

Recombination Shapes the Natural Population Structure of the Hyperthermophilic Archaeon *Sulfolobus islandicus*

Rachel J. Whitaker,*¹ Dennis W. Grogan,† and John W. Taylor*

*Department of Plant and Microbial Biology, University of California, Berkeley; and †Department of Biological Sciences, University of Cincinnati

Although microorganisms make up the preponderance of the biodiversity on Earth, the ecological and evolutionary factors that structure microbial populations are not well understood. We investigated the genetic structure of a thermoacidophilic crenarchaeal species, *Sulfolobus islandicus*, using multilocus sequence analysis of six variable protein-coding loci on a set of 60 isolates from the Mutnovsky region of Kamchatka, Russia. We demonstrate significant incongruence among gene genealogies and a lack of association between alleles consistent with recombination rates greater than the rate of mutation. The observation of high relative rates of recombination suggests that the structure of this natural population does not fit the periodic selection model often used to describe populations of asexual microorganisms. We propose instead that frequent recombination among closely related individuals prevents periodic selection from purging diversity and provides a fundamental cohesive mechanism within this and perhaps other archaeal species.

Introduction

The frequency of genetic exchange among individuals will dramatically influence the way diversity is generated and maintained within natural microbial populations (J. M. Smith, Feil, and N. H. Smith 2000). Theoretical models predict that clonally reproducing species evolve through the recurring process of periodic selection (Atwood, Schneider, and Ryan 1951). Because there is no recombination, a rise in the frequency of an adaptive allele results in the rise in frequency of its linked genome. Therefore, in response to natural selection, a purely clonal population will be purged of genetic diversity as the genotype with the highest relative fitness becomes fixed (Levin 1981). In contrast, genetic loci are unlinked in highly recombining species such as sexual eukaryotes, allowing natural selection to act upon regions of the genome independently and preserving diversity by preventing periodic genome-wide selective sweeps.

Bacteria and archaea reproduce clonally, vertically transmitting an identical copy of their genome to the next generation. However, the transfer of genetic material among clonal lineages has been shown to occur in microbial species through the uptake of free DNA from the environment or in association with the movement of viruses, plasmids, and transposable elements. Therefore, the evolutionary dynamics that define the level and structure of diversity within microbial populations combine elements of clonality and recombination. Microbial population structure will ultimately be determined by the relative frequency at which recombination breaks down clonal structure and periodic selection promotes it (Cohan 1994).

Empirical studies of bacterial species have revealed a range of population structures, from purely clonal to highly recombining (Selander et al. 1985; Souza et al. 1992; Suerbaum et al. 1998, 2001; Holmes, Urwin, and Maiden 1999; Feil et al. 2001, 2003; Koehler et al. 2003; Sarkar and Guttman 2004). These studies have pri-

marily focused on bacteria associated with agriculture or human and plant disease in which anthropogenic activity may tip the balance between periodic selection and recombination in either direction (i.e., through extreme selective regimes resulting from the use of antibiotics, facilitated dispersal, human alteration of environmental conditions, etc.). We extend fine-scaled analysis of population structure to the third domain of life and examine natural population structure in the thermophilic crenarchaeon *Sulfolobus islandicus*, which is relatively untouched by human activity.

Sulfolobus species are particularly amenable to population genetic analyses because they can be grown chemoheterotrophically on solid media in laboratory conditions, facilitating the isolation and identification of individuals—an elusive but essential property for high-resolution multilocus sequence analyses of wild microbial species. In addition, their unusual “extreme” growth requirements restrict growth of thermophilic *S. islandicus* to active geothermal regions, making population boundaries relatively easy to identify. In order to link the patterns of diversity we observed to the evolutionary mechanisms that created them, we first determined where individuals coexist by defining the biogeographical limits of five differentiated populations of *S. islandicus* from the Northern Hemisphere (Whitaker, Grogan, and Taylor 2003). Here, we investigate the genetic structure of one of these populations from the Mutnovsky Volcano on the Kamchatka Peninsula of eastern Russia to determine if recombination affects the generation and maintenance of genetic diversity.

Materials and Methods

Strain Isolation

A single water and sediment sample was taken from each of two hot springs (A and B, approximately 25 m apart) near the Mutnovsky Volcano on the Kamchatka Peninsula. The springs were similar in appearance, although they differed slightly in temperature and pH (A, 90°C, pH 3.0; B, 76°C, pH 2.0). Sixty colonies (24 from hot spring A and 36 from hot spring B) were isolated by direct plating on solid media at 75°C as described in Whitaker, Grogan, and Taylor (2003). Clonally purified cultures were isolated from each resulting colony and preserved for detailed analysis as previously described (Whitaker, Grogan, and Taylor 2003).

¹ Present address: Ecosystems Sciences, 137 Mulford Hall, University of California at Berkeley, Berkeley, CA 94720-3114.

Key words: recombination, population structure, archaea, *Sulfolobus*.

E-mail: rwhitaker@nature.berkeley.edu.

