothesized that there should be a phylogenetic signal in the way that each clade responds to environmental variation. We fit standard linear regression models of clade relative abundance (square root transformed) to the measured lake characteristics. The coefficients can be interpreted as the relationship of each clade to changes in the characteristics across lakes. Since pH was the one measured lake characteristic significantly related to our phylogeny and community composition, we plotted pH linear coefficient values aligned with the acI clade phylogeny to obtain a visual description of how each clade is related to pH and to see if closely related species exhibited similar pH-related distribution patterns (Fig. 2). In order to quantify the amount of phylogenetic signal in each clade's pH regression coefficients, we used the procedure of Helmus et al. (14). We first calculated the K^* statistic with measurement error (18), which quantifies the amount of phylogenetic signal contained within a set of species traits. Second, for the regression coefficients we fit the model $b = \beta I + \epsilon + \eta$, where β is the expected value of the pH regression coefficient for b , I is an 11×1 vector of ones, ϵ is a vector of error terms, and η is the vector of estimated errors of the coefficients (i.e., measurement error).

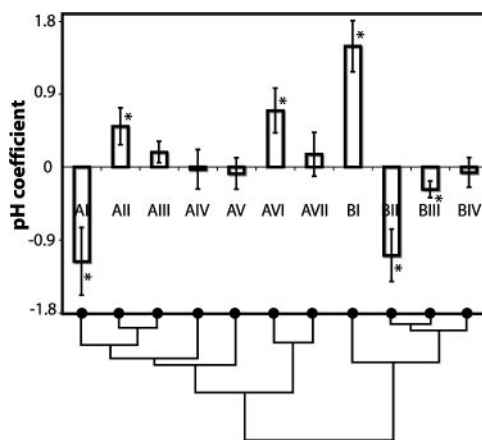


FIG. 2. pH coefficient value and standard error for each clade and clade phylogram. The pH coefficient values were calculated by determining the slope of a regression between the relative abundance of each clade in each lake and lake pH. Clades indicated by an asterisk are significant ($P \leq 0.05$).

If closely related clades respond in similar ways to pH across lakes, the pattern of correlation in values of ϵ reflects phylogenetic relatedness. If there is no phylogenetic signal, values of b are independent among clades. We conducted a statistical test for a phylogenetic signal by comparing the fit of this pH coefficient model using the correlation matrix of our phylogeny to the fit when it was assumed that there was no phylogenetic signal. We fit the two models using restricted maximum likelihood (REML), so the best-fit model of the two is the model with the higher REML log likelihood (for details, see references 14 and 18).

Nucleotide sequence accession numbers. Representative 16S rRNA gene and 16S-23S rRNA ITS sequences generated in the current study that were included in the phylogenetic analyses have been submitted to the GenBank database (accession numbers EU117556 to EU117989).

RESULTS

acI screening and sequencing. Screening the 18 constructed clone libraries revealed 560 clones with an acI affiliation among the 1,850 clones examined (~30%). The 16S rRNA gene of 434 acI-affiliated clones and the 16S-23S ITS of 411 clones were sequenced. Individual libraries exhibited a wide range of percentages of screened clones containing an acI sequence. Little Trout Lake contained the most acI-related clones (59% of the total clones), while Jimi Hendrix Bog Lake contained the least (5% of the total clones).

16S rRNA gene phylogenetic inference and distribution. The 434 16S rRNA gene sequences grouped into 11 newly defined clades (Fig. 1) within the previously defined acI clades A and B (36). The acI-A clade contained 237 of the 434 16S rRNA gene sequences, and nearly two-thirds of these sequences clustered into two newly defined clades, AI and AVI (Fig. 1). The acI-B clade contained 197 of the 434 sequences, the majority of which (165 sequences) grouped into the newly defined clade BI. These three clades contained nearly 75% of all sequenced acI clones. In addition, these three clades were the most prevalent across the Wisconsin lake landscape (Fig. 1). The 11 clades all had a within-clade 16S rRNA gene sequence identity of $\geq 97.9\%$, while the mean minimum sequence identity within the 11 clades was 99.1% (Fig. 1).

