EVOLUTION

Microbial Evolution in the Wild

A genetic study of microbes in an acid mine drainage site provide evolutionary rate estimates for wild microbe populations.

Edward F. DeLong

Microbial species thrive in every corner of Earth's biosphere. Their rapid growth rates and promiscuous gene swapping provide ample grist for the evolutionary mill over relatively short time spans. In ocean surface waters, for example, microbial doubling times of about 1 day result in an estimated production of around 10^{30} new cells per year (1). (This number exceeds that of all individual grains of sand on Earth by about 10 orders of magnitude.) These cells are not static entities, but rath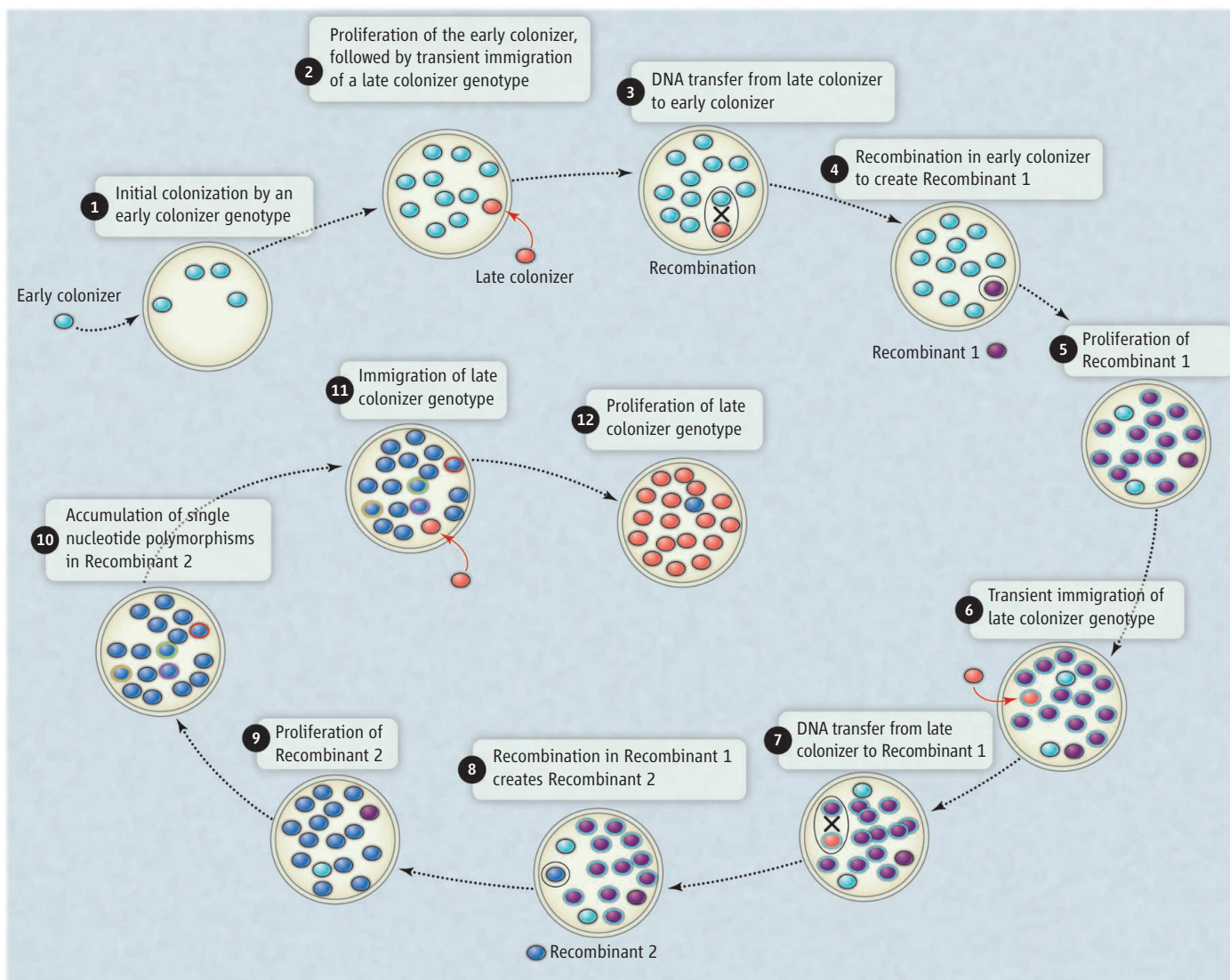er the products and perpetuators of dynamic evolutionary processes. But exactly how fast, to what degree, and by what mechanisms, do free-living microbes change and evolve over time in natural settings? On page 462 of this issue, Denef and Banfield (2) report evolutionary rate estimates from free-living microbial species in the wild.

