

Biogeographic Patterns of Divergence Between Populations of *Sulfolobus 'islandicus'*

Rachel Whitaker, Dennis Grogan* and John Taylor
University of California, Berkeley, *University of Cincinnati

rwhitaker@nature.berkeley.edu

We have developed a system to investigate microevolutionary mechanisms – dispersal, adaptation, selection, genetic drift, and recombination – occurring in natural populations of the hyperthermophilic archaeon *Sulfolobus 'islandicus'*. By comparing fine scale genetic differences between 78 strains isolated from different geothermal fields around the world, we have identified seven populations that are isolated by distance. Because these populations are differentiated and have been diverging a relatively short period of evolutionary time (0.74 - 2.8 million years), the pattern of nucleotide polymorphisms within and between them can be used to develop and test hypotheses about the mechanisms that drive their evolution.

Collection Sites

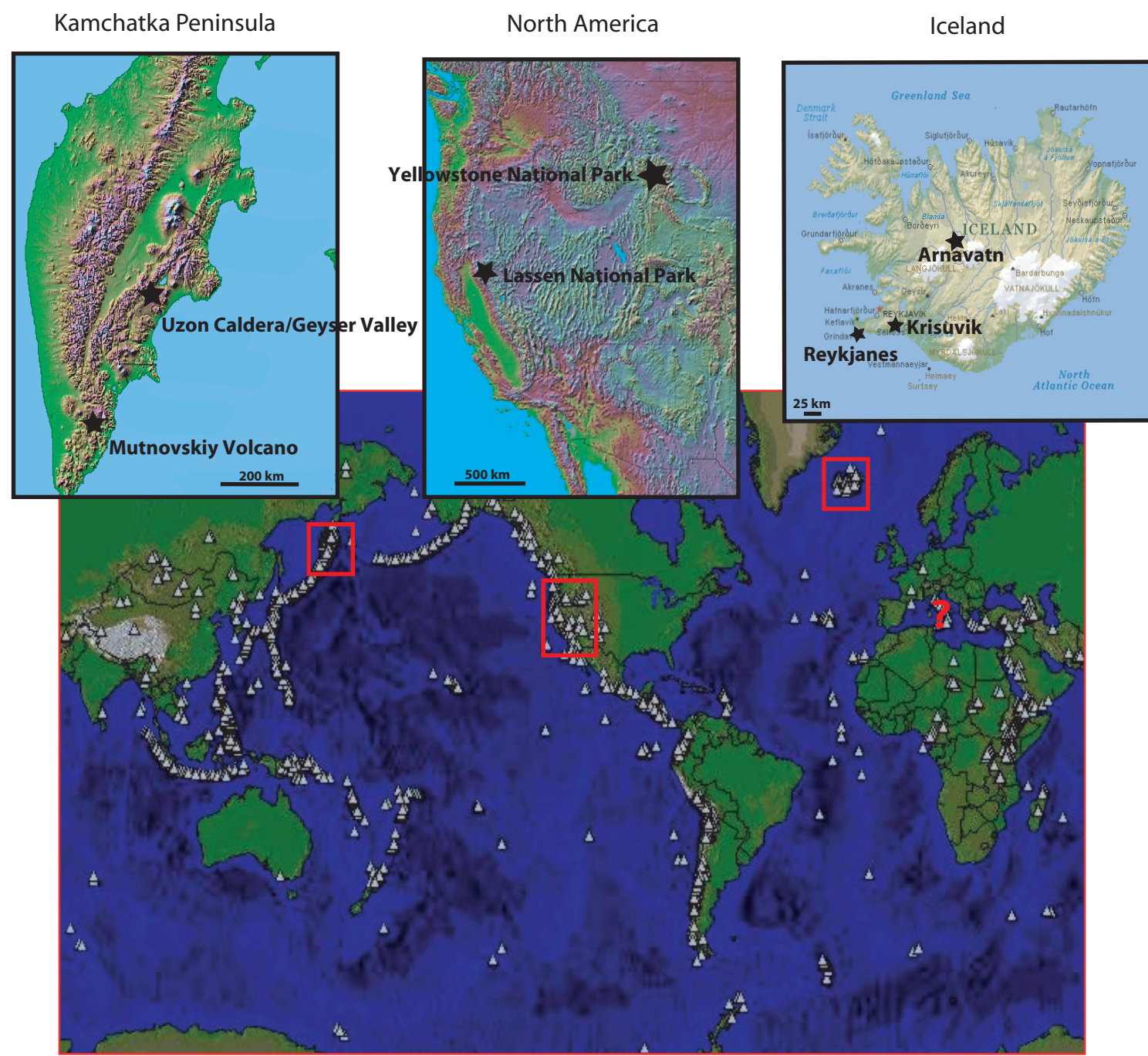


Figure 1. *Sulfolobus 'islandicus'* strains were isolated from geographically distant geothermal regions. To sample a nested hierarchy of distances, multiple areas were sampled within each region, multiple sites were sampled within each area and multiple strains were isolated from each site.²

Markers

Locus	Function	Number of base pairs sequenced	Number of polymorphic sites	Number of indels	Number of alleles
DNA polymerase B3	Replication	447	3	-	4
16S rRNA	Translation	552	10	-	13
Isocitrate lyase	Energy metabolism	616	18	1	16
Molybdopterine oxidoreductase	Energy metabolism	543	26	-	17
RNA helicase	Replication and repair	526	17	-	17
Glycogen debranching enzyme	Energy metabolism	540	13	-	11
Acetyl CoA acetyltransferase	Lipid metabolism	457	12	-	11
Permease	Transport	567	27	4	19
Chaperonin gamma subunit	Post translational modification	415	11	-	9
Total (multilocus genotypes)		4663	137	5	(65)