The BI clade, based on clone recovery, was the most abundant clade in 11 of the 18 lake clone libraries, while the AI

TABLE 2. Clone recovery by lake^a

Lake ^b	PSV	PSE	n ^c	% of clones in clade:											
				AI	AII	AIII	AIV	AV	AVI	AVII	BI	BII	BIII	BIV	
Crystal Bog	0.07	0.02	10										90	10	
Hook	0.79	0.32	17	88								6	6		
Jimi Hendrix Bog	0.78	0.58	11	42				25		17		8			8
Little Trout	0.60	0.72	53	11	6	2	2	17	15	9	38				
Firefly	0.72	0.72	38	29	3		3		26		26	8	3	3	
Ike Walton	0.55	0.65	23	61									22	4	13
Brandy	0.74	0.62	29	7	3		3			21		55	3		7
Red Cedar	0.77	0.80	18	22					6	28		44			
Trout	0.63	0.71	40	8			3	5	10	13	28	35			
Salmo Pond	0.64	0.52	27	37	7			15	4	30		7			
Little Arbor Vitae	0.66	0.64	29		7			3		17		63			
Mendota	0.76	0.76	20				5	10		35		50			
Mirror	0.76	0.76	23	13	22					39		26			
Green	0.74	0.73	20					10		15	20	55			
Delton	0.70	0.76	11	9	18					27	9	36			
Clear	0.67	0.48	17						6	6		76			
Cox Hollow	0.83	0.63	28					18		11		64			7
Twin Valley	0.83	0.65	20	11						21		63			5

^a Clone recovery was calculated for each lake by dividing the number of clones in each clade from the lake by the total number of acI clones recovered from the lake.
^b The lakes are in order from the lake with the lowest pH to the lake with the highest pH.
^c Number of sequenced clones from each lake. Ninety-three clones were screened for each lake except Jimi Hendrix Bog Lake, for which 269 clones were screened.

clade was the most abundant clade in 5 of the 18 lake clone libraries (Table 2). No clade was recovered from all 18 lakes, but clade BI was identified in 16 of the 18 lakes and clade AVI was detected in 15 of the 18 lake libraries (Table 2). Several of the clades were recovered from relatively few lakes; clade BIII was identified in only three lake libraries, and all three lakes had pH values below 6.6 (Table 2).

We identified 172 sequences from our constructed acI “communities” designed to evaluate the PCR-clone library process. The results suggested that the PCR and cloning bias between different acI clades was limited, as AII, AIII, AVI, and BI accounted for 23, 21, 26, and 30% of the constructed library, respectively, which is not significantly different from the expected 25% for each clade ($P \geq 0.05$, Student’s *t* test). Therefore, we used and reported relative clone numbers in analyses as an indication of relative abundance in lake samples.

ITS comparison. The acI 16S-23S rRNA ITS sequences do not contain obvious homologs of any known tRNAs, as is often seen in this region. The ITS length was relatively conserved for all 418 sequenced acI clones. The shortest ITS was 230 bp long, while the longest ITS was 335 bp long (see the supplemental material). The within-clade ITS length was even more conserved (see the supplemental material). Clade BI contained ITS regions with lengths that varied from 230 to 277 bp, which was a very distinct size range compared to the lengths for the other clades, which generally ranged from 287 to 335 bp. Clades often contained ITS regions with one or two dominant lengths, and the more significant length variation resulted from a few clones with distinct lengths (data not shown).

Nonphylogenetic approaches to explain community composition. While there was a strong correlation between acI clade community similarity and environmental similarity (Mantel $R = 0.48$, $P \leq 0.01$), community similarity was not related to the geographic distance among lakes (Mantel $R = -0.09$, $P \geq 0.01$). There was also no covariance between geographic and environmental distance (Mantel $R = -0.10$, $P \geq 0.01$).

Phylogenetic approaches to explain community composition. PCoA with the UniFrac web interface (23) revealed a strong relationship between community composition and lake characteristics. The first two principal coordinate axes summarized ca. 82% of the variation in acI community composition (Fig. 3), and lake pH was the strongest and only significant correlate to the first axis ($\rho = 0.63$, Spearman’s rank). Ammonia and pH were also significantly correlate to the second axis ($\rho = -0.52$ and $\rho = 0.48$, respectively, Spearman’s rank).