Mol. Biol. Evol. 22(12):2354–2361. 2005

doi:10.1093/molbev/msi233

Advance Access publication August 10, 2005

Table 1
Characteristics of Six Variable Loci

Locus	Position ^a	Description ^a	Length	SNPs ^b	<i>N</i> ^c	Number of Alleles	π (Standard Error)	θ_w ^e	<i>D</i> ^f	<i>P</i> ^g
II	1,168,340	Hypothetical isocitrate lyase	628	9	2	4	0.0065 (0.0022)	0.0042	2.900	0.001
III	1,424,702	Hypothetical molybdopterin oxidoreductase	566	12	4	2	0.0036 (0.0010)	0.0062	-0.420	0.617
IV	1,692,724	Putative ATP-dependent helicase	457	5	1	5	0.0025 (0.0011)	0.0032	0.126	0.427
VI	2,166,513	Hypothetical acetyl CoA acetyl-transferase	433	3	2	4	0.0013 (0.0009)	0.0021	-0.397	0.638
VIII	2,740,316	Hypothetical thermosome gamma subunit	332	2	1	3	0.0009 (0.0006)	0.0018	0.718	0.211
IX	385,800	Hypothetical transcription initiation factor	562	3	0	3	0.0012 (0.0008)	0.0016	0.154	0.377
Total			2,416	31	10	18	0.0030 (0.00063)	0.0032		

^a As listed in the *Sulfolobus solfataricus* P2 genome sequence annotation.

^b Number of single-nucleotide polymorphisms.

^c Number of nonsynonymous substitutions.

^d Average pairwise genetic distance (p-distance) at each locus between 60 members of the population.

^e Per-site Waterson's θ , $2N_e\mu$.

^f Tajima's *D*.

^g Proportion of simulated Tajima's *D* values greater than observed.

Because all isolates formed a monophyletic clade in comparison to other *Sulfolobus* species and differed from each other by no more than 1% of nucleotide positions across six protein-encoding loci (table 1), they were considered to be the same species (Palys et al. 2000), which has been designated *S. islandicus* informally by others (Zillig et al. 1994).

Marker Determination and Sequencing

Primers were designed to amplify 332- to 566-bp segments of six loci that were present in a single copy and were well distributed around the genome of the closely related species *Sulfolobus solfataricus* P2 (She et al. 2001). Primers and polymerase chain reaction (PCR) conditions for five of these loci (II, III, IV, VI, and VIII) were described previously (Whitaker, Grogan, and Taylor 2003). Locus IX is new to this study and was amplified using primers GCGTTAAGATGGAGAAAGT and GGGATCATAAAGAAAAGT in PCR reactions using conditions described previously for locus I of Whitaker, Grogan, and Taylor (2003). For all but 39 of the 360 total sequences, forward and reverse sequences were obtained for the entire locus segment. For the others, sequence from at least two-thirds of the locus segment was used for allele designations. A representative strain containing each unique combination of alleles recovered from 60 individuals was independently regrown from frozen stocks, and each locus was reamplified and resequenced to ensure fidelity.

Data Analysis

Maximum likelihood parameters were determined using Modeltest (Posada and Crandall 1998). The model of Felsenstein (F81) (Felsenstein 1981) was determined to be the optimal likelihood model for loci III, VI, VIII, and IX. For locus II, the optimal model was determined to be the Hasegawa-Kishino-Yano (HKY) model (Hasegawa, Kishino, and Yano 1985), with a transition to transversion ratio of 8.0744. Every substitution in locus IV was a transition, leading to an extremely high transition to transversion ratio. However, there was no difference between the topologies of maximum likelihood trees constructed for locus IV using the parameter-rich HKY model, which

accommodates the extreme transition to transversion ratio, or using the parameter-poor F81 model, which uses equal rates of transition or transversion. Position 374 in locus IV was excluded from phylogenetic analysis because it introduced homoplasy. For each locus, a maximum likelihood tree was determined from a heuristic search on 10 random replicates using the Tree Bisection-Reconnection algorithm in PAUP* 4.01b10 (Swofford 1996). Significant incongruence among trees was determined using the Shimodaira-Hasegawa likelihood ratio test (SH test) (Shimodaira and Hasegawa 1999). For loci II and IV, the SH test results were independent of the maximum likelihood model used.

Pairs of biallelic loci are classified as incompatible if the alleles are present in the data set in all four possible combinations. A compatibility matrix for each informative site in a concatenated alignment was constructed using the program SITES (Hey and Wakeley 1997). The proportion of compatible loci in pairwise comparisons between genotypes, and the significance of these values relative to a distribution of values simulated under the null hypothesis of free recombination, was determined using MultiLocus version 1.2 (Agapow and Burt 2000; <http://www.agapow.net/software/multilocus/>) with alleles designated as for I_A^S (below).

Estimates of the per-locus population recombination parameter (ρ) were determined using LDhat version 2.0 (McVean, Awadalla, and Fearnhead 2002). For all simulations, we used the segregating sites for five loci, excluding locus II due to evidence that it may be under selection (see *Results*). We input the average per-site population mutation rate (Waterson's θ) of 0.003 determined using DnaSP version 3.53 (J. Rozas and R. Rozas 1999). Genetic distances between loci were approximated using the distances between homologous loci in the genome of *S. solfataricus* P2 (She et al. 2001), or with loci in a random order separated by 500,000 bp, with no difference in results. Manually shuffling the gene order did not change the estimated recombination rate. The per-locus rate of initiation of gene conversion events is the appropriate estimate of recombination rate for asexual species (McVean, Awadalla, and Fearnhead 2002). Because the crossing-over model of LDhat includes both the beginning and the end of a gene conversion event when estimating ρ , the output from each simulation was divided by 2 for estimates of the per-locus

recombination (initiation of gene conversion) events relative to mutation. To test the significance of changes in likelihood scores for increasing ρ values, a likelihood ratio test was performed by comparing twice the difference in log-likelihood values to a χ^2 distribution with one degree of freedom. To estimate ρ/θ ratios, we used the per-locus population mutation parameter (θ per site \times number of sites per locus) averaged across all loci. To test for recombination, we used the likelihood permutation test which relies on the inverse correlation between chromosomal distance and linkage observed in recombining organisms. With this test, significance is determined by comparison of observed data to 1,000 data sets where all sites were randomly permuted (McVean, Awadalla, and Fearnhead 2002).