Table 1. Nine housekeeping genes were sequenced from 78 strains. Variation in each locus ranges from 0.67% in DNA polymerase to 4.8% in Permease and Molybdopterine oxidoreductase. 60% of the single nucleotide polymorphisms are synonymous changes. Primer design was based on the *Sulfolobus solfataricus* P2 Genome Sequencing Project, www.archbac.u-psud.fr/projects/sulfolobus.

Population Profiles

Region Name Area Name	Number of Samples	Number of Strains	Average Temp	Temp Range	Average pH	pH Range	Genetic Diversity (SE)
Mutnovskiy Volcano	5	18	85	20	2.7	1.5	0.0028 (0.0005)
Uzon-GV	5	13	76	22	4.6	3.0	0.0025 (0.0005)
Uzon	2	6	74	17	4.2	0.7	0.0016 (0.0004)
Geyser Valley	3	7	77	22	4.9	3.0	0.0028 (0.0006)
Lassen National Park	8	17	77	31	2.8	2.8	0.0012 (0.0003)
Devil's Kitchen	3	3	70	21	3.3	2.0	0.0009 (0.0003)
Tehema Caldera	5	14	81	13	2.4	1.3	0.0011 (0.0003)
Yellowstone National Park	12	26	66	46	3.6	5.3	0.0009 (0.0002)
Norris Geyser Basin	6	17	70	20	3.4	2.6	0.0007 (0.0002)
Geyser Creek	5	9	68	15	3.8	4.6	0.0008 (0.0002)

Table 2. To survey the genetic diversity, a range of temperatures and pH were sampled. Regions are highlighted in bold. Areas within regions are separated by at least 5 km. Genetic diversity is the average pairwise genetic distance per base pair (calculated with MEGA 1.01). Diversity does not reflect the range of temperature, pH, or number of sites sampled.

Differentiation Between Populations

Populations Compared	Fixed differences	Shared polymorphisms	Fst	Divergence (changes/bp)	Distance (km)
Mutnovskiy : Iceland	11	2	0.61	0.0042	7042.35
Uzon/GV : Iceland	7	1	0.51	0.0025	6813.42
Iceland : Lassen	24	0	0.82	0.0058	6464.10
Mutnovskiy : Yellowstone	18	1	0.80	0.0065	6305.46
Uzon/GV : Yellowstone	14	0	0.75	0.0042	6085.84
Iceland : Yellowstone	23	0	0.85	0.0053	5990.75
Mutnovskiy : Lassen	19	1	0.78	0.0070	5966.43
Uzon-GV : Lassen	15	1	0.72	0.0044	5774.57
Lassen : Yellowstone	1	1	0.48	0.0009	1004.84
Mutnovskiy : Uzon-GV	3	5	0.56	0.0034	253.96
Uzon : Geyser Valley	0	16	0.14	0.0004	15.05
Devil's Kitchen : Tehema	0	0	0.36	0.0006	7.97
Norris Geyser Basin : Geyser Creek	0	6	0.29	0.0003	5.71