The \overline{PSV}_{obs} was less than, but not significantly less than, the mean for null model 1 (\overline{PSV}_{null1}); however, \overline{PSV}_{obs} was significantly less than \overline{PSV}_{null2} (Table 3). \overline{PSE}_{obs} was significantly

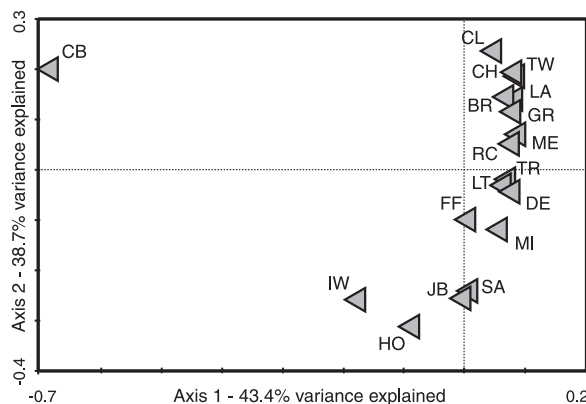


FIG. 3. acI UniFrac PCoA ordination. The ordination was constructed using UniFrac distances with weighting by the relative abundances of each unique lake-clade group (see Materials and Methods). Each triangle represents the acI phylogenetic community composition of a single lake. Of the measured environmental variables, pH was the only variable significantly correlated with the first axis ($\rho = 0.63$, Spearman’s rank), while NH_4 and pH were significantly correlated with the second axis ($\rho = 0.52$ and $\rho = 0.48$, Spearman’s rank). Lake abbreviations are defined in Table 1.

TABLE 3. PSV and PSE

Data set	Model	PSV ^a	PSE ^a
Full	Observed	0.68	0.61
	Null 1 ^c	0.71	0.54
	Null 2 ^c	<u>0.74</u>	<u>0.66^d</u>
Low pH ^b	Observed	0.71	0.67
	Null 1	0.71	0.58
	Null 2	0.73	0.68

^a Significance was assessed at α values of 0.05 and 0.10. Significant results for observed means above the null mean distribution are indicated by bold type, and significant results for observed means below the null distribution are underlined.

^b Analyses were performed with the two lakes with the lowest pH values removed (Crystal Bog and Hook Lake).

^c Mean value for null model 1 or null model 2 randomizations.

^d The test was significant at an α value of 0.10.

greater than $\overline{\text{PSE}}_{\text{null1}}$ and marginally less than $\overline{\text{PSE}}_{\text{null2}}$ ($P \leq 0.1$) (Table 3). The null model 1 tests suggested that there was a significant phylogeographic pattern in clade abundance, with the most divergent clades within the acI lineage being most abundant in the landscape (Fig. 1). On the other hand, the null model 2 tests suggested that regardless of the observed phylogeographic patterns for acI abundance and prevalence, on average communities contained more closely related clades than expected from a random construction of communities, a pattern consistent with environmental filtering for closely related clades.

Of the lake characteristics measured, lake pH was by far the best predictor of phylogenetic community composition (second-order polynomial fit for PSV, $r^2 = 0.35$ and $P \leq 0.05$; second-order polynomial fit for PSE, $r^2 = 0.69$ and $P \leq 0.01$). However, this relationship was dependent on sequences recovered from the two lakes with the lowest pH values, Crystal Bog (pH 5.1) and Hook Lake (pH 5.3). With the data sets for these two lakes removed, the polynomial regression fits decreased to $r^2 = 0.22$ and $r^2 = 0.01$ for PSV and PSE ($P \geq 0.05$), respectively. Similarly, with the Crystal Bog and Hook Lake data sets removed, null model 2 was not rejected for either PSV or PSE (Table 3); thus, there was no evidence for environmental filtering in the remaining 16-lake data set. Although Crystal Bog and Hook Lake had similar pH values, the phylogenetic compositions of these two lakes were quite different, with the clades in Crystal Bog being much more closely related than the clades in Hook Lake (Table 2). The PSV values for these two lakes differed greatly; Crystal Bog had the lowest PSV value (0.07, 3.6 standard deviations below $\overline{\text{PSV}}_{\text{obs}}$), and Hook Lake had the second highest PSV value (0.79, 0.6 standard deviation above $\overline{\text{PSV}}_{\text{obs}}$). In contrast, Crystal Bog (0.02) and Hook Lake (0.32) had the lowest PSE values in the entire data set (Table 3).

