 doi:10.1111/j.1462-2920.2008.01584.x

Volcanic calderas delineate biogeographic provinces among Yellowstone thermophiles

Cristina Takacs-Vesbach,[1]* Kendra Mitchell,[1] Olan Jackson-Weaver[1] and Anna-Louise Reysenbach[2]
[1]*Department of Biology, University of New Mexico, MSC03 2020, 1 UNM, Albuquerque, NM 87131, USA.*
[2]*Department of Biology, Portland State University, PO Box 751 Portland, OR 97207, USA.*

Summary

It has been suggested that the distribution of microorganisms should be cosmopolitan because of their enormous capacity for dispersal. However, recent studies have revealed that geographically isolated microbial populations do exist. Geographic distance as a barrier to dispersal is most often invoked to explain these distributions. Here we show that unique and diverse sequences of the bacterial genus *Sulfurihydrogenibium* exist in Yellowstone thermal springs, indicating that these sites are geographically isolated. Although there was no correlation with geographic distance or the associated geochemistry of the springs, there was a strong historical signal. We found that the Yellowstone calderas, remnants of prehistoric volcanic eruptions, delineate biogeographical provinces for the *Sulfurihydrogenibium* within Yellowstone (χ^2: 9.7, $P = 0.002$). The pattern of distribution that we have detected suggests that major geological events in the past 2 million years explain more of the variation in sequence diversity in this system than do contemporary factors such as habitat or geographic distance. These findings highlight the importance of historical legacies in determining contemporary microbial distributions and suggest that the same factors that determine the biogeography of macroorganisms are also evident among bacteria.

Introduction

Many studies have examined the extent to which the identity, abundances and distributions of microorganisms vary with contemporary environmental conditions, but few have explored how their distribution may be influenced by historical factors (Martiny *et al*., 2006). Baas-Becking's statement 'everything is everywhere, but the environment selects' held that the distributions of microorganisms are cosmopolitan because of their enormous capacity for long-distance dispersal, whereas influences of habitat and environmental variables determine where local populations occur (de Wit and Bouvier, 2006). However, recent studies have revealed the existence of geographically isolated microbial populations (Papke *et al*., 2003; Whitaker *et al*., 2003). Geographic distance as a barrier to dispersal is most often invoked to explain these distributions (Martiny *et al*., 2006).

The geographic distributions of organisms reflect both the effects of contemporary environmental conditions and the legacies of historical geological and climatic conditions on the origins and spread of lineages (for a review see Martiny *et al*., 2006). A large body of literature shows how the distributions of macroorganisms can be classified into biogeographic provinces, which reflect the historic effects of plate tectonics, sea-level changes and barriers to dispersal such as mountain ranges (Hedges *et al*., 1996; Wares and Cunningham, 2001; de Bruyn *et al*., 2005). Methodological difficulties have hindered a comparable treatment of microbial biogeography. Advances in molecular techniques have found significant correlations of microbial distributions with environmental factors such as salinity or oxygen (Franklin *et al*., 1999; Casamayor *et al*., 2002), but there is little evidence of the latitudinal patterns of diversity or geographic provinces seen in macroorganisms (Morris *et al*., 2002; Knittel *et al*., 2005). If dispersal of microorganisms is limited by geographical barriers, biogeographical provinces should be apparent, such as those reported for terrestrial thermophiles from separate continents (Papke *et al*., 2003; Whitaker *et al*., 2003).

Recently, a framework for microbial biogeography was proposed to consider the relative influence of habitat (contemporary environmental factors) and province (historical legacies) by evaluating four alternatives hypotheses: (i) the null hypothesis that microorganisms are randomly distributed, (ii) that contemporary environmental variation is necessary and sufficient to account for current distributions, (iii) that contemporary distributions show clear legacies of historical events, such as tectonic events and geographic barriers, and (iv) that the distribution of microorganisms is due to both environmental effects and

Received 13 September, 2007; accepted 25 January, 2008. *For correspondence. E-mail cvesbach@unm.edu; Tel. (+1) 505 277 3418; Fax (+1) 505 277 0304.

historical events (Martiny *et al*., 2006). The apparent influence of habitat and province is affected by the scale of sampling (Martiny *et al*., 2006). On smaller scales, environmental factors are believed to be of primary importance in determining presence or absence of patterns of genetic variation, whereas on an intercontinental scale, isolation distance is believed to override environmental effects (Papke *et al*., 2003). It has been suggested that the joint influence of both environmental factors and historical contingencies should be most apparent at intermediate scales (10–3000 km) (Martiny *et al*., 2006).

Here we apply this framework to determine the relative influence of habitat and province on the distribution of members of the bacterial genus *Sulfurihydrogenibium* in thermal springs of Yellowstone National Park. Using molecular phylogenetic approaches, combined with environmental data and dispersal-vicariance analyses, we found geographically isolated microbial populations within the Yellowstone thermal ecosystem. The distribution pattern we detected was correlated with the boundary of Yellowstone's calderas (or volcanic craters) and suggests that volcanic eruptions in the past 2 million years explain more of the DNA sequence divergence than do contemporary factors such as habitat or geographic distance. These findings highlight the importance of historical legacies in determining contemporary microbial distributions and suggest that the same factors that determine the biogeography of macroorganisms are also evident among bacteria.