Measuring evolutionary rates of individual species in complex natural microbial assemblages is a daunting task. Yet Denef and Banfield faced even more formidable challenges. The hot, humid, acidic, low-oxygen conditions found at their acid mine drainage (AMD) study site is, as the authors put it, "close to the limit of human endurance" (2). This would seem an unlikely place for conducting detailed evolutionary studies—but in some ways it is ideal. The AMD site harbors a very simple ecosystem containing only a handful of microbial species (3). *Leptospirillum*, a hardy and versatile bacterium that can live in sulfuric acid, eat iron, and fix carbon dioxide, often dominates AMD biofilms (3, 4). The very low diversity and complexity of the AMD biofilm allowed Denef and Banfield to directly follow the succession of genotypes in *Leptospirillum* populations over time, and to track recombination events and the accumulation of genomic nucleotide substitutions in populations collected over the course of 5 years (see the figure).

By comparing assembled genomes from different genotypes against total population DNA, the authors found that the AMD bio-

Departments of Biological Engineering and of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02138, USA. E-mail: delong@mit.edu

Microbial microevolution in a biofilm. This hypothetical scenario of evolutionary events in AMD biofilms illustrates some of the processes observed and inferred by Denef and Banfield. The white circles do not necessarily represent the same biofilm. The dotted black arrows do not imply direct succession in the same biofilm; rather, different amounts of time, and different spatial locations, may be involved in the transitions, which could take years to decades.

film populations contained six distinct *Leptospirillum* genotypes (types I to VI). The time series was dominated by *Leptospirillum* genotype III, which appeared to be a recombinant hybrid of genotypes I and VI. Time-series population DNA sequencing of *Leptospirillum* genotype III biofilms sampled over the course of 5 years yielded a single-nucleotide substitution rate of 1.4×10^{-9} substitutions per nucleotide per generation. This rate estimate, along with evolutionary history reconstructions, suggested that the six *Leptospirillum* genotypes diverged from a common ancestor in a matter of decades. The data also indicated that the major recombinations reflected single events that led to the domination of specific recombinant genotypes. Population genomic analyses suggested that *Leptospirillum* genotype evolution (see the figure) consisted of a complex menu of immigration, horizontal gene transfer (HGT), homologous recombination, fixation, and selective sweeps that successively generated and replaced different genotypes and, presumably, phenotypes.

The AMD *Leptospirillum* nucleotide substitution rate of 1.4×10^{-9} per nucleotide per generation is at the high end of previous estimates, which have ranged from 7.2×10^{-11} to 4.0×10^{-9} (5–7). One reason may be the use of a different approach. In perhaps the most elegant and comprehensive study of its kind to date, Denef and Banfield performed cultivation-independent population genomic analyses. In addition, they leveraged a unique combination of data that included in situ proteomic data sets (8), three distinct *Leptospirillum* population genotypes (assembled directly from natural populations) and a population genomic time series.

Unlike Denef and Banfield's study, other estimates of bacterial evolutionary rates have relied on isolation, propagation, or comparison of pure cultures or laboratory populations (5, 7, 9). These approaches may be selective and may not reflect native community complexity. On the other hand, the AMD habitat is itself unique; its unusual conditions and low biological complexity may affect the evolutionary rate estimates. Denef and Banfield point out, however, that their estimates are consistent with independent theoretical predictions of a universal mutation rate (10).

A central goal in microbial population biology is to identify key genomic changes

that lead to adaptation and selection—the "differences that make a difference" (*11–13*). Denef and Banfield speculate that phenotypic selection in the *Leptospirillum* populations may be due to differences of only a few nucleotides, but there are some caveats. Their approach captured major homologous recombination events, but nonhomologous gene acquisitions from HGT may have gone undetected, because the *Leptospirillum* population analyses relied on comparisons of genes present in reference genomes. Without complete and contiguous genome assemblies from the population genomes, large blocks of newly acquired foreign genes could be invisible to this approach.

Recent studies suggest that HGT and illegitimate recombination can be common and frequent in native microbial populations (*13–15*).

It turns out that "genome" is a verb, not a noun—a process, not a product. Current and emerging microbial population genomic studies that capture this dynamic (*2, 6, 7, 9, 13–15*) promise to help paint a more integrated motion picture of microbial ecology and evolution in action.