We examined the distribution of individual acI clades among lakes as a function of pH. The relative abundance of 6 of the 11 acI clades was significantly related to pH (i.e., the pH coefficient was significantly different from zero [$\alpha = 0.05$] [Fig. 2]). The relative abundance of three of these clades, AI, BII, and BIII, increased with decreasing pH, and the relative abundance of three different clades, AII, AVI, and BI, increased with increasing pH (Fig. 2). The K^* value for the pH regression coefficients (0.57) was slightly less than 1, indicating that there

was less phylogenetic signal in these data than expected by a Brownian motion model of trait evolution across our phylogeny (14). This was expected since there were some divergent clades with similar coefficient values (e.g., AI and BII) (Fig. 2). However, while the signal may not have been strong, it was significant since the pH coefficient model with phylogenetic correlation generated a better fit than the model without correlation (model with phylogenetic correlation, REML log likelihood = -8.7699 ; model without phylogenetic correlation, REML log likelihood = -8.7996).

DISCUSSION

Increased phylogenetic resolution for the acI lineage. The acI lineage of freshwater *Actinobacteria* is one of the most ubiquitous and abundant groups of lake bacterioplankton (28, 37). Our results support this conclusion, since all 18 lakes studied contained acI and more than 30% of the clones recovered were acI affiliated. Despite this ubiquity, the number of publicly available 16S rRNA gene sequences for acI-affiliated organisms was relatively low (97 sequences that were $\geq 1,200$ bp long in the RDP database on 28 January 2007), which hindered our ability to examine the factors controlling the distribution and community composition of these important organisms among and within lakes. Examination of these controlling factors was also potentially limited by the phylogenetically broad nature of the monophyletic acI lineage ($\sim 93\%$ 16S rRNA gene identity) and its more phylogenetically refined clades, acI-A, acI-B, and acI-C ($\sim 95\%$ 16S rRNA gene identity within each clade). Our retrieval of more than four times the current number of nearly full-length acI 16S rRNA gene sequences and statistical evidence for sequence saturation at 99% 16S rRNA gene identity provide a much-needed framework for future study of this cosmopolitan freshwater lineage.

The dependence of perceived patterns on the taxonomic scale used during community analyses can influence a researcher's ecological conclusions. In a study of a north central Florida plant community, increasing the taxonomic scale from a single genus to all plants shifted the community phylogeny pattern from overdispersion resulting from species interactions to underdispersion resulting from environmental selection (7). Likewise, different clades with $\sim 98\%$ 16S rRNA gene identity within the freshwater *Polynucleobacter* cluster ($\sim 95\%$ 16S rRNA gene identity) showed markedly distinct dynamics and habitat ranges (41). Using only broadly defined operational taxonomic units to examine "populations" within a larger community severely reduces the ability to link bacterial dynamics to their environmental drivers and may result in a completely different relationship than would be obtained with a more refined phylogenetic view (12, 32). Our current acI phylogenetic framework is comprised of 11 monophyletic clades with $\geq 97.9\%$ 16S rRNA gene sequence identity (Fig. 1). We treat these clades as species or ecotypes and understand that the still somewhat broad definition of these clades may hide some evidence for forces shaping acI communities. Additional effort to more thoroughly sample populations at finer levels of phylogenetic resolution is required to determine if the observed phylogenetic patterns do not change across operational taxonomic unit definitions. However, defining these 11 clades revealed new potential relationships between environmental

drivers (lake pH), community composition, and clade distribution (Fig. 2 and Table 3) for this lineage.

