Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea

Rachel J. Whitaker,1* Dennis W. Grogan,2 John W. Taylor1

Barriers to dispersal between populations allow them to diverge through local adaptation or random genetic drift. High-resolution multilocus sequence analysis revealed that, on a global scale, populations of hyperthermophilic microorganisms are isolated from one another by geographic barriers and have diverged over the course of their recent evolutionary history. The identification of a biogeographic pattern in the archaean *Sulfolobus* challenges the current model of microbial biodiversity in which unrestricted dispersal constrains the development of global species richness.

It has recently been argued that limits to dispersal do not affect unicellular organisms because of their small size, enormous abundance, and metabolic plasticity (1). This view is supported by environmental surveys using both 16S ribosomal RNA (rRNA) signature sequences and phenotypic characters; these surveys have repeatedly identified apparently identical microorganisms in similar environments as far apart as the polar oceans (2, 3). In addition, several studies using multilocus analysis have shown that pathogenic bacterial species and *Bacillus* spore formers have global panmictic distributions (4–6). But such a ubiquitous distribution seems implausible for “extremophiles,” whose growth requires inter-temperate habitats that are often discontinuous and distant (7, 8). *Sulfolobus* species, for example, are archaea that inhabit solfataric geothermal springs and grow optimally at about 80°C and pH 3 (9). Abrupt temperature shifts induce cell cycle arrest and chromosomal DNA degradation in liquid *Sulfolobus* cultures (10), and no resistant spore state has been observed for any *Sulfolobus* species. It is therefore surprising that the same *Sulfolobus* species has been isolated from geothermal hot springs throughout the Northern Hemisphere (11). How organisms with such specific growth requirements can survive dispersal across the large, inhospitable distances that separate geothermal regions has intrigued those studying thermophiles since their discovery (12). To address this question, we examined the population structure in *Sulfolobus* and tested whether barriers to dispersal or ecological selection are primarily responsible for its development.

*Sulfolobus* strains were isolated from water and sediment samples collected from a nested hierarchy of geographic locations (13). At the largest scale, we sampled five regions separated by distances of >250 km: the Mutnovsky Volcano and the Uzon Caldera/Geyser Valley region on the Kamchatka Peninsula in Eastern Russia, Lassen Volcanic National Park and Yellowstone National Park in North America, and the solfataric region of western Iceland. Within each region, we focused our collections on two to three geothermal areas separated by 6 to 15 km (table S1). Within each area, we collected samples from up to seven different hot springs and isolated multiple individual colonies from each sample.

Segments of nine chromosomal loci (table S2) were sequenced from 78 individuals (13). Loci were selected that code for proteins with a variety of putative cellular functions and an even distribution around the genome of *Sulfolobus solfataricus* P2 (14). Of 4663 base pairs sequenced for each strain, we identified 138 variable positions. Although many divergent *Sulfolobus* species are capable of growth under our isolation conditions, all strains isolated in this study are closely related to *RenH1*, a strain that has been informally named *Sulfolobus “islandicus”* (15). All strains (including RenH1) are at least 99.8% identical in 16S rRNA sequence and differ by at most 1.05% across all nine loci.

Weir and Cockerham’s *FST* parameter (16) provides a measure of population differentiation based on variance in genetic diversity within and between groups of strains (17). Pairwise comparisons between strains grouped by region (Mutnovsky, Uzon/Geyser Valley, Lassen, Yellowstone, and Iceland), using a concatenated alignment of all loci, result in large, significant *FST* values (Table 1). Analyzed individually, the majority of loci support differentiation between regional populations (Table 1).
Within regions, significant $F_{ST}$ values were also found between Uzon Caldera and Geyser Valley in Kamchatka, Norris Geyser Basin and Geyser Creek in Yellowstone National Park, and Devil’s Kitchen and Brokeoff Caldera in Lassen National Park. These values indicate that there is a small but significant level of genetic differentiation between populations that inhabit different areas within the same geothermal region. To highlight the importance of multilocus analysis, it is worth noting that similar analyses of the sequence data from any single locus did not provide enough resolution to identify this level of population subdivision.

Within areas, $F_{ST}$ values calculated for pairwise comparisons between hot springs were not significantly different from zero, indicating that populations from local springs were not genetically differentiated. Identical genotypes were isolated from different hot springs in the same area, again suggesting that there is frequent gene flow among springs separated by small distances (<50 m). As shown in table S1, there is substantial genetic diversity within each hot spring. This microheterogeneity is consistent with the possibility of multiple microenvironments within a single spring, but we were unable to test for differentiation at this scale (18).