Results and discussion

We surveyed 18 thermal springs (Table 1) from essentially every major thermal area in the Yellowstone. Sites separated by 5–79 km were screened for bacterial 16S rRNA gene sequences belonging to the *Sulfurihydrogenibium*, which are members of the order *Aquificales*. *Sulfurihydrogenibium* spp. often dominate near-neutral Yellowstone springs (Barns *et al*., 1994; Hugenholtz *et al*., 1998; Reysenbach *et al*., 2000; Spear *et al*., 2005) and have been found in many terrestrial thermal systems worldwide, often contributing up to 95% of the biomass (Reysenbach *et al*., 2000; Takacs *et al*., 2001). Thirty-three unique 16S rRNA gene sequences (of 1728 screened clones) were retrieved from eight springs. Similarity of the *Sulfurihydrogenibium* sequences ranged from 97.1% to > 99% over 1446 nucleotides (based on uncorrected distance). One explanation for the observed sequence divergence is that the unique sequences reflect temperature ecotypes. However, there was no relation between the GC content of the sequences and the temperature of the spring from which they were collected ($r^2 = 0.062$, $n = 33$, $P = 0.14$).

Sulfurihydrogenibium sequences from the eight sites were distinct, suggesting that geographic isolation exists

Table 1. Location and selected geochemical parameters of thermal springs surveyed in this study.

Sample ID	Area	Northing[b]	Easting[b]	Temperature (°C)	pH	Conductivity (μS)	Alkalinity[c]	$SO_4^=$	As (total)	Fe (total)	SiO_2	H_2S
03YMAM001[a]	Mammoth	4979165.2	522886.6	73.5	6.43	2230	782	546	0.186	0.008	51	0.000
03YGIB023	Gibbon Hill	4948811.6	520789.8	73.8	8.51	2230	115	107	2.456	0.02	222	0.023
03YLOW027	Lower Geyser Basin	4934077.7	513245.8	86.2	8.43	1487	210	29	1.12	0.002	265	0.079
03YCOF035	Coffee Pots	4955983.6	554771.2	65.2	5.75	555	8	195	0.002	5.42	91	0.092
03YJOS038[a]	Joseph's Coat	4954349.2	553827.6	67.3	6.04	1680	144	242	1.24	0.023	134	0.000
03YHSB042[a]	Hot Spring Basin	4955653.2	558889.2	64.5	5.90	1678	40	701	0	0.007	234	4.650
04YLST010[a]	Lonestar	4918086.0	515156.4	73.6	7.90	1253	55	17	0.874	0.015	184	0.021
04YUPP012	Upper Geyser Basin	4923398.0	513006.0	78.4	8.80	2260	407	19	1.86	0	416	0.255
04YMID014	Midway Geyser Basin	4928874.8	514753.5	74.6	8.47	1475	380	14	1.5	0.004	239	0.034
04YSMH024[a]	Seven Mile Hole	4955774.6	547165.5	69.7	7.21	2050	172	161	3.7	0.017	187	1.167
04YSMJ030[a]	Smoke Jumper	4918098.3	503549.8	70.5	7.34	2130	0	72	0	0.042	223	2.417
04YHRT036	Heart Lake	4905826.8	538070.8	69.3	8.93	1500	251	146	1.06	0.008	308	0.654
04YHRT039	Heart Lake	4904299.0	539046.2	84.8	9.06	1660	254	104	0.446	0.009	177	0.035
04YSHO041[a]	Shoshone	4910920.5	516275.7	77.0	6.49	1122	251	50	0.4	0.018	256	4.583
04YSHO044	Shoshone	4910969.2	515601.5	80.9	8.32	1365	460	36	0.6	0.005	275	0.072
04YSHO046	Shoshone	4910961.6	515877.0	72.5	8.49	1402	382	35	0.824	0.01	292	0.243
04YBEC050	Bechler	4903641.8	509696.1	80.6	7.81	1003	454	21	0.178	0.037	171	0.065
04YCAS055[a]	Cascade Corner	4905566.2	496745.8	52.6	5.63	160	66	11	0	0.086	94.8	0.000

a. Sites where *Sulfurihydrogenibium* sequences were detected.
b. Northing and Easting given in UTMs, NAD 1983, Grid 12.
c. As HCO_3^-.
Concentration given as mg l^{-1} unless indicated otherwise.

Table 2. F_{ST} values estimated from 16S rRNA gene sequences and tested for significance against 1000 randomized bootstrap replicates with ARLEQUIN 3.1.

			F_{ST}
Between areas within regions			
Inside the calderas	Lonestar	Shoshone	0.420
		Smoke Jumper	0.642
		Cascade Corner	0.651
	Shoshone	Smoke Jumer	0.682
		Cascade Corner	0.691
Outside the calderas	Seven Mile Hole	Hot Spring Basin	0.468
		Joseph's Coat	0.666
		Mammoth	0.674
	Hot Spring Basin	Joseph's Coat	0.672
	Mammoth	Hot Spring Basin	0.680
		Joseph's Coat	0.914
Between regions			
Inside:outside	Lonestar	Seven Mile Hole	0.422
		Hot Spring Basin	0.430
		Joseph's Coat	0.620
		Mammoth	0.632
	Shoshone	Seven Mile Hole	0.459
		Hot Spring Basin	0.466
		Joseph's Coat	0.663
		Mammoth	0.672
	Smoke Jumper	Seven Mile Hole	0.684
		Hot Spring Basin	0.690
		Mammoth	0.894
		Joseph's Coat	0.926
	Cascade Corner	Seven Mile Hole	0.693
		Hot Spring Basin	0.670
		Mammoth	0.903
		Joseph's Coat	0.938

Significant ($P < 0.001$) differentiation was detected between all sites.

among the sites we assayed. The level of genetic variation among the eight sites was computed by Weir and Cockram's F_{ST} parameter (Wier and Cockerham, 1984; Ronquist, 1997), which is a measure of differentiation among samples. All F_{ST} values were large and significantly different, indicating that each of the sites was genetically distinct ($P < 0.05$; F_{ST} values ranged from 0.420 to 0.938, Table 2). The amount of genetic variation among sites is consistent with a hypothesis of isolation and diversification. Our F_{ST} values were in the same range as values computed between regions separated by 254–6305 km in a biogeographical study of the archaeon *Sulfolobus* (Whitaker *et al.*, 2003), and higher than values reported for bacteria associated with recent volcanic deposits 42–300 years old (Dunfield and King, 2004; 2005).