References
1. W. B. Whitman, D. C. Coleman, W. J. Wiebe, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6578 (1998).
2. V. J. Denef, J. F. Banfield, *Science* **336**, 462 (2012).
3. G. W. Tyson *et al*., *Nature* **428**, 37 (2004).
4. G. W. Tyson *et al*., *Appl. Environ. Microbiol.* **71**, 6319 (2005).
5. J. E. Barrick *et al*., *Nature* **461**, 1243 (2009).
6. N. A. Moran, H. J. McLaughlin, R. Sorek, *Science* **323**, 379 (2009).
7. L. Yang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7481 (2011).
8. V. J. Denef *et al*., *Environ. Microbiol.* **11**, 313 (2009).
9. T. D. Lieberman *et al*., *Nat. Genet.* **43**, 1275 (2011).
10. J. W. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7160 (1991).
11. G. Bateson, *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology* (Univ. of Chicago Press, Chicago, 1972).
12. C. Fraser, E. J. Alm, M. F. Polz, B. G. Spratt, W. P. Hanage, *Science* **323**, 741 (2009).
13. B. J. Shapiro *et al.,* *Science* **336**, 48 (2012).
14. J. H. Hehemann *et al*., *Nature* **464**, 908 (2010).
15. C. S. Smillie *et al*., *Nature* **480**, 241 (2011).

Sr-rich phases therefore add to the list of potential mineral biosignatures in the fossil record.

References and Notes

1. A. Herrero, E. Flores, Eds., *The Cyanobacteria: Molecular Biology, Genomics and Evolution* (Caister Academic, Wymondham, Norfolk, UK, 2008).
2. C. Jansson, T. Northen, *Curr. Opin. Biotechnol.* **21**, 365 (2010).
3. R. Riding, *Geobiology* **4**, 299 (2006).
4. W. Altermann, J. Kazmierczak, A. Oren, D. T. Wright, *Geobiology* **4**, 147 (2006).
5. J. Farquhar, A. L. Zerkle, A. Bekker, *Photosynth. Res.* **107**, 11 (2011).
6. G. Arp, A. Reimer, J. Reitner, *Science* **292**, 1701 (2001).
7. M. Merz, *Facies* **26**, 81 (1992).
8. N. Planavsky, R. P. Reid, T. W. Lyons, K. L. Myshrall, P. T. Visscher, *Geobiology* **7**, 566 (2009).
9. C. Dupraz *et al.*, *Earth Sci. Rev.* **96**, 141 (2009).
10. M. Obst *et al.*, *Geochim. Cosmochim. Acta* **73**, 4180 (2009).
11. M. R. Badger, D. Hanson, G. D. Price, *Funct. Plant Biol.* **29**, 161 (2002).
12. E. Kupriyanova *et al.*, *Microbiology* **153**, 1149 (2007).
13. See supplementary materials on *Science* Online.
14. J. Kaźmierczak *et al.*, *Facies* **57**, 543 (2011).
15. E. Couradeau *et al.*, *PLoS ONE* **6**, e28767 (2011).

In Situ Evolutionary Rate Measurements Show Ecological Success of Recently Emerged Bacterial Hybrids

Vincent J. Denef[1]* and Jillian F. Banfield[1,2]†

Few data are available on how quickly free-living microorganisms evolve. We analyzed biofilms collected from a well-defined acid mine drainage system over 9 years to investigate the processes and determine rates of bacterial evolution directly in the environment. Population metagenomic analyses of the dominant primary producer yielded the nucleotide substitution rate, which we used to show that proliferation of a series of recombinant bacterial strains occurred over the past few decades. The ecological success of hybrid bacterial types highlights the role of evolutionary processes in rapid adaptation within natural microbial communities.

Microbial communities, which drive Earth's geochemical cycles (*1*), can rapidly respond to change, but the proportion of this response that can be attributed to evolutionary processes, rather than species composition or gene expression shifts, remains an unresolved

question (2). Most evolutionary rate estimates are available for nucleotide substitution rates and derive from laboratory measurements (3, 4). It is difficult to know how relevant these rates are for geochemical environments, because studies on natural populations have been restricted to pathogens (5–7) and endosymbionts (8).

Our approach to measuring evolutionary rates for free-living bacterial populations involved tracking genome change in a time series of natural microbial community samples. Few systems are suitable for such a study; requirements include the availability of discrete, well-defined, and relatively homogeneous microbial community samples from a location where a reproducible community forms over time, with very restricted microbial input from other regions. The microbial communities catalyzing the formation of acid mine drainage (metal-rich, low-pH solutions) in the Richmond Mine (Iron Mountain, CA) provide such a system. Biofilms that develop at the air-solution interface on standing pools and slow-flowing underground streams have been used for more than a decade to study evolutionary processes and ecological complexity in nature (9). Key to the success of this environment as a model system has been its low species richness, which enables detailed culture-independent community genomics analyses of the genetic structure and dynamics of natural populations (10–13). Most biofilms within this system are dominated by the chemolithoautotrophic *Leptospirillum* group II. Previous metagenomic studies on two biofilms led to the reconstruction of genomes for *Leptospirillum* group II, types I and VI, the genes of which share ~94% average nucleotide identity (10, 13, 14). Inferences based on proteomics data indicate that large-scale homologous recombination events occurred between two parental *Leptospirillum* group II types, resulting in six recombinant (hybrid) genotypes (14, 15). One genotypic group, comprising types II through VI, predominates during initial colonization, whereas type I becomes abundant in later successional stages (16).