The standardized index of association (I_A^S) was determined with LIAN version 3.1 (Haubold and Hudson 2000). I_A^S measures the degree of association between alleles at different loci based on the variance in genetic distance between genotypes. I_A^S values were calculated for each hot spring individually and for a combined set of all strains. Values of I_A^S including all individuals were compared to those determined when duplicate genotypes were removed (clone corrected). It should be noted that disregarding the number of nucleotide differences when designating alleles can incorrectly bias results toward association because I_A^S values will reflect variance in genetic distance between individuals that is disproportionately increased by single-nucleotide substitutions. To minimize this problem, I_A^S was calculated using a single position to represent each major variant of loci II, III, IV, and IX (alleles 1 and 2 in fig. 1A) and the most informative position in loci VI and VIII (closest to 50:50 in allele frequency) (Burt et al. 1996). The significance of I_A^S was determined by comparison to the null hypothesis of free recombination simulated by 1,000 randomized reshufflings of alleles for each locus between individuals.

Tajima's D estimates were determined using Arlequin version 2.0 (Schneider, Roessli, and Excoffier 2000) and compared to the distribution of values determined from coalescent simulations, assuming both neutrality and population equilibrium. The MacDonald-Kreitman test for selection was performed in DnaSP version 3.53 (J. Rozas and R. Rozas 1999), using homologous sequences from the *S. solfataricus* P2 (She et al. 2001) genome for comparisons between species.

Prior analysis of *S. islandicus* populations yielded no evidence of barriers to gene flow between hot springs less than 15 km apart (Whitaker, Grogan, and Taylor 2003). To further ensure that barriers to gene flow between two hot spring populations did not bias our results, we performed each test for recombination on strains from each hot spring sample individually, as well as from the combined sample. Because results were similar at both levels of analysis, we conclude that there are no significant barriers to gene flow between our two hot spring samples and report values only for the combined analyses.

Results

We identified 34 polymorphic sites from a total of 2,978 bp sequenced for each of 60 *S. islandicus* strains

(table 1). Two to five alleles were identified at each of six variable loci, which contain 2–12 cosegregating single-nucleotide polymorphisms, as shown in figure 1A. Seventeen unique combinations of the alleles (genotypes) were identified from 60 individuals (fig. 1B). The frequency at which we recovered each genotype is shown in figure 1C. The most frequent genotype (genotype number 1) was recovered from 23% (14 of 60) of the strains.

We tested for evidence of recombinant genotypes in the population by evaluating whether the genealogical relationships described by each locus were congruent (Dykhuizen and Green 1991). Figure 2 shows that the maximum likelihood trees for each locus resolve different relationships among the 17 unique genotypes. The SH test on genealogies constructed with the 17 unique genotypes revealed significant ($P < 0.05$) incongruence in 53% of 30 reciprocal, pairwise comparisons among six loci (table 2). Because strictly clonal evolutionary histories should show the same relationships among individual genotypes at each locus, the significant conflict among single-gene phylogenies suggests that recombination has occurred in this population.

To test for evidence of recombination within a locus, we examined each pair of informative sites for compatibility (see *Materials and Methods*). Figure 3 shows that with one exception (position 374 in locus IV) polymorphic sites are compatible within a locus, indicating that nearly all such positions are linked. In contrast, 73% of all pairwise comparisons between loci (groups of linked sites) are incompatible, suggesting, as above, that there has been significant reassortment of alleles among genotypes. In addition to the incompatible position in locus IV, a shared polymorphism among variants was observed in locus IX (position 369, symbol \wedge in fig. 1A), suggesting that linkage within a locus may not be complete.

In order to quantify the importance of recombination to the generation of diversity within the population, we compared estimates of relative rates of recombination and mutation using two different analyses. Feil et al. (1999) suggest that the effect of recombination relative to mutation can be estimated by examining the number of nucleotide differences between alleles in individuals that differ from one another at a single locus (single-locus variants, SLVs). They reason that, in the short evolutionary time over which SLVs develop, the probability of more than one mutation occurring in the same locus or the same mutation occurring independently in multiple individuals is small. Feil et al. infer that SLVs with more than one cosegregating polymorphic site or that share a single mutation with alleles uncovered elsewhere in the population are likely to be acquired through recombination. We identified 12 pairs of SLVs among 17 genotypes. In seven of these 12 SLVs, variant alleles differ by more than one nucleotide substitution. Two pairs of genotypes (1 and 2, 6 and 8) differed by a single polymorphism that is shared by a different allele (\wedge in fig. 1A). Following the reasoning of Feil et al. (1999), this results in a ratio of alleles changed by recombination to mutation (r/m) of 9:3, suggesting that any one locus is three times more likely to change as a result of recombination than from mutation in this species. On a per-site basis, the r/m ratio is a measure of the diversity of