Spatial differentiation among the surveyed sites is supported by phylogenetic analysis. Sequences from six geographically distinct thermal areas of the park (Mammoth, Hot Spring Basin, Seven Mile Hole, Smoke Jumper, Shoshone and Cascade Corner) formed clades well supported by bootstrap analysis regardless of whether parsimony, maximum likelihood or neighbor-joining algorithms were used to construct the tree. The remainder of the tree topology was composed of sequences that did not form geographically distinct clades, which suggests that these sequences are cosmopolitan or that dispersal may occur between these sites (Fig. 1).

Unlike previous biogeographical studies of microorganisms, we did not find that geographical distance or geochemistry correlated significantly with the distribution of sequence types (Mantel test of F_{ST} values or genetic distance, compared with geographical distance and major ion chemistry, $P = 0.174$, pH, temperature, alkalinity, conductivity and sulfate combined, $P = 0.691$ or temperature alone, $P = 0.231$). From these results it would seem that there was no geographical or ecological pattern in the distribution of *Sulfurihydrogenibium*. However, a plot of genetic divergence as a function of geographic distance indicated a strong linear relationship ($r^2 = 0.55$, $P < 0.0001$) among sites up to 49 km apart, compared with a much weaker pattern across all sites ($r^2 = 0.13$ $P < 0.0001$, Fig. 2). Recognizing that this pattern may result because environmental parameters may be autocorrelated to distance at sites < 49 km, we repeated the Mantel tests described above for these sites, but did not find a significant relationship ($P = 0.128$–0.882). Therefore, these genetic–geographic patterns suggest that (i) dispersal is largely responsible for the genetic divergence we observed at distances < 49 km, and (ii) there is a thresh-

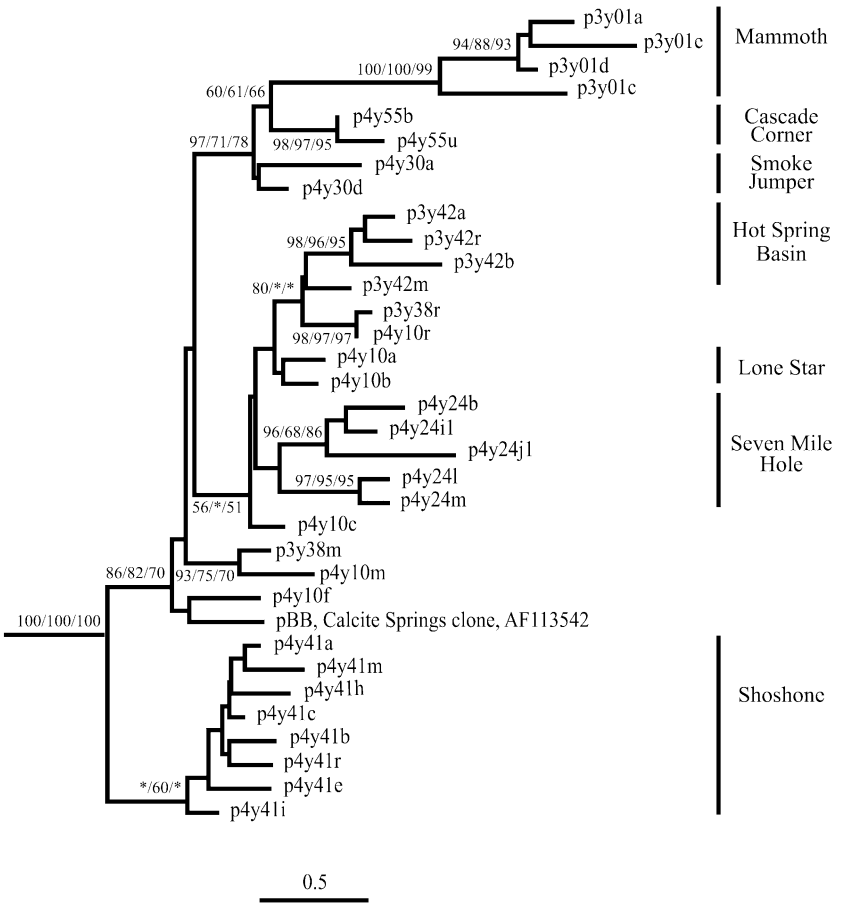

Fig. 1. Phylogenetic tree of sequences detected in this study. Trees were constructed by maximum parsimony, neighbor-joining and maximum likelihood methods, which resulted in the same overall topology. Tree given was generated by maximum parsimony, numbers at nodes represent bootstrap proportions for clades with significant support based on 1000 or 100 resamplings for maximum likelihood (parsimony/neighbor-joining/maximum likelihood, the asterisks represent bootstrap proportions below 50%). Tree was rooted with *Aquifex aeolicus* (AJ309733) and *Thermocrinis ruber* (AJ005640). Accession numbers for the remainder of sequences analysed are DQ906005 to DQ906039 and AJ113542 (pBB Calcite Springs clone). Scale bar represents 0.5 fixed mutations per 100 nucleotide positions. Clades that were comprised solely of sequences from a single geographic region are indicated by the bars with region names on the side of the figure. Sequences that did not cluster geographically were interpreted to be cosmopolitan.