Biogeographic and phylogeographic patterns of acI community composition. Physical isolation as a force shaping community structure has been explored for freshwater bacterial communities (for a review, see reference 9). In one of the studies, Yannarell and Triplett (43) found a significant difference in the bacterial community similarity between northern and southern Wisconsin lakes (minimum distance between regions, 150 km). This result suggests that the lakes sampled by Yannarell and Triplett were not in a single biogeographic region. We show that there was no effect of geographic distance (maximum distance, 362 km; minimum distance, 2.4 km) on community similarity even though our 18 lakes were sampled in the study of Yannarell and Triplett. This apparent contradiction is most likely due to the fact that we focused on a single bacterial lineage, while Yannarell and Triplett surveyed complete bacterial communities defined by automatic ribosomal intergenic spacer analysis. Our lakes may therefore be thought of as a single clade pool or as part of a single biotic region (27), in which it is assumed that all lakes have an equal chance of containing any of the acI clades. This is especially relevant because one implicit assumption of our community analyses is that any relationships found are not simply driven by unrelated spatial autocorrelation. Furthermore, the lack of distance-based biogeographical patterns for acI, despite the patterns present in the whole bacterial community from the lakes (as assessed by automatic ribosomal intergenic spacer analysis [43]), implies that different freshwater bacterial clades have different landscape distribution patterns or dispersion capabilities.

Our phylogenetic analyses of community composition revealed a phylogeographic pattern within the clade pool, with the most divergent clades being the most abundant clades in the clone libraries. This pattern was likely produced by the phylogenetically disparate AI, AVI, and BI clades, all of which were highly represented in the data set (Fig. 1). This pattern is consistent with the hypothesis that these three clades either have greater dispersal abilities than the other clades and that dispersal ability is a trait that does not show phylogenetic conservatism or that these three clades represent the surviving clades of the most recent selective sweep in the acI lineage. However, at this early stage in our understanding of forces structuring acI communities we can only speculate. Additional efforts to assess the prevalence of each clade in potential source communities, such as air or rain, groundwater, and surface flow and examination of predictive ecotype models, such as those developed by Cohan and Perry (8), are needed to test these possibilities.

Evidence for pH as an environmental filter. Two recent studies on the filamentous SOL clade of *Bacteroidetes* (32) and on the *Polynucleobacter* clade of *Betaproteobacteria* (42) identified water chemistry parameters as potential drivers for phenotypic differentiation among the corresponding cosmopolitan freshwater bacteria. Fierer and Jackson (10) also presented evidence that pH is a predictor of bacterial community composition in soils. Likewise, our analyses revealed a significant correlation between community similarity and environmental similarity, and the main environmental factor that explained this correlation was lake pH. Lake pH was also significantly

related to both of our phylogenetic relatedness metrics and indicated that more acidic lakes tend to contain more closely related acI clades. Similarly, our acI communities had significantly less PSV and marginally less PSE than expected, which was a further indication that environmental filtering is a predominant force structuring the acI community phylogeny (Table 3, null model 2) and adds another community to the growing list of communities (7, 15) whose phylogenies are influenced by environmental filtering.

The filtering effect of pH in our study was apparent only when the two lakes with the lowest pH values, Crystal Bog and Hook Lake, were in the data set. This is similar to what Helmus et al. (13a) showed for fish communities in the same geographic region as our lakes. Lakes with the lowest pH values also appeared to drive this filtering effect (13a). Therefore, it is possible that lake pH is a common filter for a variety of taxonomic groups, and future work on bacteria and/or other aquatic groups should focus on obtaining data from more low-pH lakes.

Although Crystal Bog and Hook Lake are both low-pH lakes and together drove the filtering effect in our data, the acI clade compositions of these lakes were very different (Table 2). Crystal Bog contained only clades that were highly related (BII and BIII), while Hook Lake contained quite divergent clades (AI, BI, and BII). These two lakes illustrate that even though sister acI clades tend to have similar preferences for acidity (Fig. 2), this trait is not phylogenetically conserved deep in the phylogeny. Instead, the distribution of pH coefficients across the phylogeny (Fig. 2), with clades AI, BII, and BIII strongly associated with acidic lakes and clades AII, AVI, and BI strongly associated with alkaline lakes, suggests that there have been instances of either convergent evolution or genetic exchange among clades and is an indication that these clades are distinct ecotypes.