We have collected biofilms within the Richmond Mine over approximately 9 years. Sampling required field expeditions to locations where ambient conditions are often close to the limit of human endurance (e.g., 100% humidity, ~48°C, with depressed O_2 levels) at sites only accessible at times of the year when flow rates are low and danger from underground collapse is minimized. Working within these constraints, we collected microbial biofilms (both time series and spatially resolved) from six underground locations between March 2002 and December 2010 at intervals of 3 weeks to more than 1 year (Fig. 1, table S1, and

figs. S1 and S2) (17). We selected 13 samples from the C75 location, a site indicated by proteomics-based inferences to be typically dominated by the *Leptospirillum* group II, type III genotype (15). This allowed us to analyze single-nucleotide substitution accumulation over 5 years, and thereby to estimate the substitution rate. On the basis of the same proteomics-based inferences, we also selected nine biofilms that provided multiple samplings of most of the other *Leptospirillum* group II genotypes (types I, III, IV, V, and VI). Together with the two sequence data sets from which the type I and type VI reference genomes had been reconstructed, this allowed us to reconstruct a lineage history (fig. S1).

We generated community genomic data sets comprising ~63 billion base pairs of sequence from 24 samples, each of which contained ~10^8 *Leptospirillum* group II cells (10). Using sequences obtained from the first five C75 location biofilms, for which longer read lengths were available, we reconstructed a type III genome de novo (17). The genome is indeed a recombinant hybrid of the previously reconstructed type I and type VI genomes, as deduced from proteomic analysis (15). All identified recombination points were located within genes (table S2 and fig. S3). All *Leptospirillum* group II sequences were accounted for (fig. S4) (17), excluding the possibility of other genotypes in these five samples.

The sequencing reads from each of the 13 C75 samples collected over 5 years were recruited to the previously reconstructed *Leptospirillum* group II, type I and type VI genomes (13, 18) to identify

the recombinant block structure of the genome at each time point at this location. Only alignment of reads to the sequence with the highest similarity was permitted. Twelve biofilms contained only the type III genotype (>99% of cells); the remaining one contained ~84% type III and ~16% type I genotype (Fig. 1 and figs. S5 and S6A). Because 1.7 million to 8 million Illumina sequencing reads were obtained per sample (table S1), and because each read is an independent measurement of population-level sequence variation, we could unambiguously identify high-frequency single-nucleotide polymorphisms (SNPs) (>90% of all reads).

Nucleotide by nucleotide, we compared the type I–like segments of the type III genome to the type I genome sampled in 2002. Similarly, type VI–like regions of the type III genome were compared to the type VI genome sampled in 2005. In total, relative to the reference sequences, 64 mutations were fixed in all populations sampled at the C75 location (fig. S7A).

We then tracked genome change at the C75 location between August 2006 and December 2010. In the 11 C75 samples taken after August 2006, we identified up to six additional high-frequency substitutions relative to the 2006 type III genome (fig. S7A). Notably, the number of high-frequency mutations increased over time (fig. S8, $R^2 = 0.78$). Using this information, we calculated a rate of 1.4×10^{-9} (SE, ±0.2×10^{-9}) substitutions per nucleotide per generation (table S3). This number is in the higher range of previously reported bacterial genome-wide short-term substitution rates [7.2×10^{-11} to 4.0×10^{-9} substitutions per
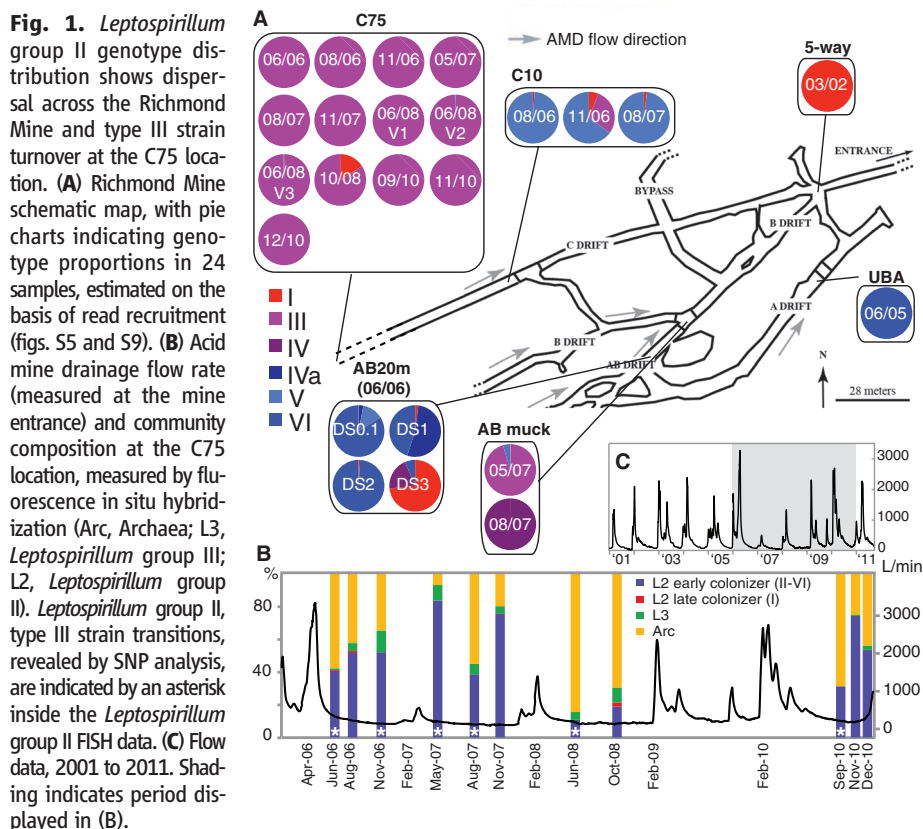
[1]Department of Earth and Planetary Science, University of California, Berkeley, CA 94720, USA. [2]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA.