We performed a nested hierarchical analysis of molecular variance (AMOVA), using all strains grouped by areas within regions, to test the significance of the population structure as a whole (13). AMOVA analysis of nine loci, concatenated, partitioned 73% of the molecular variance among regions, 5% among areas within regions, and 22% within areas. Random permutation of the data between areas and regions allows us to reject a null model of panmixia ($P < 0.01$).

Using a concatenated alignment of the nine loci, we performed phylogenetic analysis to determine the evolutionary relationships among strains. A maximum likelihood phylogenetic tree resolves five distinct clades with significant bootstrap support (Fig. 1). These clades clearly correspond to the five geographic regions sampled, showing that strains within a region share a common evolutionary history distinct from strains found in other regions.

Baas-Becking’s formula of microbial biogeography, “everything is everywhere; the environment selects” (19), is widely accepted by microbiologists today. In this model, ecological characteristics are primarily responsible for the spatial patterning of microbial diversity. Several investigations have identified specific ecological parameters responsible for structuring bacterial communities (20–22). In contrast, despite the range of environments tested in each geothermal region (table S1), Fig. 1 shows that Sulfolobus strains cluster by geographic locale rather than by hot spring character. When the effect of geographic distance is removed, we find no significant correlation between genetic distance and absolute difference in hot spring pH or temperature in pairwise comparisons between strains (pH: Mantel $r = 0.06, P = 0.238$; temperature, $r = 0.05, P = 0.231$).

Hot spring geochemistry is largely determined by host rock composition. Although the host rock composition of Lassen National Park and both regions of Kamchatka are similar (andesites and basalts), Lassen National Park strains are more genetically similar to strains isolated from the geothermally distinct (rhyolitic) sites in Yellowstone National Park (23–26). Papke et al. (27) similarly describe region-specific cyanobacterial lineages that are recovered from environments with a wide range of geochemical parameters. From these qualitative assessments it is parsimonious to conclude that geographic isolation is primarily responsible for development of global population structure.

![Fig. 1. Maximum likelihood tree for concatenated alignment of all loci for all strains. Phylogenetic analysis was performed with PAUP* 4.0b10 (73). Numbers next to principal branches show bootstrap support of >70% using maximum likelihood and maximum parsimony (parentheses) algorithms. Individual strains are identified by site name from table S1 followed by a letter. Colors denote strains from the same sample. Boxes designate identical genotypes. Shaded areas highlight geographic regions. Scale bar indicates 1 substitution per 1000 sites.](www.sciencemag.org)

**Table 1.** Comparisons between populations reveal significant differentiation and divergence. $F_{ST}$ values were estimated using third codon positions and tested for significance against 1000 randomized bootstrap replicates with Arlequin 2.000, assuming no differentiation (33). All values are significantly different from zero ($P < 0.01$). $N_{gt}$, number of individual loci showing significant differentiation; I, divergence per 100 nucleotides ± SE based on 500 bootstrap resamplings of each population, estimated using MEGA version 2.1 (34); D, geographic distance (13).