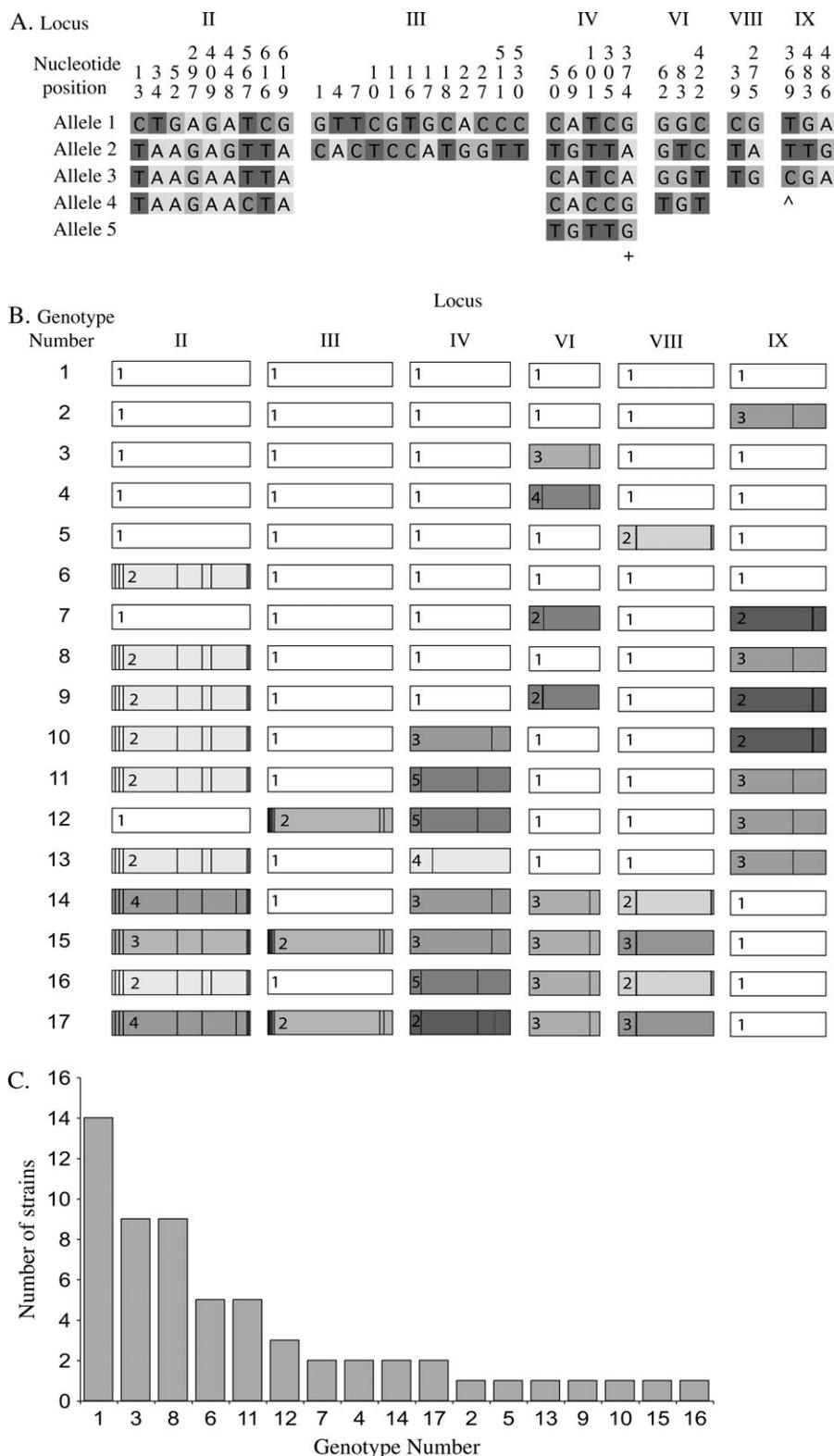


FIG. 1.—Distribution of variation among alleles and genotypes from the Mutnovsky *Sulfolobus islandicus* population. (A) The variable sites for each allele for each locus. Loci are listed in the order found in the genome of the closely related species *Sulfolobus solfataricus*. Nucleotide position is counted from the first position of the fragment used in analysis. “+” indicates incompatible site in locus IV. ^ indicates a shared polymorphism between alleles 1 and 2 in locus IX. (B) Distribution of alleles at six loci among 17 genotypes. Shading highlights different alleles. Numbers within the boxes show allele number as in (A). (C) The number of times each genotype was recovered from the population; genotype numbers as in (B).