old beyond which dispersal no longer dominates community assembly (although some dispersal is evident among sites separated by > 49 km, e.g. the 'cosmopolitan' sequences in Fig. 1 and the less divergent sequences in Fig. 2). This led us to explore alternate explanations for the distribution of sequence types, especially among sites greater than 49 km apart.

Geographic barriers and geological events offer two potential explanations for the abrupt breakdown of the genetic divergence–geographic distance relationship for

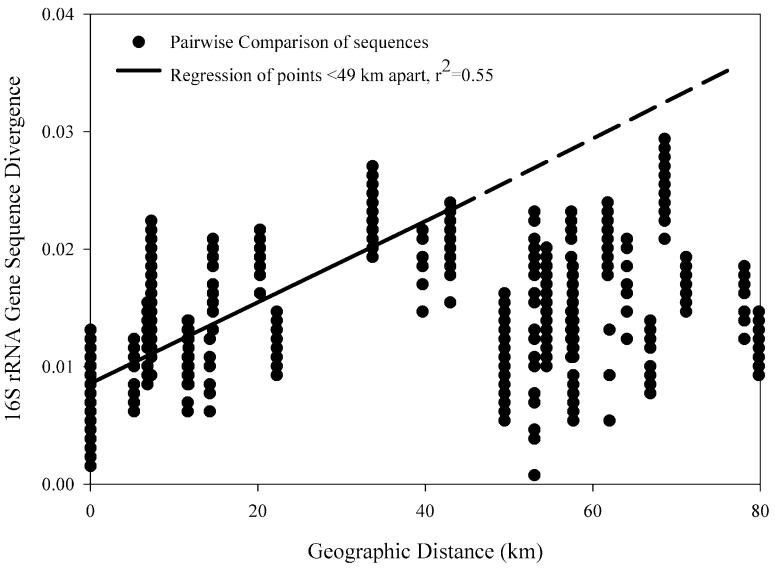

Fig. 2. Pairwise comparisons of uncorrected genetic divergence plotted against geographic distance (km). Regression line was fitted to data points < 49 km apart ($r^2 = 0.55$, $P < 0.001$). Dashed portion of line is given to illustrate the lack of correlation for distances > 49 km apart. Note that sequence comparisons with low divergence from sites > 49 km apart result from the cosmopolitan sequence types found in Fig. 1.

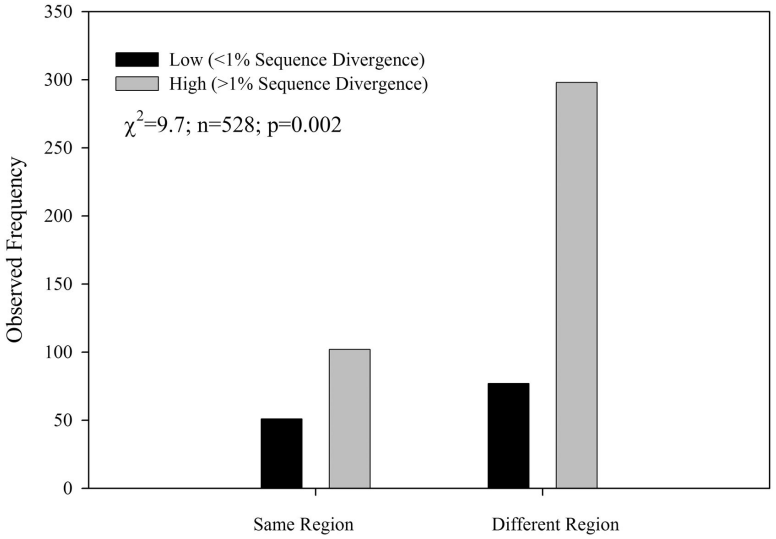

Fig. 3. Plot of observed frequencies used in χ^2 test. Number of pairwise comparisons of 16S rRNA gene sequences with low and high genetic divergence from sites found within the calderas (Same Region) or when at least one of the sites was located outside the calderas (Different Region). Four sites were inside the calderas (Cascade Corner, Smoke Jumper, Lone Star and Shoshone) and four sites were outside the calderas (Mammoth, Seven Mile Hole, Joseph's Coat and Hot Spring Basin). A significant relationship was found between level of genetic divergence and geographic position with respect to calderas (χ^2: 9.7, $n = 528$, $P = 0.002$) and was supported by a Monte Carlo simulation based on 5000 permutations ($P = 0.001$).

the more distantly separated sites. However, there were no apparent geographic barriers or corridors, such as mountain ranges or rivers, that explained our data, nor did the prevailing south-west wind across Yellowstone (Houston, 1973) or rhyolite lava flows and faults (Christiansen, 2001) correlate with the distributions. Subterranean dispersal is one factor that is not possible to assess, but is not believed to be important across the Yellowstone ecosystem because of the ring fracture system formed by the calderas, the high temperatures at shallow depth (~350°C) and lack of hydraulic connectivity between recharge zones (Fournier, 1989; Bonheyo *et al*., 2005). The eight sites where *Sulfurihydrogenibium* sequences were detected occur in either the northern or southern region of Yellowstone, suggesting a latitudinal trend. However, analysis of molecular variance (AMOVA) indicated no genetic structure among sites when placed into north or south groups (Fixation index = –0.004, $P = 0.565$). More of the molecular variation among sites could be attributed to between and within site variation, rather than variation between north and south groups of sites.