Conclusions. In this study we showed that the acI lineage is comprised of significantly more monophyletic clades (11 clades) than had previously been identified. Three of these clades, AI, AVI, and BI, accounted for >70% of the clones from the 18 clone libraries. Using a combination of community composition analyses and novel community phylogeny techniques, we found evidence that both phylogeographic patterns in the landscape and environmental filtering by lake pH, but not physical isolation or current competition within lakes, contributed to the acI community structure. Although environmental filtering was evident in the data set, it played a large role only in the community structure of lakes with pH values below 6, with clades AI, BII, and BIII preferring more acidic lakes. The remaining lakes (with pH values above 6) contained acI communities with uneven clade distribution patterns in the landscape, where three clades, AII, AVI, and BI, preferred more alkaline lakes. These phylogeographic patterns are consistent with the hypothesis that clade radiations from basal survivors of the last selective sweep or clade dispersal capability differences play a large role in shaping the acI community in Wisconsin. This study is an example of how we can statistically relate phylogenies to environmental data and begin assigning ecological significance to bacterial clades when the actual clade traits are not known. With the improved understanding of the acI lineage composition, we believe that it is now possible to

target specific acI populations and understand their dynamics and contributions to processes that occur in lake ecosystems.

ACKNOWLEDGMENTS

We thank Anthony Yannarell for collecting the samples for this study, Jacqueline Minasso for initiating the cloning and sequencing process, and Aaron Jones for providing valuable technical assistance. We also thank all members of the McMahon lab for helpful discussions and two anonymous reviewers for their suggestions.

This research was supported by National Institutes of Health Biotechnology Training Program Grant 5 T32 G08349 (R.J.N.) and by the University of Wisconsin—Madison Graduate School (K.D.M.).