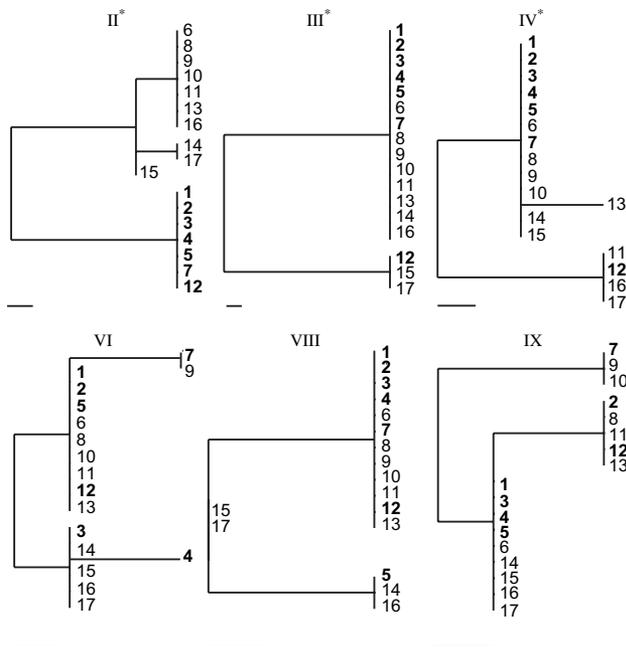


FIG. 2.—Incongruence among genealogies. Maximum likelihood phylogenetic trees for each locus showing the relationship between 17 genotypes. Scale bars represent 1 nucleotide change per 1,000 bp. A set of genotypes with the same allele for locus II are highlighted in bold in each genealogy to show the difference in the relationships among genotypes described by each locus. Genotypes are numbered as in figure 1. “*” shows topologies that are in significant conflict with the topology of each of the five other loci (see table 2). Site 374 in locus IV was excluded resulting in three alleles rather than the five shown in figure 1A.

recombinant alleles. The minimum ratio of sites changed by recombination compared to those changed by mutation in this population is 36:3, suggesting that, at a minimum, any site is 12 times more likely to change by recombination than mutation.

The r/m method is especially susceptible to underestimating rates of recombination in data sets with low diversity due to difficulty in distinguishing mutations from recombination events that introduce a single-nucleotide change (Feil et al. 1999). Diversification that occurs primarily through the stepwise accumulation of mutations predicts a positive correlation between the number of nucleotide differences between alleles and the number of allele differences between genotypes (Feil et al. 2003). We see no such relationship in all pairwise comparisons between genotypes (data not shown), supporting the result that recombination is the primary mechanism generating genotypic diversity in this population.

We further assessed the rate of recombination within the population using the coalescent-based method of McVean, Awadalla, and Fearnhead (2002). This method estimates the composite likelihood scores for the data across a range of per-locus recombination rates ($\rho = 2N_c r$). Across a range of 0–100, we found that the likelihood scores increased to a ρ value of 20 above which we found no significant increase in likelihood. Based on this analysis, we approximate the lower bound for the rate of initiation of recombination events to be 10 ($2N_c r/2$, see *Materials and Methods*). The average per-locus population

Table 2
Shimodaira-Hasegawa Test for Conflict Among Tree Topologies for Different Loci

Locus	II	III	IV	VI	VIII	IX
II	—	0.000^a	0.029	0.071	0.085	0.041
III	0.003	—	0.029	0.224	0.444	0.209
IV	0.001	0.000	—	0.224	0.074	0.209
VI	0.002	0.000	0.029	—	0.090	1.000
VIII	0.001	0.000	0.029	0.089	—	1.000
IX	0.002	0.000	0.029	1.000	0.778	—

^a P values for the difference in log-likelihood scores between the maximum likelihood tree topology for each locus (columns) and the data for that locus constrained by the maximum likelihood topology of other loci (rows). P values < 0.05 are bold.

mutation rate ($\theta = 2N_c \mu$) was determined to be approximately 1.5. The resulting ρ/θ ratio for this analysis is on the order of 6.6:1. Using this method, we rejected the null model of no recombination ($P < 0.05$).

We were unable to resolve the optimal track length using the gene conversion model of recombination provided in the method of McVean, Awadalla, and Fearnhead (2002). Future studies including additional loci from a larger range of genetic distances across the chromosome will better resolve the size of the DNA fragments that are incorporated during genetic exchange in this species.

We gauged the extent to which recombination affects the overall structure of a population by measuring the standardized index of association (I_A^S) (Smith et al. 1993; Haubold and Hudson 2000). I_A^S is expected to be zero if populations are freely recombining and greater than zero if there is association between alleles. We estimated I_A^S to be 0.076 when all 60 strains were included in the data set. This I_A^S value is significantly different from zero ($P = 0.001$) and would thus reject the null hypothesis of free recombination. When duplicate genotypes were removed from the data set, however, I_A^S decreased to 0.033, which

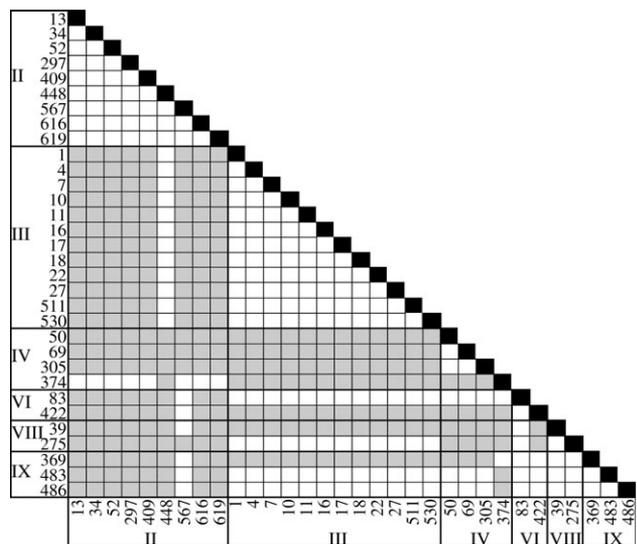


FIG. 3.—Compatibility matrix. Each informative site is listed on x and y axes. Pairs of sites are incompatible if they are observed in all four possible combinations. Shaded blocks show pairs of sites that are incompatible. White blocks are sites that are compatible.

was not significantly different from values expected under the null hypothesis of free recombination ($P = 0.132$). The difference in I_A^S between complete and clone-corrected data sets provides evidence of an “epidemic” population structure in which certain clones rise in frequency relative to the rest of the recombinant population (Smith et al. 1993).