Inspection of Yellowstone geological maps revealed that the distance between sample sites that are inside and outside of the historic caldera boundaries was greater than 49 km, which suggested a relationship between genetic divergence among sequences and the calderas. Most geothermal activity in Yellowstone occurs within calderas that formed from volcanic eruptions during the past 2 million years. The most recent was 640 000 years ago, following earlier eruptions 1.3 and 2 million years ago (the 1.3 million year caldera lies to the west of the Park boundaries, Fournier, 1989). We found a significant relationship (χ^2: 9.7, $P = 0.002$, Pearson's phi = 0.113, $n = 528$) between the level of genetic divergence among the sequences (i.e. pairwise comparison of sequences that had a low < 1% and high > 1% difference) and the origin of the sequence with respect to the calderas (i.e. inside or outside of the calderas). Four sites were inside the caldera (Cascade Corner, Smoke Jumper, Lone Star and Shoshone) and four sites were outside the calderas (Mammoth, Seven Mile Hole, Joseph's Coat and Hot Spring Basin). The χ^2 result was confirmed by performing a Monte Carlo simulation upon the contingency table (5000 permutations, $P = 0.001$). Randomization procedures such as Monte Carlo provide a means for verifying a statistical pattern that is not dependent on the number of samples or any of the underlying assumptions of a parametric test such as the χ^2 (Eddington, 1995). Examination of the pairwise comparison of the sequences showed that the more divergent sequences most often resulted when one sequence was from outside the calderas (Fig. 3) and genetic divergence was the greatest when sites on opposite sides of the caldera boundaries were compared. Thus, it appears that historic volcanic eruptions have left their imprint on the genetic variation of thermophilic *Sulfurihydrogenibium* within Yellowstone. Calderas, surface expressions of past volcanic activity delineate biogeographical provinces among these microorganisms, although dispersal at sites < 49 km apart also seems to play a role in the distribution of Yellowstone *Sulfurihydrogenibium* sequences.

The phylogeographic history of a taxon may be inferred from the distribution of the most basal members. Two contrasting conclusions have been drawn: either the most basal members should occupy the region of its origin or they survive only in the peripheral areas (reviewed in Briggs, 2004). We reconstructed the history of Yellowstone *Sulfurihydrogenibium* sequence distribution by a dispersal vicariance approach based on the distribution and phylogeny of the 16S rRNA gene sequences using

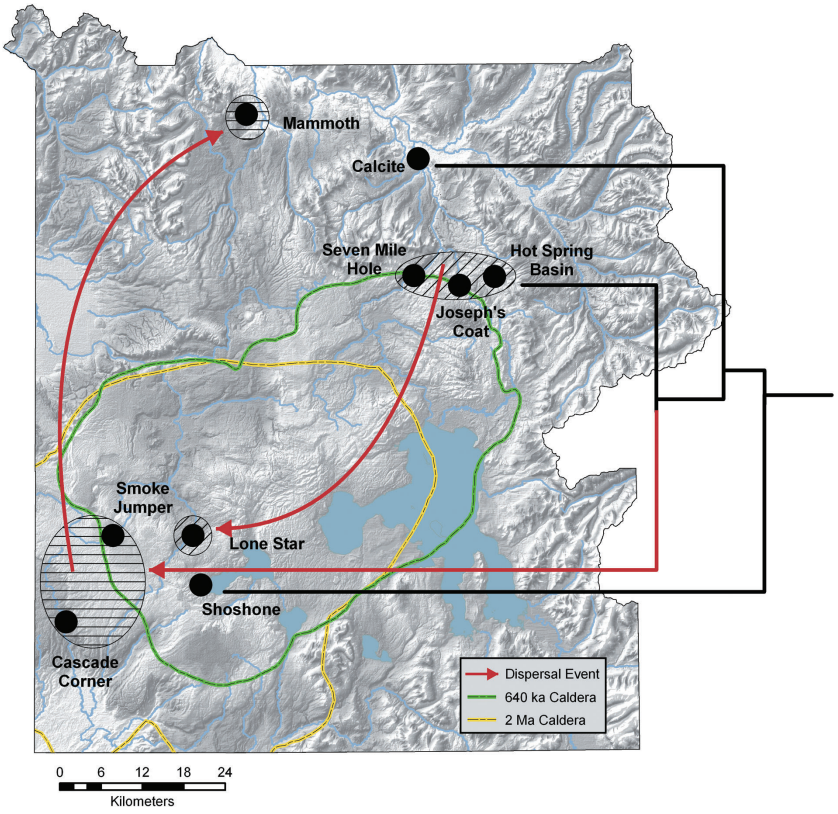

Fig. 4. Map of sampling sites where *Sulfurihydrogenibium* 16S rRNA gene sequences were detected. Exact geographic location of the sites is given in Table 1. Green and yellow lines represent the approximate topographic rim of the 640 ka and 2 my calderas today respectively. Ring fracture and rim of the actual calderas lie within these boundaries. Therefore, Seven Mile Hole and Joseph's Coat are classified as outside of the calderas. Four sites were inside the caldera (Cascade Corner, Smoke Jumper, Lone Star, and Shoshone) and four sites were outside the calderas (Mammoth, Seven Mile Hole, Joseph's Coat and Hot Spring Basin). The sites circled on the map indicate locations that are phylogenetically similar and collectively are areas of genetic sources and sinks (i.e. the Mirror plateau sequences clustered phylogenetically and were the ancestral source of the population now found in the Lone Star site). Backbone of the overall phylogeny among the sequences is used to illustrate the biogeographic history of the sites sampled, including dispersal events (red lines), inferred by DIVA analysis.

