

dated as 16.13 ± 0.12 ky B.P., 24.03 ± 0.37 ky B.P., 39.54 ± 0.38 ky B.P. and 48.39 ± 0.38 ky B.P., respectively, in the WD2014 chronology (25) [2σ uncertainty (8)]. Our proposed ages for Heinrich events 1 and 2 match the timing of maximum IRD observed in a stack of North Atlantic sediment records (26) (Fig. 2), and closely correspond to calibrated ^{14}C -based estimates (7). Despite these encouraging signs, we cannot be certain of the phasing of events. For example, the scenario of Northern Hemisphere cooling leading to CH_4 increase that we describe could potentially occur before the Heinrich event, maybe even triggering the event. This seems an unlikely possibility, given the requirement for another forcing mechanism, the abrupt nature of the CH_4 signals, and their unique occurrence in Heinrich stadials, but it cannot be ruled out.

The direct impact of a Heinrich event must cease once the ensuing interstadial commences and Northern Hemisphere climate warms (Fig. 1). We calculated maximum durations for the climatic impact of individual Heinrich events, perhaps largely on tropical hydrology, as ranging from 740 to 1520 years (Fig. 2). These estimates are significantly higher than the best guess for the duration of the actual iceberg armada (1 to 500 years) (7). We therefore suggest that this duration is related to a prolonged state of severely slowed or collapsed AMOC, maintained by extensive North Atlantic sea ice. This does not preclude the likelihood that AMOC was already relatively weak before Heinrich events, but rather explains why AMOC strength remained low through an extended stadial period when most coupled models suggest recovery of the overturning circulation within a few hundred years (19). The late HS plateaus in CH_4 concentration, particularly in HSs 4 and 5, are remarkably stable in comparison to interstadials (Fig. 1), suggesting limited variability in tropical precipitation and a relatively stable AMOC state. It is tempting to speculate on whether we capture information about the actual iceberg armada duration in the ~90- to 190-year CH_4 overshoots resolved during HSs 1, 4, and 5 (Fig. 2). However, short-duration CH_4 overshoots are not unique to these events and are in fact a feature of nearly all interstadial onsets, possibly resulting from an intrinsic response of the atmospheric system to rapid reorganization or intensive methanogenesis in newly inundated wetlands.

We suggest that our results provide motivation for climate modeling experiments to further investigate the difference between regular stadials and Heinrich stadials in the North Atlantic region and their respective impacts on tropical climate. Moreover, it is intriguing that only Heinrich events 1, 2, 4, and 5 are associated with Southern Hemisphere CH_4 signals. These four Heinrich events have relatively thick and spatially extensive sediment deposits, rich in detrital carbonate (Fig. 1) sourced from the Hudson Strait, and are linked to surging of the Laurentide ice sheet (7). Were the remaining Heinrich events too insignificant in scale to greatly perturb tropical climate, or did they not influence the same climate-sensitive location?

REFERENCES AND NOTES

- S. R. Hemming, *Rev. Geophys.* **42**, RG1005 (2004).
- P. U. Clark, S. W. Hostetler, N. G. Pisias, A. Schmittner, K. J. Meissner, in *Geophysical Monograph Series*, A. Schmittner, J. C. H. Chiang, S. R. Hemming, Eds. (American Geophysical Union, Washington, DC, 2007), vol. 173, pp. 209–246.
- G. Bond et al., *Nature* **365**, 143–147 (1993).
- W. S. Broecker, *Nature* **372**, 421–424 (1994).
- E. W. Wolff, H. Fischer, R. Röthlisberger, *Nat. Geosci.* **2**, 206–209 (2009).
- R. H. Rhodes et al., *Earth Planet. Sci. Lett.* **368**, 9–19 (2013).
- E. J. Brook, S. Harder, J. Severinghaus, E. J. Steig, C. M. Sucher, *Global Biogeochem. Cycles* **14**, 559–572 (2000).
- Materials and methods are available as supplementary materials on Science Online.
- J. Chappellaz et al., *Clim. Past* **9**, 2579–2593 (2013).
- Y. J. Wang et al., *Science* **294**, 2345–2348 (2001).
- S. A. Marcott et al., *Nature* **514**, 616–619 (2014).
- P. Kindler et al., *Clim. Past* **10**, 887–902 (2014).
- G. Deplazes et al., *Nat. Geosci.* **6**, 213–217 (2013).
- J. A. Collins et al., *Clim. Past* **9**, 1181–1191 (2013).
- S. A. Carolin et al., *Science* **340**, 1564–1566 (2013).
- J. Muller et al., *Quat. Sci. Rev.* **27**, 468–475 (2008).
- J. P. Severinghaus, R. Beaudette, M. A. Healy, K. Taylor, E. J. Brook, *Science* **324**, 1431–1434 (2009).
- J. C. H. Chiang, C. M. Bitz, *Clim. Dyn.* **25**, 477–496 (2005).
- M. Kageyama et al., *Clim. Past* **9**, 935–953 (2013).
- L. A. Parsons, J. Yin, J. T. Overpeck, R. J. Stouffer, S. Malyshev, *Geophys. Res. Lett.* **41**, 2013GL058454 (2014).
- M. Guillemin et al., *Clim. Past* **10**, 2115–2133 (2014).
- P. Bergamaschi et al., *J. Geophys. Res. Atmos.* **114**, D22301 (2009).
- B. P. Walter, M. Heimann, E. Matthews, *J. Geophys. Res. Atmos.* **106**, 34207–34219 (2001).
- I. Cvijanovic, P. L. Langen, E. Kaas, P. D. Ditlevsen, *J. Clim.* **26**, 4121–4137 (2013).
- C. Buizert et al., *Clim. Past* **11**, 153–173 (2015).