High rates of recombination would allow selection to act on individual loci. We tested for evidence of selection at each locus using Tajima’s D statistic (Tajima 1989). As shown in table 1, only locus II had a Tajima’s D value that was significantly greater than 1, which is consistent with diversifying selection. This result reflects the fact that two divergent alleles at locus II are maintained in the Mutnovsky population at a higher frequency than would be expected under neutrality. An alternative explanation for a positive Tajima’s D statistic could be population substructure. However, population substructure prevents mixing between individuals in different populations and would be expected to affect all loci in the genome. The fact that we did not observe the positive Tajima’s D value for any other locus indicates that only locus II (encoding a putative isocitrate lyase) is under diversifying selection. We found no significant signal for positive selection at any locus by comparing the ratio of nonsynonymous to synonymous substitutions within *S. islandicus* to that found between *S. islandicus* and *S. solfataricus* using the MacDonald-Kreitman test (MacDonald and Kreitman 1991). As is observed in many populations, negative or purifying selection against amino acid changes appears to be occurring at all loci as shown by the fact that only 29% of total nucleotide polymorphisms were replacement (nonsynonymous) substitutions (table 1).

Discussion

It has been suggested that periodic selection is a fundamental force in the development of genetic diversity in microbial populations by providing cohesion within clusters of individuals inhabiting the same ecological niche, while promoting divergence among clades adapted to different environments (Cohan 2001). The genetic structure of the Mutnovsky population of *S. islandicus* does not fit the periodic selection model. Recombination occurs at rates greater than mutation, and the majority of the diversity is in the form of mosaic combinations of variant alleles. These data suggest that recombination not only generates diversity but also maintains recombinants within the population by preventing selection from acting to fix a single adaptive genotype.

Recently, several empirical studies of bacteria and archaea have identified a surprising level of genetic heterogeneity within local populations, which the authors note are difficult to reconcile with the theoretical predictions of the periodic selection model (Acinas et al. 2004; Papke et al. 2004; Tyson et al. 2004; Venter et al. 2004). It has been suggested that heterogeneity results from neutral mutations that accumulate either because selection events are rare or because microbial populations are too vast to suffer the full effects of genetic drift (Giovannoni 2004; Thompson et al. 2005). If, as has been suggested for marine bacterioplankton populations (Thompson et al. 2005),

S. islandicus genetic heterogeneity was similarly maintained because periodic selection events are rare, we would expect that each unique allele would be associated with a unique genotype and never in the mosaic combinations shown in figure 1. In light of our data and other studies of pathogenic bacteria showing frequent recombination (Souza et al. 1992; Suerbaum et al. 1998, 2001; Holmes, Urwin, and Maiden 1999; Feil et al. 2001; Koehler et al. 2003), we suggest that in many natural microbial populations, genetic heterogeneity may be maintained because recombination prevents periodic selective sweeps from purging diversity.

The maintenance of diversity in this *S. islandicus* population might also be facilitated by an environment that is temporally or spatially heterogeneous, or both. Temporal heterogeneity, in which the environment changes faster than a single adaptive genotype can become fixed in a population, may be consistent with the hydrological dynamics of hot springs that result in great variations in temperature, pH, and geochemical conditions over short timescales (Brock 1978). Spatial heterogeneity may also maintain genetic diversity through niche specialization, which would prevent competitive exclusion and periodic selection (Turner, Souza, and Lenski 1996). However, the evidence presented here posits a prominent role for recombination in making this *Sulfolobus* population into a cohesive, albeit heterogeneous, species, whether or not spatial or temporal heterogeneity assists in maintaining diversity by reducing competition between individuals.

The capacity for recombination within *S. islandicus* may be an adaptive strategy with evolutionary consequences that parallel those proposed to explain the evolution of sex (Redfield 2001). Frequent recombination prevents periodic selection from purging genomic diversity, but it also allows deleterious mutations to be purged from the populations without the risk of losing linked adaptive alleles (Muller 1964; Duarte et al. 1992). Furthermore, reassortment of adaptive alleles into new combinations accelerates the adaptive process (Goddard, Godfray, and Burt 2005) and has been shown to be especially advantageous in highly variable environments (Burger 1999; Lenormand and Otto 2000). Finally, genetic exchange and recombination could also serve as mechanisms for maintaining chromosomal integrity in response to the high levels of DNA damage induced by a thermoacidic environment (Grogan 1998).

Although recombination occurs within this *S. islandicus* population, there is an overrepresentation of certain genotypes (e.g., the most frequent genotype was recovered from 24% of strains rather than the 13% predicted from allele frequencies), which is reflected in the I_A^S values calculated for the complete data set. This observation indicates that a single clone may rise in frequency relative to the rest of the recombinant population but does not entirely exclude other genotypes within the population, as would occur in a complete selective sweep. Similar “epidemic” population structures have been observed in global collections of some bacterial pathogens (Feil et al. 2001). It has been suggested that this epidemic population structure may result from selection of certain clonal types that have increased pathogenicity, antibiotic resistance, or ability to evade the immune system or through neutral microepidemics resulting from

transmission in local populations (Feil and Spratt 2001; Fraser, Hanage, and Spratt 2005). The fact that we observe a similar epidemic structure in *S. islandicus* indicates that this type of structure is not limited to the unique lifestyle of bacterial pathogens. Epidemic clonal expansions may occur continuously in microbial populations without the severe selective constraints imposed by human activity.