the DIVA software package (Ronquist, 1997). DIVA, a parsimony method, was used because it does not make assumptions about the existence or shape of a biogeographic pattern and is appropriate even when area relationships may be reticulate, rather than hierarchic. Furthermore, DIVA searches for optimal reconstruction of ancestral distributions by assuming vicariant speciation, but incorporating the potential for dispersal and extinction events in shaping contemporary distributions. The DIVA analysis suggested two possible ancestral distributions of the *Sulfurihydrogenibium* sequences: (i) within the youngest caldera, specifically the Shoshone spring, or (ii) within both the youngest caldera and totally outside the calderas, specifically the Shoshone and Mirror Plateau springs. Shoshone appears to be a special case within the calderas; all the other springs within the calderas were founded by dispersal events from outside the calderas. Additionally, the DIVA analysis indicated that three dispersal events were responsible for the distribution (Fig. 4). One dispersal event occurred from out of the calderas to the ancestor of the communities at Cascade Corner and Smoke Jumper, followed by a second dispersal back out of the caldera to Mammoth. The third dispersal was from out of the caldera to Lone Star. Either the ancestor of the Shoshone sequences is now extinct everywhere except Shoshone or it now exists in some other spring that has not been surveyed yet.

One scenario that could explain the DIVA results and that is consistent with our hypothesis that the calderas delineate biogeographical provinces for *Sulfurihydrogenibium* sequences in Yellowstone is that with each of the volcanic eruptions in Yellowstone, ancestral thermophiles went extinct within the calderas and as new springs formed, they were subsequently colonized from peripheral sites that survived outside the calderas. While we recognize that infinite dispersal and extinction events may not be accounted for in the static approach of using phylogeographic analysis tools such as DIVA, this approach provides a means for developing and testing phylogeographic hypotheses of microbial distributions. An alternate explanation for the relationship between *Sulfurihydrogenibium* sequence distribution and the calderas is that this observation reflects some as yet unrecognized underlying factor that is autocorrelated with the outline of the calderas.

Environmental factors appear to have little influence on the distribution of *Sulfurihydrogenibium* in the thermal springs that were the focus of this study. This finding may be an artefact of our site selection (near-neutral thermal springs only), or the limited number of sites assayed in this study. We recognize that the patterns we have detected are based on the current data set only, but the data and analyses presented here provide strong evidence that Yellowstone calderas delineate *Sulfurihydro-*

genibium biogeographical provinces. Although future analysis of additional sites will be necessary to further explore our findings, this report provides a framework for hypothesis development and future study. The precise mechanism by which Yellowstone calderas form barriers between the thermal springs is not clear, but *Sulfurihydrogenibium* sequences from these regions appear to reflect the historical events that formed them. If similar relationships are found for other groups of thermophiles in Yellowstone, then vicariance rather than a barrier to dispersal would be indicated for thermal communities. Finally, our study provides evidence that a legacy of geological history, which has been instrumental in describing the biogeography of macroorganisms, is also evident among present-day distributions of microorganisms.

Experimental procedures

Sampling sites

The microbial communities of 18 thermal springs within Yellowstone National Park were sampled during 2003 and 2004 (Table 1). Prior to collecting biomass, spring water was collected for geochemical analysis by sampling the water directly overlying the biomass. All water was filtered (0.2 μm) and preserved as appropriate for the analysis to be performed. Geochemical analysis, including anions, cations, trace metals, sulfide and various gases, was conducted using standard USGS methods (McCleskey *et al*., 2005). Environmental DNA was extracted from samples as described previously (Auchtung *et al*., 2006).