*Present address: Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.
†To whom correspondence should be addressed. E-mail: jbanfield@berkeley.edu

Fig. 1. *Leptospirillum* group II genotype distribution shows dispersal across the Richmond Mine and type III strain turnover at the C75 location. (**A**) Richmond Mine schematic map, with pie charts indicating genotype proportions in 24 samples, estimated on the basis of read recruitment (figs. S5 and S9). (**B**) Acid mine drainage flow rate (measured at the mine entrance) and community composition at the C75 location, measured by fluorescence in situ hybridization (Arc, Archaea; L3, *Leptospirillum* group III; L2, *Leptospirillum* group II). *Leptospirillum* group II, type III strain transitions, are indicated by an asterisk inside the *Leptospirillum* group II FISH data. (**C**) Flow data, 2001 to 2011. Shading indicates period displayed in (B).

nucleotide per generation (4, 7, 8)]. It is similar to the rate of 1.3×10^{-9} substitutions per nucleotide per generation predicted for the type III genome size using the universal mutations-per-genome rate suggested by Drake (3). Although the highest rates have been reported for endosymbionts and pathogens, which have small effective population sizes, correspondence with Drake's prediction may indicate that genetic drift accelerated by population bottlenecks is not the main factor affecting the substitution rate in *Leptospirillum* group II.

To assess potential biases in the measured substitution rate introduced by spatial variation within the contiguous biofilms covering the C75 acid mine drainage pool, we sampled three loca-

tions ~1 m apart in June 2008. Samples 1 and 3 were dominated by the same genotype (with four SNPs relative to the type III genome), whereas sample 2 was dominated by a variant of this genotype lacking one high-frequency polymorphism (fig. S7A). The three samples also had only 93 (±8) replicated SNPs at frequency greater than 0.05 (average frequency 0.08). The low levels of variation, both within populations and across space, provide confidence that nucleotide changes in populations at the C75 location over time can be used to calculate the substitution rate.

Read recruitment of community genomic data from sites located across the underground system (five-way, UBA, C75, C10, AB muck, and AB20 locations) to types I and VI yielded