The mechanism of genetic exchange and recombination in natural populations of *S. islandicus* remain mysterious. Species of crenarchaea have components of homologous recombination systems that are more similar to those found in eukaryotes than bacteria (Seitz, Haseltine, and Kowalczykowski 2001). Potential agents of gene transfer have been identified for *Sulfolobus* species (Schleper et al. 1995; Martusewitsch, Sensen, and Schleper 2000; Rice et al. 2001), and genetic marker exchange has been demonstrated under laboratory conditions (Grogan 1996). The mechanism of gene transfer in the hot spring environment through any of these systems has not been determined, and the frequency with which they may facilitate recombination is unknown.

If homologous recombination efficiency in archaea decreases with genetic divergence (Majewski and Cohan 1998; Daubin, Moran, and Ochman 2003), the model of microbial population dynamics described above predicts that recombination provides cohesion among closely related individuals in a process analogous to the cohesion in biological species of sexual eukaryotes (Dykhuizen and Green 1991). Evidence for this pattern of genetic exchange has been observed in two recent studies of the species from the euryarchaeal kingdom of the archaeal domain that identified evidence for recombination but did not assess its effect on natural population structure. The first study recovered recombinant genome fragments of *Ferroplasma* type II from the community genome sequence of a biofilm growing on acid mine drainage (Tyson et al. 2004). By reconstructing the genome sequences from three euryarchaeal members of this community, Tyson et al. inferred that recombination break points occurred approximately once every 5kb when nucleotide sequences were >98% similar (i.e., within *Ferroplasma* type II species), but found very few instances of recombination between more diverged sequences, (i.e., between *Ferroplasma* type I and II species), which averaged 77% nucleotide identity. More recently, Papke et al. (2004) demonstrated evidence for recombination in the halophilic euryarchaeon *Halorubrum* isolated from Spanish salterns. Although significant nucleotide diversity was identified among strains isolated for this study, here too, the majority (all but one) of recent recombination events (as seen by SLVs) were identified between individuals that showed very little sequence divergence. In combination with these studies, our demonstration of high rates of recombination in *Sulfolobus islandicus*, shows that recombination is a fundamental force maintaining diversity with natural microbial populations and providing cohesion within species of the archaeal domain.

Supplementary Material

All sequences have been submitted to GenBank with accession numbers AY740403–AY740423.

Acknowledgments

We thank J. Banfield, E. Feil, E. Holmes, G. McVean, J. Townsend, E. Turner, and S. Wald and two anonymous reviewers for helpful comments on the manuscript and G. Bell, J. Hansen, and R. Olarte for assistance with isolations. This study was supported by a National Aeronautics and Space Administration Graduate Student Research Fellowship (R.J.W.), National Science Foundation (NSF) grant MCB9733303 (D.W.G.), and NSF grant DEB0316710 (J.W.T.).

Literature Cited

- Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz. 2004. Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**:551–554.
- Agapow, P. M., and A. Burt. 2000. MultiLocus. Version 1.2. Ascot, Berks.
- Atwood, K. C., L. K. Schneider, and F. J. Ryan. 1951. Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **37**:146–155.
- Brock, T. D. 1978. *Thermophilic Microorganisms and Life at High Temperatures*. Springer-Verlag, New York.
- Burger, R. 1999. Evolution of genetic variability and the advantage of sex and recombination in changing environments. *Genetics* **153**:1055–1069.
- Burt, A., D. A. Carter, G. L. Koenig, T. J. White, and J. W. Taylor. 1996. Molecular markers reveal cryptic sex in the human pathogen *Coccidioides immitis*. *Proc. Natl. Acad. Sci. USA* **93**:770–773.
- Cohan, F. M. 1994. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *Am. Nat.* **143**:965–986.
- . 2001. Bacterial species and speciation. *Syst. Biol.* **50**:513–524.
- Daubin, V., N. Moran, and H. Ochman. 2003. Phylogenetics and the cohesion of bacterial genomes. *Science* **301**:829–832.
- Duarte, E., D. Clarke, A. Moya, E. Domingo, and J. Holland. 1992. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. USA* **89**:6015–6019.
- Dykhuizen, D. E., and L. Green. 1991. Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* **173**:7257–7268.
- Feil, E. J., J. E. Cooper, H. Grundmann et al. (12 co-authors). 2003. How clonal is *Staphylococcus aureus*? *J. Bacteriol.* **185**:3307–3316.
- Feil, E. J., E. C. Holmes, D. E. Bessen et al. (12 co-authors). 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. USA* **98**:182–187.
- Feil, E. J., M. C. J. Maiden, M. Achtman, and B. G. Spratt. 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol. Biol. Evol.* **16**:1496–1502.
- Feil, E. J., and B. G. Spratt. 2001. Recombination and the population structures of bacterial pathogens. *Annu. Rev. Microbiol.* **55**:561–590.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- Fraser, C., W. P. Hanage, and B. G. Spratt. 2005. Neutral micro-epidemic evolution of bacterial pathogens. *Proc. Natl. Acad. Sci. USA* **102**:1968–1973.