<table>
<thead>
<tr>
<th>Between regions</th>
<th>$F_{ST}$</th>
<th>$N_{gt}$</th>
<th>$D$ (km)</th>
<th>I (± SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mutnovsky:Yellowstone</td>
<td>0.84</td>
<td>9</td>
<td>6305</td>
<td>6.5 (0.10)</td>
</tr>
<tr>
<td>Uzon/Geyser Valley:Yellowstone</td>
<td>0.79</td>
<td>9</td>
<td>6086</td>
<td>4.2 (0.7)</td>
</tr>
<tr>
<td>Mutnovsky:Lassen</td>
<td>0.82</td>
<td>8</td>
<td>5966</td>
<td>7.0 (1.0)</td>
</tr>
<tr>
<td>Uzon/Geyser Valley:Lassen</td>
<td>0.77</td>
<td>9</td>
<td>5775</td>
<td>4.4 (0.8)</td>
</tr>
<tr>
<td>Lassen:Yellowstone</td>
<td>0.50</td>
<td>6</td>
<td>1005</td>
<td>0.9 (0.3)</td>
</tr>
<tr>
<td>Mutnovsky/Uzon/Geyser Valley</td>
<td>0.59</td>
<td>7</td>
<td>254</td>
<td>3.4 (0.6)</td>
</tr>
<tr>
<td>Between areas within regions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uzon Caldera:Geyser Valley</td>
<td>0.14</td>
<td>2</td>
<td>15</td>
<td>0.4 (0.2)</td>
</tr>
<tr>
<td>Devil’s Kitchen:Brokeoff Caldera</td>
<td>0.36</td>
<td>3</td>
<td>8</td>
<td>0.6 (0.3)</td>
</tr>
<tr>
<td>Norris Geyser Basin:Geyser Creek</td>
<td>0.37</td>
<td>3</td>
<td>6</td>
<td>0.3 (0.1)</td>
</tr>
</tbody>
</table>
For neutral markers, genetic divergence between populations is inversely proportional to the level of gene flow between them. A model in which geographic distance increasingly restricts gene flow therefore predicts a positive correlation between genetic divergence and geographic distance between populations and provides a quantitative test for the effects of geographic isolation (28). Table 1 and Fig. 2 show the relationship between genetic divergence and geographic distance among *Sulfolobus* populations. A large, highly significant correlation coefficient was found in a Mantel test of all pairwise comparisons of genetic and geographic distance between strains (*r* = 0.82, *P* = 0.001) (13). This pattern of divergence is consistent with a model in which geographic distance (and the physiological challenges to survival that accompany it) contributes to the differentiation among populations.

Although biogeographical patterns are commonly found in multicellular eukaryotic organisms, they are unexpected for microorganisms. Several authors have observed geographic patterns in bacterial distributions, but their data lack the resolution needed to explicitly test the importance of geographic barriers (7, 29–31). The only other empirical test of microbial geographic structure is based on differences in the genomic fingerprint of repetitive elements (32). These data are difficult to interpret in an evolutionary context because they depend on the untested assumption of homology between similar-sized bands, and because there is no model for the evolution of repetitive elements that substantiates their use as a measure of genetic divergence between populations. The patterns of genetic divergence presented here are based on nucleotide substitutions, which are well characterized by classical evolutionary theory.

Our data show that gene flow among *Sulfolobus* populations is limited. Either the highly specialized growth requirements of these hyperthermophilic acidophiles prevent dispersal, or immigrants are unable to persist in established endemic populations, or both. This type of population subdivision will markedly decrease effective population size, thereby increasing the effect of genetic drift. In addition, genetically isolated populations have an increased potential for local adaptation to specific environmental conditions. Although this analysis has focused on extreme environments, it is not difficult to imagine other barriers to dispersal that may foster the development of geographically isolated populations (geotypes) in a broad array of microbial specialists. If true, this would suggest that microorganisms harbor substantially greater biodiversity and species richness than current estimates imply. Careful investigation of fine-scale microbial population structure promises to enhance both our understanding of microbial diversity in nature and our ability to analyze the evolutionary mechanisms that shape it.

References and Notes

13. Information on materials and methods is available online.
17. *F*ST values range from 0 to 1, with 0 indicating no genetic differentiation. Values greater than 0.15 indicate “great” divergence according to Wright’s model.
33. S. Schneider, D. Roessli, L. Excoffier, Arlequin: A Software for Population Genetics Data Analysis (Genetics and Biometry Laboratory, Geneva, Switzerland, 2000).
34. S. Kumar, K. Tamura, M. Nei, MEGA: Molecular Evolutionary Genetics Analysis (Pennsylvania State University, University Park, PA, 1993).
35. We thank M. Bidartondo, B. Bohannan, S. Dawson, W. F. Doolittle, P. Hugenholtz, J. Hughes, T. Pavlovská, E. Turner, and S. Wald for comments on this manuscript; G. Bell, J. Hansen, and B. Kinkle for assistance with sampling and isolations; and W. Zillig and K. Stedman for generously providing Sulfolobus strains from Iceland. Nucleotide sequences have been deposited in GenBank (accession numbersAY247838 toAY247949).Supported by a NASA Graduate Student Research Fellowship (R.J.W.), NSF grantMCB9733303 (D.W.C.), and NIH grants (J.W.T.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1086909/DC1
Materials and Methods
Table S1 and S2

References
16 May 2003; accepted 10 July 2003
Published online 24 July 2003;
10.1126/science.1086909
Include this information when citing this paper.