six genotypes (Fig. 1 and fig. S9): type I, type VI, and four other types (III, IV, IVa, and V) that each consist of a mosaic of type I and type VI genome blocks tens to hundreds of kilobases in length (Fig. 2, A and B, and figs. S6B and S9). This again confirmed previous strain-resolved proteomic analysis–based inferences (14, 15). The presence of exactly the same transition points between the recombinant blocks in each of the genotypes in every sample indicates that each major recombination event occurred in a single cell, whose descendants rose to fixation (Fig. 2). Each genotype's fixed mutations were identified by read recruitment to type III (Fig. 3 and fig. S7B) (17) and were used to construct a phylogenetic tree (Fig. 2C). Because read recruitment was performed
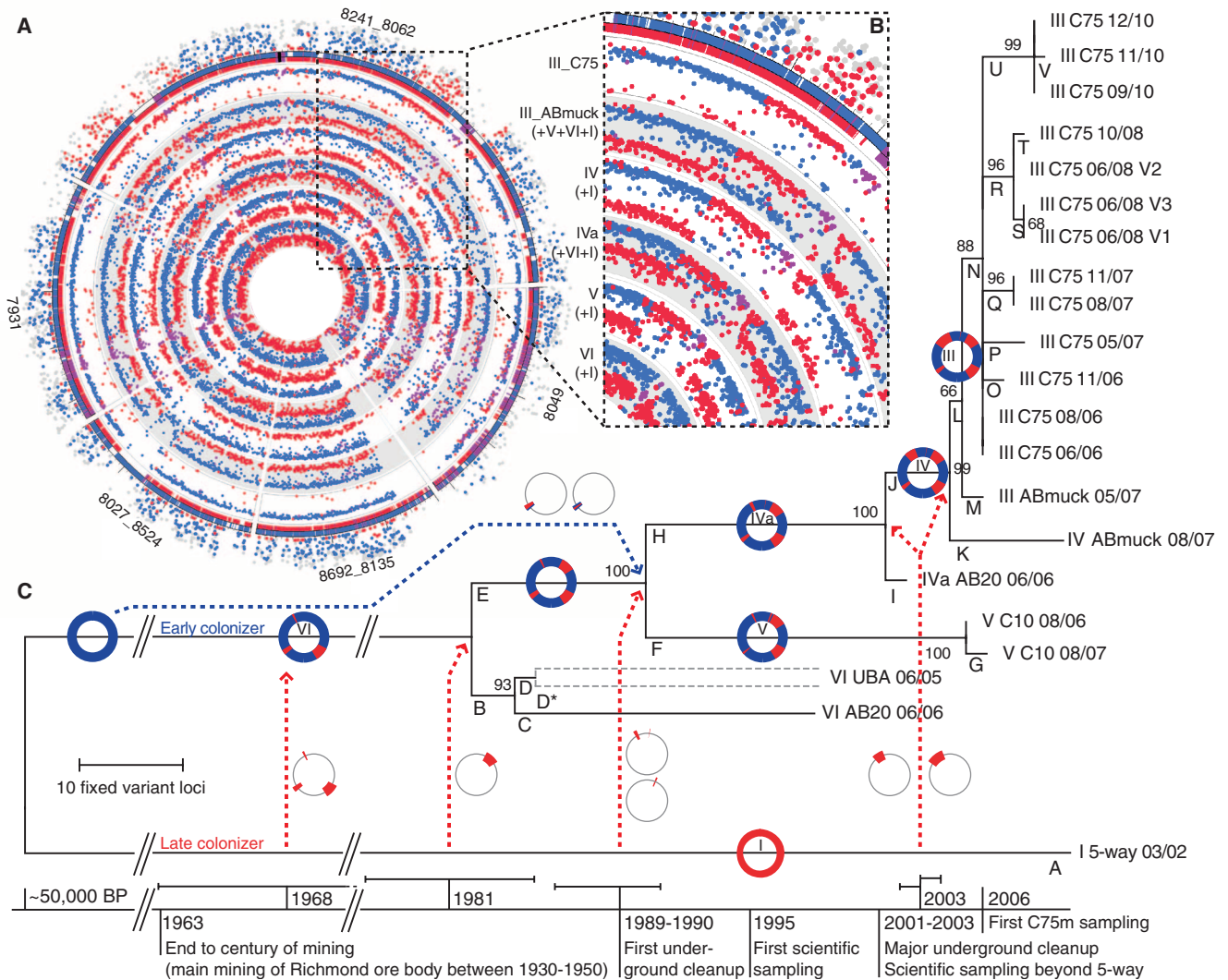

Fig. 2. Reconstruction of the recent evolutionary history of *Leptospirillum* group II. (**A**) The outer circle plots the counts of peptides unique to the type I (red) and type VI (blue) reference genomes at each protein locus (dots) for the C75 type III population [proteomics-inferred genotyping (PIGT); data from (15)]. Inner circles show Illumina read recruitment to the reference genomes, revealing the prevalence of type III, IV, IVa, V, and VI recombinants. (**B**) Correspondence of PIGT and recruitment data, highlighting the identical recombination point in genotypes III to V. (**C**) Evolutionary history of the sampled genotypes, based on

the variant loci (table S7), inferred using the maximum parsimony method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test is shown next to the branches. Dotted arrows indicate recombination events; circle schematics represent the regions affected. The timeline indicates the calculated time ranges of recombination events (table S3) as well as historical events (fig. S2) (18, 21). Branch D* is presented as two strains, each assigned half the total number of UBA 06/05 SNPs, because low incidence of SNPs precluded their linkage. BP, years before the present.

relative to the type III genome, positions within type I–like regions not shared by type IV, V, or VI genotypes were designated as gaps for tree construction, and these gaps were treated as missing data in the calculation of tree branch length. Notably, the tree topology indicates that the larger the fraction of type I–like sequence recombined into the type VI background, the more recently the specific genotype emerged (Fig. 2C).