- Giovannoni, S. 2004. Evolutionary biology: oceans of bacteria. *Nature* **430**:515–516.
- Goddard, M. R., H. C. J. Godfray, and A. Burt. 2005. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**:636–640.
- Grogan, D. W. 1996. Exchange of genetic markers at extremely high temperatures in the archaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* **178**:3207–3211.
- . 1998. Hyperthermophiles and the problem of DNA instability. *Mol. Microbiol.* **28**:1043–1049.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**:160–174.
- Haubold, B., and R. R. Hudson. 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. *Bioinformatics* **16**:847–849.
- Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination rate. *Genetics* **145**:833–846.
- Holmes, E. C., R. Urwin, and M. C. J. Maiden. 1999. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* **16**:741–749.
- Koehler, A., H. Karch, T. Beikler, T. F. Flemmig, S. Suerbaum, and H. Schmidt. 2003. Multilocus sequence analysis of *Porphyromonas gingivalis* indicates frequent recombination. *Microbiology* **149**:2407–2415.
- Lenormand, T., and S. Otto. 2000. The evolution of recombination in a heterogeneous environment. *Genetics* **156**:423–438.
- Levin, B. R. 1981. Periodic selection, infectious gene exchange and the genetic structure of *Escherichia coli* populations. *Genetics* **99**:1–23.
- Majewski, J., and F. M. Cohan. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* **148**:13–18.
- Martusewitsch, E., C. W. Sensen, and C. Schleper. 2000. High spontaneous mutation rate in the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by transposable elements. *J. Bacteriol.* **182**:2574–2581.
- McDonald, H., and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**:652–654.
- McVean, G., P. Awadalla, and P. Fearnhead. 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. *Genetics* **160**:1231–1241.
- Muller, H. J. 1964. The relation of recombination to mutational advance. *Mutat. Res.* **1**:2–9.
- Palys, T., E. Berger, I. Mitrica, L. K. Nakamura, and F. M. Cohan. 2000. Protein-coding genes as molecular markers for ecologically distinct populations: the case of two *Bacillus* species. *Int. J. Syst. Evol. Microbiol.* **50**:1021–1028.
- Papke, R. T., J. E. Koenig, F. Rodriguez-Valera, and W. F. Doolittle. 2004. Frequent recombination in a saltern population of *Halorubrum*. *Science* **306**:1928–1929.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Redfield, R. J. 2001. Do bacteria have sex? *Nat. Rev. Genet.* **2**:634–639.
- Rice, G., K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott, and M. J. Young. 2001. Viruses from extreme thermal environments. *Proc. Natl. Acad. Sci. USA* **98**:13341–13345.
- Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.
- Sarkar, S. F., and D. S. Guttman. 2004. Evolution of the core genome of *Pseudomonas syringae*, a highly clonal, endemic plant pathogen. *Appl. Environ. Microbiol.* **70**:1999–2012.
- Schleper, C., I. Holz, D. Janekovic, J. Murphy, and W. Zillig. 1995. A multicopy plasmid of the extremely thermophilic archaeon *Sulfolobus* effects its transfer to recipients by mating. *J. Bacteriol.* **177**:4417–4426.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin ver. 2.000: a software for population genetic analysis. Genetics and Biometry Laboratory, Geneva, Switzerland.
- Seitz, E., C. Haseltine, and S. Kowalczykowski. 2001. DNA recombination and repair in the archaea. *Adv. Appl. Microbiol.* **50**:101–169.
- Selander, R. K., R. M. McKinney, T. S. Whittam, W. F. Bibb, D. J. Brenner, F. S. Nolte, and P. E. Pattison. 1985. Genetic structure of populations of *Legionella pneumophila*. *J. Bacteriol.* **163**:1021–1037.
- She, Q., R. K. Singh, F. Confalonieri et al. (31 co-authors). 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **98**:7835–7840.
- Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**:1114–1116.
- Smith, J. M., E. J. Feil, and N. H. Smith. 2000. Population structure and evolutionary dynamics of pathogenic bacteria. *Bioessays* **22**:1115–1122.
- Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**:4384–4388.
- Souza, V., T. T. Nguyen, R. R. Hudson, D. Piñero, and R. E. Lenski. 1992. Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? *Proc. Natl. Acad. Sci. USA* **89**:8389–8393.
- Suerbaum, S., M. Lohrengel, A. Sonnevend, F. Ruberg, and M. Kist. 2001. Allelic diversity and recombination in *Campylobacter jejuni*. *J. Bacteriol.* **183**:2553–2559.
- Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* **43**:A7.
- Swofford, D. L. 1996. Paup*: phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel, and M. F. Polz. 2005. Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**:1311–1313.
- Turner, P. E., V. Souza, and R. E. Lenski. 1996. Tests of ecological mechanisms promoting stable coexistence of two bacterial genotypes. *Ecology* **77**:2119–2129.
- Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. Rubin, D. S. Rokhsar, and J. F. Banfield. 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**:37–43.
- Venter, J. C., K. Remington, J. F. Heidelberg et al. (23 co-authors). 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**:66–74.
- Whitaker, R. J., D. W. Grogan, and J. W. Taylor. 2003. Geographic barriers isolated endemic population of hyperthermophilic archaea. *Science* **301**:976–978.
- Zillig, W., A. Kletzin, C. Schleper, I. Holz, D. Janekovic, J. Hain, M. Lanzendoerfer, and J. K. Kristjansson. 1994. Screening for *Sulfolobales*, their plasmids and their viruses in Icelandic solfataras. *Syst. Appl. Microbiol.* **16**:609–628.

Edward Holmes, Associate Editor

Accepted July 25, 2005