Table 3. Measures of population differentiation between regions and between areas within regions show the degree of genetic isolation. The number of fixed differences is inversely proportional to the number of shared polymorphisms and is greater between more highly differentiated populations. F_{st} is a measure of differentiation based on the difference in variance of pairwise genetic distance within and between populations (values > 0.15 indicate "great" divergence according to the classic model of Wright³). All F_{st} values are significant (P < 0.05) (calculated using Arlequin v 2.00). Divergence is the average pairwise genetic distance between populations corrected for the genetic diversity within populations.

This work was funded by a NASA Graduate Student Research Fellowship

Acknowledgements: Iceland strains used in this study were generously donated by Dr. Kenneth Stedman. Greg Bell isolated the *Sulfolobus* strains from Yellowstone National Park and many of the isolates from the Kamchatka.

References: ¹ Zillig et al., 1994, Systematic and Applied Microbiology 16:609. ² Maps from left to right were edited from www.avo.alaska.edu/avo4/atlas/kamchap.htm, birrell.org/andrew/reliefMaps, www.bartleby.com/151/c112.html, www.volcano.si.edu/gvp/volcano/0000_map.jpg. ³ Wright, 1978, Evolution and the Genetics of Populations. Vol. 4. Variability Within and Among Natural Populations. University of Chicago Press, Chicago. ⁴ Sequences from NCBI: D85507, D14053, X90480, D26489, X89852. ⁵ Ochman et al. 1999, PNAS, 96:12638.