Phylogenetic analysis of 16S rRNA gene sequences

Amplification of the 16S rRNA gene from environmental DNAs was performed with universal primers 8F (5′-AGAGTTTGAT-CCTGGCTCAG-3′) and 1492R (5′-GGT TAC CTT GTT ACG ACT-3′) and cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA). Libraries were screened (96 clones each) by restriction fragment length polymorphism analysis and clones with unique patterns were fully sequenced (both strands, 2X coverage). Assembled sequences were submitted to Check Chimera (http://rdp8.cme.msu.edu/cgis/chimera.cgi) to test for the existence of PCR artefacts. GenBank accession numbers of the sequences determined in this study are DQ906005 to DQ906039. The 16S rRNA gene sequences were aligned with published sequences from *Thermocrinis ruber* (AJ005640), *Aquifex pyrophilus* (M83548), and the uncultured Yellowstone relative, pBB (AJ113542). The alignment was manually checked using established secondary structure and conserved sequence regions of the 16S rRNA, thus ensuring that only homologous regions were compared. Of the approximately 1600 nucleotides sequenced for each sequence type, 1446 bases were used in the distance matrices, phylogenetic analysis and ARLEQUIN analysis; sites that were not conserved in more than 50% of the aligned sequences or that were ambiguously aligned were not considered. Phylogenetic tree topologies were explored by parsimony, neighbor-joining and maximum likelihood analyses using heuristic search in PAUP (Swofford, 2003). A general time reversible model, plus gamma = 0.5 was used for the neighbor-joining and maximum likelihood analyses. Bootstrap proportions were determined on the final tree from 1000 resamplings of the data (100 resamplings for the maximum likelihood tree). We were concerned that the high level of within-site sequence variation was contributing to our results and repeated all analyses possible using one representative sequence from each site, which did not change or dampen the conclusions made here. Additionally, we conducted the complete analysis presented here using sequences downloaded from public sequence databases from other Yellowstone sites such as Joseph's Coat (W. Inskeep, unpublished) and Washburn and Obsidian springs (Spear *et al*., 2005). Again, the overall results of our analyses were the same, but these data were not included here because they tended to heavily skew the data set towards sites outside the calderas. Furthermore, the methods for identifying unique sequences among our data sets were not identical and a portion of the database sequences were not completely sequenced. Although our results are based on one sample per spring, Yellowstone *Sulfurihydrogenibium* (Reysenbach *et al*., 2005) communities are known to be temporally stable and the sequence types we detected were consistent with previous studies (Barns *et al*., 1994; Hugenholtz *et al*., 1998; Reysenbach *et al*., 2000; Bonheyo *et al*., 2005; Spear *et al*., 2005).

Geographic positioning and visualization

Geographic location of sample sites was determined in the field using a Trimble GeoXM (< 3 m accuracy) and differentially corrected to the GPS base station at Old Faithful using Pathfinder Office. Caldera layers were obtained from the Yellowstone Volcano Observatory (http://volcanoes.usgs.gov/yvo/) and represent the approximate topographic rim of the 640 ka and 2 my calderas today. The ring fracture and actual rim of the calderas lie within these boundaries (J.B. Lowenstern, USGS, pers. comm.). Therefore, Seven Mile Hole and Joseph's Coat are classified as outside the calderas.

Statistical analysis

Chi-squared test was performed on uncorrected pairwise sequence differences (1446 bp) grouped into categories (columns) that were distinguished by whether the two sequences came from sites within the same region (inside – Cascade Corner, Smoke Jumper, Lone Star and Shoshone, or outside of the calderas – Mammoth, Seven Mile Hole, Joseph's Coat and Hot Spring Basin) or different regions where at least one sequence was from outside the calderas and into groups of low (< 1%) and high (> 1%) levels of genetic divergence (rows). Monte Carlo simulations were performed to estimate a significance level of the chi-squared statistic. Contingency tables with the same marginal sums were generated 5000 times and the computed chi-squared statistic was compared with the observed chi-squared statistic. Site pairwise differences (F_{ST}

values, shown in Table 2) and AMOVA were performed in ARLEQUIN version 3.1 (Schneider *et al*., 2005), a software program for population genetics analysis that is appropriate for F_{ST} estimation regardless of the ploidy of an organism (L. Excoffier, pers. comm.). F_{ST} values were estimated for each site and were tested for significance against 1000 randomized bootstrap resamplings. The Mantel test was performed on matrices of (i) F_{ST} values or a matrix of uncorrected distance of one sequence representative from each site, (ii) geographical distance between sites and (iii) geochemistry. Geochemistry included major ions, pH, temperature, conductivity, alkalinity, and sulfate or temperature alone. These parameters were selected using non-metric multidimensional scaling analysis and principal components analysis of over 40 geochemical parameters measured for the sites, and by examining a piper diagram of major ion geochemistry.

History of the observed distribution was reconstructed using a dispersal vicariance approach with the DIVA software package (Ronquist, 1997) based on the distribution and phylogeny of the 16S rRNA gene sequences (Fig. 1). Biogeographic methods most often search for a single branching relationship among areas of endemism although the sites may have different histories and reticulate relationships. Additionally, dispersal and extinction are rarely incorporated into these models. DIVA is a unique tool in that it does not rely on area cladograms. DIVA searches for optimal reconstruction of ancestral distributions by assuming vicariant speciation, but incorporates the potential for dispersal and extinction events in shaping contemporary distributions. Ancestral states of a character at each node are inferred based on parsimony, but unlike other methods, it does not restrict widespread distributions to terminals or force ancestral distributions to one area. As with all biogeographic inference methods, it is a static approach based on the data at hand. The addition of data from other sites may radically change the outcome, but nevertheless, DIVA is a robust hypothesis testing tool. In this case the ancestral state of interest was the geographic location of the ancestor. The three possible states were (i) within the 2 my caldera, (ii) within the 600 k caldera or (iii) outside of the calderas. The phylogenetic tree in Fig. 1 was used for the analysis, each terminal node was assigned one of the three possible geographic states. This kind of analysis may, and in this case did, result in different but equally parsimonious states at a given node.

Acknowledgements

We thank J.H. Brown for comments and discussions regarding the manuscript. The Yellowstone Center for Resources (NPS) provided logistical and GIS support. This work was funded by a grant from the National Science Foundation's (NSF) Biotic Surveys and Inventories programme (DEB 02030677) to C.T.-V. and A.-L.R., a UNM RAC to C.T.-V., and NSF PEET (0328326) to A.-L.R. An American Society for Microbiology undergraduate fellowship provided additional funding to O.J.-W., and the UNM sequencing facility is operated in part by a grant from the National Institutes of Health (Grant 1P20RR18754).