Using the substitution rate calculated from the C75 time series and assuming equal generation times for the different genotypes, we estimated that the times to coalescence for the suite of recombinant genotypes ranged between 2 and 44 years (Fig. 2 and table S4). The causes of the success of these hybrid genotypes are hard to establish. In Bacteria and Archaea, little is known about the importance of hybridization in adaptation to change. However, one recent study hypothesized that agricultural practices led to proliferation of hybrids of *Campylobacter jejuni* and *Campylobacter coli* formed by homologous recombination (19). In Eukarya, hybrids are often reported to have little ecological success (20). However, increased fitness relative to parental types has been observed when the adaptive landscape changes as a result

of natural- or human-induced environmental alterations (21, 22). The Richmond Mine site is subject to both natural (seasonal and year-to-year variation; Fig. 1C) and human perturbations (fig. S2) (17), a subset of which may have provided opportunity for hybrid proliferation.

Although we cannot connect specific events to emergence of the hybrid genotypes, we can evaluate the role of neutral versus selection-based evolutionary processes in lineage divergence. For example, extreme population bottlenecks such as flushing events that remove biofilms during the rainy season may purge diversity independent of the fitness of particular strains. However, there are several lines of evidence indicating that hybrid genotypes arose as the result of selective processes. We detected evidence for positive selection (McDonald-Kreitman test, $P < 0.05$) for mutations specific to one type VI population (branch C; see Fig. 2), type V (branch F), and the ancestor of types III to V (branch E) (table S5). Selection was also supported by the disproportionately high number of signal transduction genes (Fig. 3, functional category T) and transcriptional regulation genes (Fig. 3, category K), global regulators in particular, that were affected by

fixed mutations (Fig. 2, branches E, F, D*, and H; see also tables S5 and S6). Previous observations in laboratory settings (23–25) and a natural system (16) similarly identified evolution of gene expression as an important factor underpinning early ecological divergence. Finally, an earlier study analyzing the distribution of the different *Leptospirillum* group II genotypes in the context of geochemical conditions indicated selection among type I through type VI recombinants (15). Evidence for selection also emerged from the current study, where two sites located ~140 m apart along a single flow path were compared (C10 and C75). The type III population, which dominated all upstream C75 biofilms, was found in only one of three biofilms sampled at the same time at the downstream C10 location. Even when the type III population was present at both sites, the genomes were different (Fig. 1 and fig. S10). This result suggests selection between genotypes that differ by only a few nucleotides—a finding consistent with prior laboratory studies showing increased fitness due to adaptive SNPs (4, 25). It also indicates that lack of diversity is not the explanation for dominance of many biofilms by a single genotype.