Phylogenetic Relationships

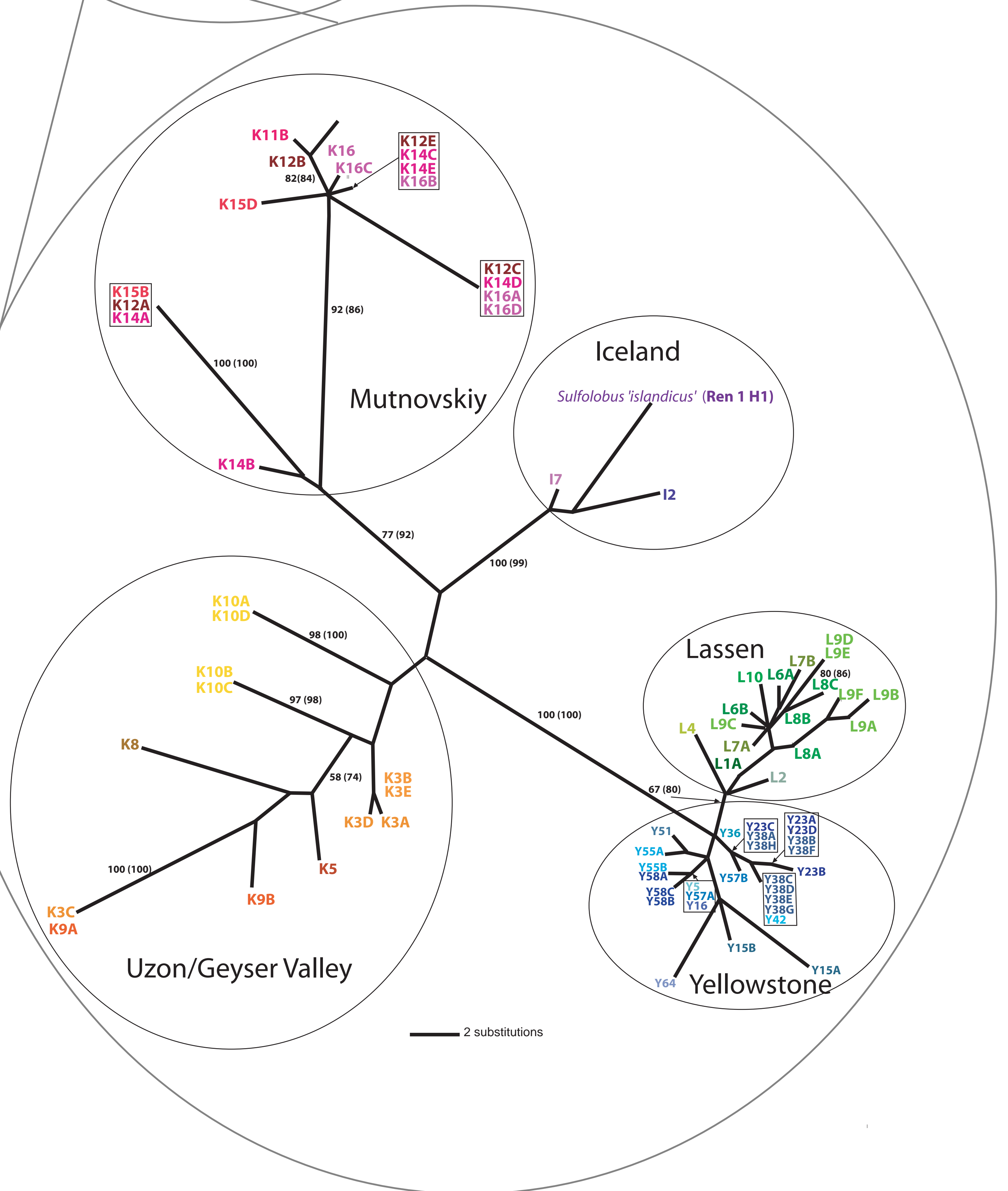