References

Auchtung, T.A., Takacs-Vesbach, C.D., and Cavanaugh, C.M. (2006) 16S rRNA phylogenetic investigation of the candidate division '*Korarchaeota*'. *Appl Environ Microbiol* **72:** 5077–5082.

Barns, S.M., Fundyga, R.E., Jeffries, M.W., and Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* **91:** 1609–1613.

Bonheyo, G.T., Frias-Lopez, J., and Fouke, B.W. (2005) A test for airborne dispersal of thermophilic bacteria from hot springs. In *Geothermal Biology and Geochemistry in Yellowstone National Park*. Inskeep, W.P., and McDermott, T.R. (eds). Bozeman, MT, USA: Thermal Biology Institute, pp. 327–340.

Briggs, J.C. (2004) Older species: a rejuvenation on coral reefs? *J Biogeogr* **31:** 525–530.

de Bruyn, M., Nugroho, E., Hossain, M.M., Wilson, J.C., and Mather, P.B. (2005) Phylogeographic evidence for the existence of an ancient biogeographic barrier: the Isthmus of Kra Seaway. *Heredity* **94:** 370–378.

Casamayor, E.O., Massana, R., Benlloch, S., Ovreas, L., Diez, B., Goddard, V.J., *et al.* (2002) Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. *Environ Microbiol* **4:** 338–348.

Christiansen, R.L. (2001) *The Quaternary and Pliocene Yellowstone Plateau Volcanic Field of Wyoming, Idaho, and Montana. Geology of Yellowstone National Park, Professional Paper 729-G*. Reston, VA, USA: Department of the Interior, U.S. Geological Survey.

Dunfield, K.E., and King, G.M. (2004) Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposits. *Appl Environ Microbiol* **70:** 4242–4248.

Dunfield, K.E., and King, G.M. (2005) Analysis of the distribution and diversity in recent Hawaiian volcanic deposits of a putative carbon monoxide dehydrogenase large subunit gene. *Environ Microbiol* **7:** 1405–1412.

Eddington, E.S. (1995) *Randomization Tests*. New York, NY, USA: Marcel Dekker.

Fournier, R.O. (1989) Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. *Annu Rev Earth Planet Sci* **17:** 13–53.

Franklin, R.B., Taylor, D.R., and Mills, A.L. (1999) The distribution of microbial communities in anaerobic and aerobic zones of a shallow coastal plain aquifer. *Microbial Ecol* **38:** 377–386.

Hedges, S.B., Parker, P.H., Sibley, C.G., and Kumar, S. (1996) Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381:** 226–229.

Houston, D.B. (1973) Wildfires in northern Yellowstone National Park. *Ecology* **54:** 1111–1117.

Hugenholtz, P., Pitulle, C., Hershberger, K.L., and Pace, N.R. (1998) Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* **180:** 366–376.

Knittel, K., Losekann, T., Boetius, A., Kort, R., and Amann, R. (2005) Diversity and distribution of methanotrophic archaea at cold seeps. *Appl Environ Microbiol* **71:** 467–479.

McCleskey, R.B., Ball, J.W., Nordstrom, D.K., Holloway, J.M., and Taylor, H.E. (2005) *Water Chemistry Data for Selected Hot Springs, Geysers, and Streams in Yellowstone National Park, Wyoming, 2001–2002: U.S. Geological Survey Open-File Report 2004–1316.* Boulder, CO, USA: U.S. Department of the Interior, U.S. Geological Society.

Martiny, J.B., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4:** 102–112.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Papke, R.T., Ramsing, N.B., Bateson, M.M., and Ward, D.M. (2003) Geographical isolation in hot spring cyanobacteria. *Environ Microbiol* **5:** 650–659.

Reysenbach, A.-L., Banta, A., Civello, S., Daly, J., Mitchell, K., Lalonde, S., *et al.* (2005) The *Aquificales* in Yellowstone National Park. In *Geothermal Biology and Geochemistry in Yellowstone National Park*. Inskeep, W.P., and McDermott, T.R. (eds). Bozeman, MT, USA: Thermal Biology Institute, pp. 129–142.

Reysenbach, A.L., Ehringer, M., and Hershberger, K. (2000) Microbial diversity at 83 degrees C in Calcite Springs, Yellowstone National Park: another environment where the *Aquificales* and '*Korarchaeota*' coexist. *Extremophiles* **4:** 61–67.

Ronquist, F. (1997) Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Syst Biol* **46:** 195–203.

Schneider, S., Kueffer, J.-M., Roessli, D., and Excoffier, L. (2005) ARLEQUIN Ver. 3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1:** 47–50.

Spear, J.R., Walker, J.J., McCollom, T.M., and Pace, N.R. (2005) Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc Natl Acad Sci USA* **102:** 2555–2560.

Swofford, D.L. (2003) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Version 4. Sunderland, MA, USA: Sinauer Associates.

Takacs, C.D., Ehringer, M., Favre, R., Cermola, M., Eggertsson, G., Palsdottir, A., and Reysenbach, A.-L. (2001) Phylogenetic characterization of the blue filamentous bacterial community from an Icelandic geothermal spring. *FEMS Microbiol Ecol* **35:** 123–128.

Wares, J.P., and Cunningham, C.W. (2001) Phylogeography and historical ecology of the North Atlantic intertidal. *Evolution* **55:** 2455–2469.

Whitaker, R.J., Grogan, D.W., and Taylor, J.W. (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* **301:** 976–978.

Wier, B.S., and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38:** 1358–1370.

de Wit, R., and Bouvier, T. (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environ Microbiol* **8:** 755–758.