Fig. 3. Overview of the *Leptospirillum* group II high-frequency variants, based on read recruitment to the C75 June 2006 type III genotype. Dots indicate variant nucleotides relative to the last common ancestor of types III to VI. Functional categories (table S6) are indicated only for nonsynonymous substitutions and only for the first occurrence, starting from the most recent genotype (outside circle). Shading in inner circles indicates regions that were excluded for SNP analysis in a particular genotype because of recombination or indel events.
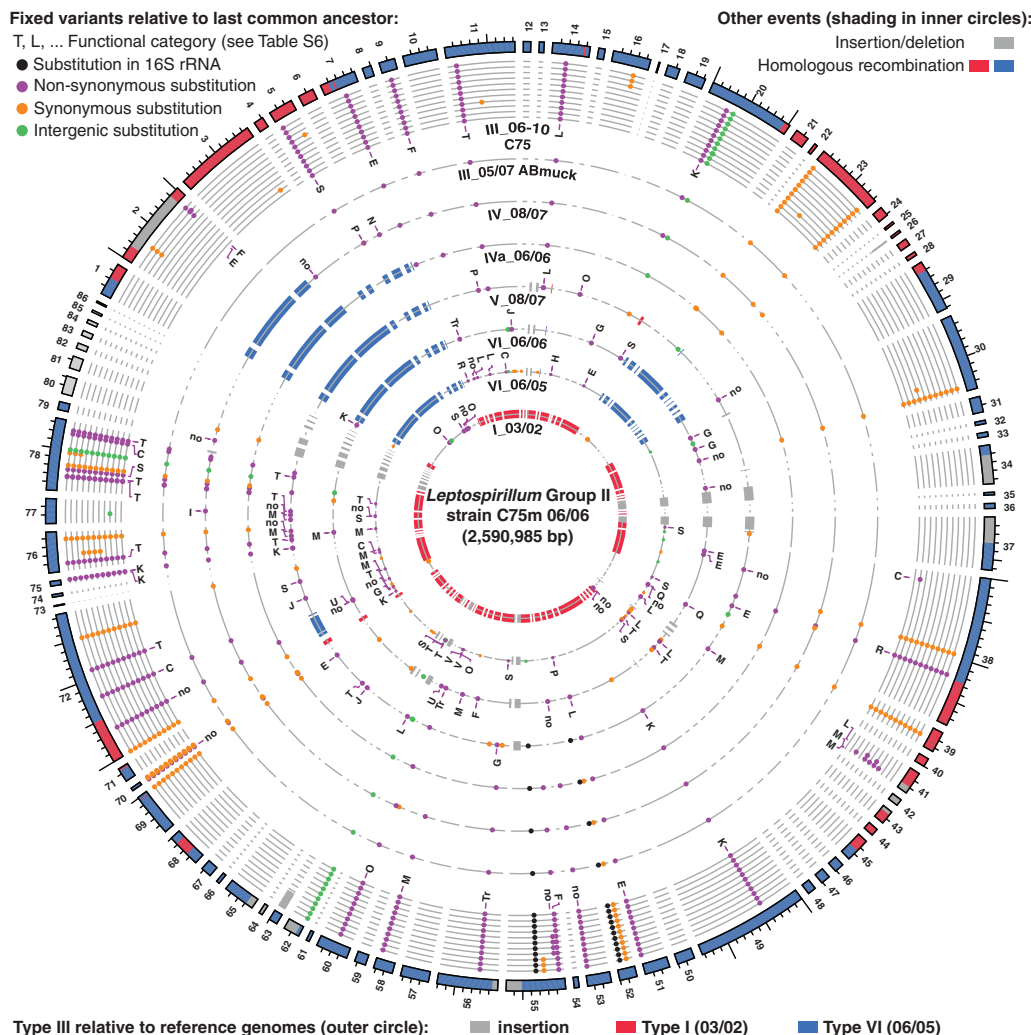
Our population genomic analyses provided data on the accumulation of genome change in free-living populations that underpin geochemical cycles. Application of these rates revealed that a series of important divergence events occurred over a time scale of years to decades within the natural model system of the Richmond Mine. We attribute these to major selection episodes, although we cannot conclude whether they were mediated by natural or human factors. Rapid adaptive evolution may have assisted in the maintenance of *Leptospirillum* group II as both the dominant primary producer and most active iron oxidizer responsible for acid mine drainage formation. Our results contribute to the development of a predictive understanding of how microbial systems respond to both natural and anthropogenic change (*2*).

References and Notes

1. P. G. Falkowski, T. Fenchel, E. F. Delong, *Science* **320**, 1034 (2008).
2. A. Reid, M. Buckley, *Microbial Evolution* (American Academy of Microbiology, Washington, DC, 2011).
3. J. W. Drake, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7160 (1991).
4. J. E. Barrick *et al.*, *Nature* **461**, 1243 (2009).
5. G. Morelli *et al.*, *PLoS Genet.* **6**, e1001036 (2010).
6. U. Nübel *et al.*, *PLoS Pathog.* **6**, e1000855 (2010).
7. L. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7481 (2011).
8. N. A. Moran, H. J. McLaughlin, R. Sorek, *Science* **323**, 379 (2009).
9. V. J. Denef, R. S. Mueller, J. F. Banfield, *ISME J.* **4**, 599 (2010).
10. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).
11. E. E. Allen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1883 (2007).
12. J. M. Eppley, G. W. Tyson, W. M. Getz, J. F. Banfield, *Genetics* **177**, 407 (2007).
13. S. L. Simmons *et al.*, *PLoS Biol.* **6**, e177 (2008).
14. I. Lo *et al.*, *Nature* **446**, 537 (2007).
15. V. J. Denef *et al.*, *Environ. Microbiol.* **11**, 313 (2009).
16. V. J. Denef *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2383 (2010).
17. See supplementary materials on *Science* Online.
18. D. S. A. Goltsman *et al.*, *Appl. Environ. Microbiol.* **75**, 4599 (2009).
19. S. K. Sheppard, N. D. McCarthy, D. Falush, M. C. J. Maiden, *Science* **320**, 237 (2008).
20. T. Hatfield, D. Schluter, *Evolution* **53**, 866 (1999).
21. B. R. Grant, P. R. Grant, *Ecology* **77**, 500 (1996).
22. A. P. Hendry *et al.*, *Proc. Biol. Sci.* **273**, 1887 (2006).
23. S. F. Elena, R. E. Lenski, *Nat. Rev. Genet.* **4**, 457 (2003).
24. T. F. Cooper, D. E. Rozen, R. E. Lenski, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1072 (2003).
25. A. Giraud *et al.*, *PLoS Genet.* **4**, e2 (2008).