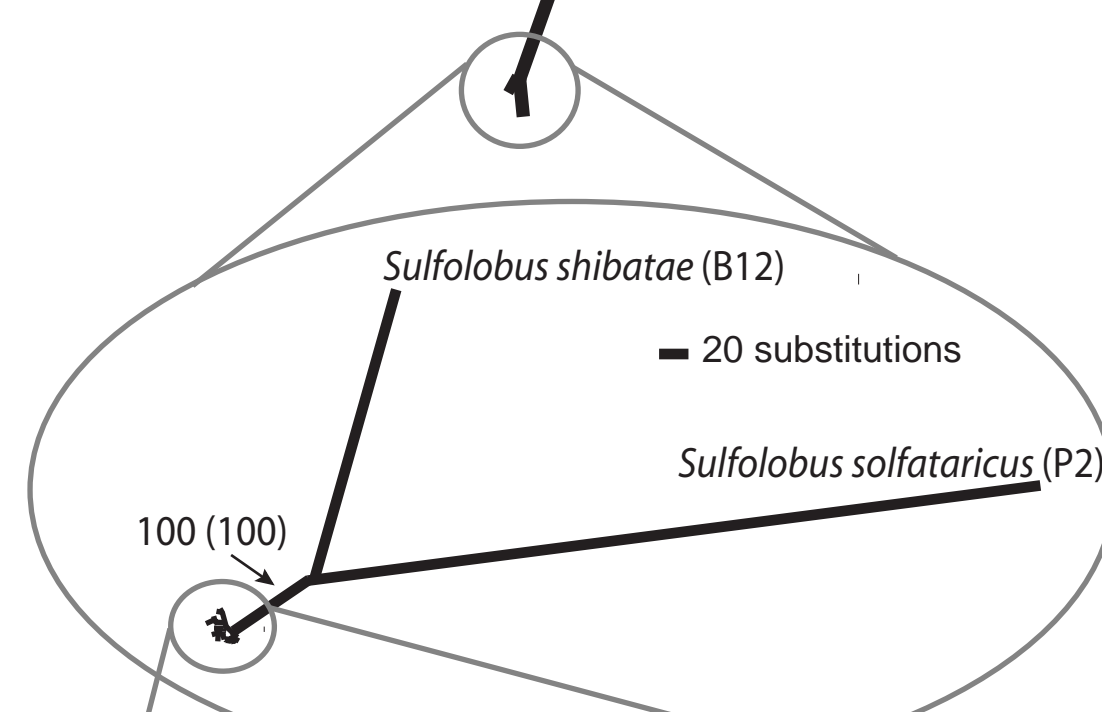
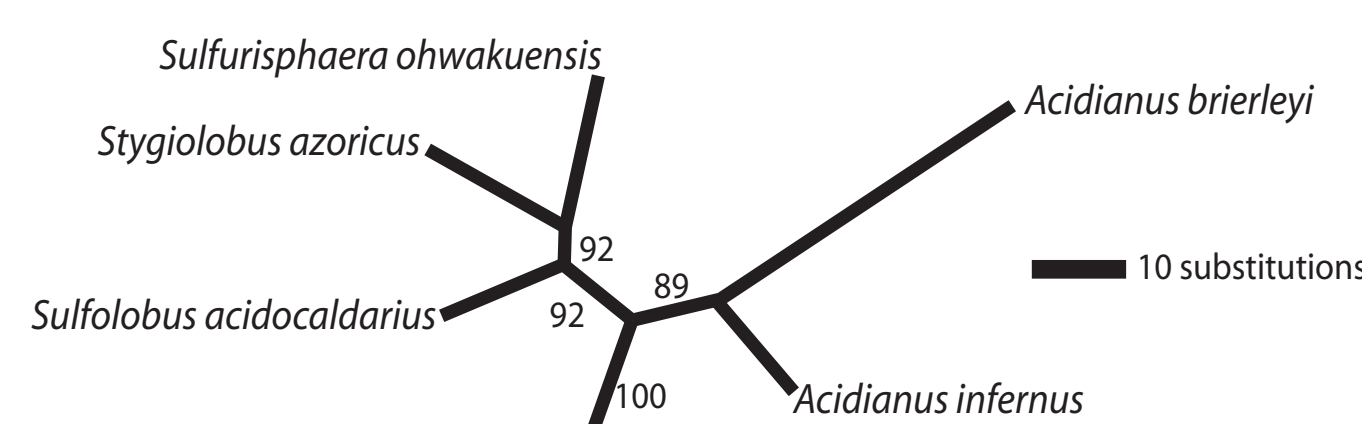