REFERENCES

- Allgaier, M., and H. P. Grossart. 2006. Diversity and seasonal dynamics of *Actinobacteria* populations in four lakes in northeastern Germany. *Appl. Environ. Microbiol.* **72**:3489–3497.
- Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.* **71**:7724–7736.
- Ashelford, K. E., N. A. Chuzhanova, J. C. Fry, A. J. Jones, and A. J. Weightman. 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* **72**:5734–5741.
- Blomberg, S. P., T. Garland, and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**:717–745.
- Bonfield, J. K., K. F. Smith, and R. Staden. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**:4992–4999.
- Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz. 2004. Phylogenetic overdispersion in Floridian oak communities. *Am. Nat.* **163**:823–843.
- Cavender-Bares, J., A. Keen, and B. Miles. 2006. Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* **87**:S109–S122.
- Cohan, F. M., and E. B. Perry. 2007. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* **17**:R373–R386.
- Dolan, J. R. 2005. An introduction to the biogeography of aquatic microbes. *Aquat. Microb. Ecol.* **41**:39–48.
- Fierer, N., and R. B. Jackson. 2006. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **103**:626–631.
- Gevers, D., F. M. Cohan, J. G. Lawrence, B. G. Spratt, T. Coenye, E. J. Feil, E. Stackebrandt, Y. Van de Peer, P. Vandamme, F. L. Thompson, and J. Swings. 2005. Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* **3**:733–739.
- Hahn, M. W., M. Pockl, and Q. L. L. Wu. 2005. Low intraspecific diversity in a *Polynucleobacter* subcluster population numerically dominating bacterioplankton of a freshwater pond. *Appl. Environ. Microbiol.* **71**:4539–4547.
- Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford, United Kingdom.
- Helmus, M. R., T. J. Bland, C. K. Williams, and A. R. Ives. 2007. Phylogenetic measures of biodiversity. *Am. Nat.* **169**:E68–E83.
- Helmus, M. R., K. Savage, M. W. Diebel, J. T. Maxted, and A. R. Ives. 2007. Separating the determinants of phylogenetic community structure. *Ecol. Lett.* **10**:917–925.
- Hornor-Devine, M. C., and B. J. M. Bohannan. 2006. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**:S100–S108.
- Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**:2317–2319.
- Hugenholtz, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**:1–8.
- Ives, A. R., P. E. Midford, and T. Garland. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* **56**:252–270.
- Jezbera, J., K. Hornak, and K. Simek. 2005. Food selection by bacterivorous protists: insight from the analysis of the food vacuole content by means of fluorescence in situ hybridization. *FEMS Microbiol. Ecol.* **52**:351–363.
- Kent, A. D., S. E. Jones, G. H. Lauster, J. M. Graham, R. J. Newton, and K. D. McMahon. 2006. Experimental manipulations of microbial food web interactions in a humic lake: shifting biological drivers of bacterial community structure. *Environ. Microbiol.* **8**:1448–1459.
- Lindstrom, E. S., M. P. Kamst-Van Agterveld, and G. Zwart. 2005. Distribution of typical freshwater bacterial groups is associated with pH, temperature, and lake water retention time. *Appl. Environ. Microbiol.* **71**:8201–8206.
- Lovette, I. J., and W. M. Hochachka. 2006. Simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. *Ecology* **87**:S14–S28.
- Lozupone, C., M. Hamady, and R. Knight. 2006. UniFrac—an online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**:371.
- Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Villbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
- Mantel, N. 1967. The detection of disease clustering and a general regression approach. *Cancer Res.* **27**:209–220.
- Martin, A. P. 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* **68**:3673–3682.
- 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. Ovreas, A. L. Reysenbach, V. H. Smith, and J. T. Staley. 2006. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**:102–112.
- Newton, R. J., A. D. Kent, E. W. Triplett, and K. D. McMahon. 2006. Microbial community dynamics in a humic lake: differential persistence of common freshwater phylotypes. *Environ. Microbiol.* **8**:956–970.
- Papke, R. T., N. B. Ramsing, M. M. Bateson, and D. M. Ward. 2003. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**:650–659.
- Pernthaler, J., T. Posch, K. Simek, J. Vrba, A. Pernthaler, F. O. Glockner, U. Nubel, R. Psenner, and R. Amann. 2001. Predator-specific enrichment of actinobacteria from a cosmopolitan freshwater clade in mixed continuous culture. *Appl. Environ. Microbiol.* **67**:2145–2155.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
- Schauer, M., C. Kamenik, and M. W. Hahn. 2005. Ecological differentiation within a cosmopolitan group of planktonic freshwater bacteria (SOL cluster, *Saprospiraceae*, *Bacteroidetes*). *Appl. Environ. Microbiol.* **71**:5900–5907.
- 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.
- Shade, A., A. D. Kent, S. E. Jones, R. J. Newton, E. W. Triplett, and K. D. McMahon. 2007. Inter-annual dynamics and phenology of bacterial communities in a eutrophic lake. *Limnol. Oceanogr.* **52**:487–494.
- Thompson, J. R., M. A. Randa, L. A. Marcelino, A. Tomita-Mitchell, E. Lim, and M. F. Polz. 2004. Diversity and dynamics of a North Atlantic coastal *Vibrio* community. *Appl. Environ. Microbiol.* **70**:4103–4110.
- Warnecke, F., R. Amann, and J. Pernthaler. 2004. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environ. Microbiol.* **6**:242–253.
- Warnecke, F., R. Sommaruga, R. Sekar, J. S. Hofer, and J. Pernthaler. 2005. Abundances, identity, and growth state of actinobacteria in mountain lakes of different UV transparency. *Appl. Environ. Microbiol.* **71**:5551–5559.
- Webb, C. O., D. D. Ackerly, M. A. McPeck, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annu. Rev. Ecol. Syst.* **33**:475–505.
- Westoby, M. 2006. Phylogenetic ecology at world scale, a new fusion between ecology and evolution. *Ecology* **87**:S163–S166.
- Whitaker, R. J., D. W. Grogan, and J. W. Taylor. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301**:976–978.
- Wu, Q. L., and M. W. Hahn. 2006. Differences in structure and dynamics of *Polynucleobacter* communities in a temperate and a subtropical lake, revealed at three phylogenetic levels. *FEMS Microbiol. Ecol.* **57**:67–79.
- Wu, Q. L., and M. W. Hahn. 2006. High predictability of the seasonal dynamics of a species-like *Polynucleobacter* population in a freshwater lake. *Environ. Microbiol.* **8**:1660–1666.
- Yannarell, A. C., and E. W. Triplett. 2005. Geographic and environmental sources of variation in lake bacterial community composition. *Appl. Environ. Microbiol.* **71**:227–239.