Figure 2. Three phylogenetic trees resolve relationships between strains and species in the *Sulfolobaceae* clade⁴. Top is a maximum parsimony tree of the 16S rRNA gene. Middle is a maximum likelihood tree constructed with seven of the nine loci listed in Table 1. Bottom is a maximum likelihood tree showing the relationships between strains using the nine loci in Table 1. The colors and numbers highlight strains isolated from the same site. Multiple strains from the same site are distinguished by the letter at the end of the name. Note that multiple genotypes were found in a single site and identical genotypes were isolated from different sites. Numbers next to branches are bootstrap values - likelihood (parsimony). All trees were constructed using Paup v. 4.0b 10.

Conclusions

The global *Sulfolobus 'islandicus'* population is highly structured by geographic distance indicating that there is very limited dispersal between geothermal regions. Even areas separated by as few as 5.7 km show differentiation.

There are cohesive populations within areas that include sites of different temperature and pH.

Genetic diversity is influenced more by geography than by selection of pH or temperature.

There is a strong correlation between genetic and geographic distance suggesting that new populations are established by individuals from the nearest geothermal area.

The Mutnovskiy and Uzon/Geyser Valley populations are significantly more diverse than the Yellowstone and Lassen populations.

Isolation by Distance

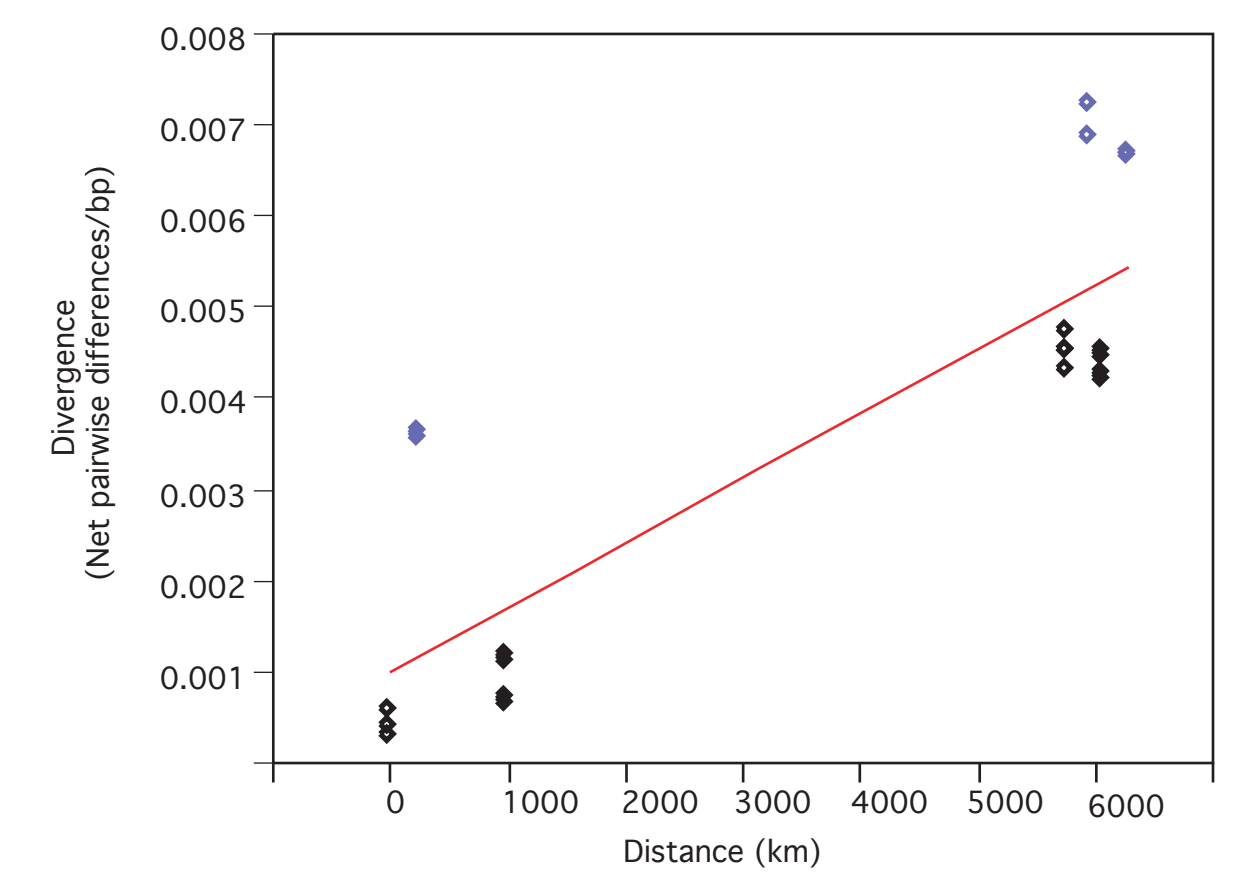


Figure 3. There is a significant correlation between geographic distance and genetic divergence between populations. Each point represents a single pairwise comparison between populations. Blue points represent comparisons to the Mutnovskiy population.

Discussion

Having defined the global population structure of *Sulfolobus 'islandicus'*, we now have a system in which to investigate the differences between isolated populations. The following three testable models for the difference in diversity between populations of *Sulfolobus 'islandicus'* describe their natural history. (Dates of isolation are determined by multiplying neutral genetic divergence by an estimate of 10⁻⁹ synonymous substitutions/year⁵.)

- 1) The Kamchatka populations is ancestral. About 1.55 million years ago a small founder population from Kamchatka established in North America. The low diversity in these populations is a result of a bottleneck caused by the founder effect. About 700,000 years ago, the Lassen and Yellowstone populations began to diverge from one another.
- 2) The Yellowstone and Lassen populations have recently gone through a selective sweep that has purged them of diversity. This sweep occurred roughly 700,000 years ago through both populations.
- 3) The diversity in Kamchatka is the result of migrants from a population that was not included in this study. Dispersal from this putative population does not reach the populations of North America.