- J. V. Stern, L. E. Lisiecki, *Geophys. Res. Lett.* **40**, 3693–3697 (2013).
- A. Svensson et al., *Clim. Past* **4**, 47–57 (2008).
- J. Southon, A. L. Noronha, H. Cheng, R. L. Edwards, Y. Wang, *Quat. Sci. Rev.* **33**, 32–41 (2012).
- L. C. Kanner, S. J. Burns, H. Cheng, R. L. Edwards, *Science* **335**, 570–573 (2012).
- D. A. Hodell, J. E. Channell, J. H. Curtis, O. E. Romero, U. Röhl, *Paleoceanography* **23**, PA4218 (2008).
- X. Wang et al., *Nature* **432**, 740–743 (2004).

ACKNOWLEDGMENTS

This work was supported by NSF grants 1043518, 1142041, 0944552, 0839093, and 1142166. Data will be made available at <http://nsidc.org/data/> and <http://ncdc.noaa.gov/paleo/>. L. Layman, N. Chellman, D. Pasteris, M. Sigl, and C. Stowasser assisted with measurements. Thanks to J. Ahn for collated Siple Dome CH_4 data, J. Rosen for use of the Oregon State University firm air model, and T. Sowers for fruitful discussion. The authors appreciate the support of the WAIS Divide Science Coordination Office at the Desert Research Institute of Reno Nevada and the University of New Hampshire for the collection and distribution of the WAIS Divide ice core and related tasks (NSF grants 0230396, 0440817, 0944348, and 0944266). We are grateful to all participants in the field effort led by K. Taylor. The NSF Office of Polar Programs also funded the Ice Drilling Program Office and Ice Drilling Design and Operations group, the National Ice Core Laboratory, Raytheon Polar Services, and the 109th New York Air National Guard.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6238/1016/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 and S2
References (32–53)

2 October 2014; accepted 30 April 2015
10.1126/science.1262005

MICROBIAL DIVERSITY

Fine-scale diversity and extensive recombination in a quasisexual bacterial population occupying a broad niche

Michael J. Rosen,¹ Michelle Davison,² Devaki Bhaya,^{2*}† Daniel S. Fisher^{1,3*}†

Extensive fine-scale genetic diversity is found in many microbial species across varied environments, but for most, the evolutionary scenarios that generate the observed variation remain unclear. Deep sequencing of a thermophilic cyanobacterial population and analysis of the statistics of synonymous single-nucleotide polymorphisms revealed a high rate of homologous recombination and departures from neutral drift consistent with the effects of genetic hitchhiking. A sequenced isolate genome resembled an unlinked random mixture of the allelic diversity at the sampled loci. These observations suggested a quasisexual microbial population that occupies a broad ecological niche, with selection driving frequencies of alleles rather than whole genomes.

The traditional view of clonal bacterial populations in laboratory or clinical settings contrasts with the genetic diversity observed in environmental populations (1–4). Clusters of genotypes are often found (5, 6), and neutral drift alone could, in principle, account for them. However, even moderate rates of homologous recombination (HR) can prevent clustering on genomic scales (7), whereas without recombination, occasional periodic selection would eliminate the diversity. An alternative sce-

nario is that microbial species contain numerous subpopulations occupying discrete ecological niches, or ecotypes (8). In marine *Prochlorococcus*,

¹Applied Physics Department, Stanford University, Stanford, CA 94305, USA. ²Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA.

³Bioengineering Department, Stanford University, Stanford, CA 94305, USA.

*Corresponding author. E-mail: dbhaya@stanford.edu (D.B.); dsfisher@stanford.edu (D.S.F.). †These authors contributed equally to this work.

a recent study revealed the presence of multiple genomic backbones—i.e., clusters of core genomes associated with different ecological conditions and with low between-cluster rates of recombination (7). More extensive recombination is inferred in other populations. For marine *Vibrio* and thermophilic archaea (3, 4), a scenario has been proposed in which recombination is fast enough to enable genes to sweep through environments in which they confer fitness without purging genome-wide diversity. Ecological differentiation takes place when the colonization of a new microenvironment generates recombination barriers between subtypes, ultimately leading to distinguishable genomic clusters resembling incipient speciation (9). Here, we produce evidence that some microbial populations may maintain a cloud of fine-scale genotypic and phenotypic diversity, with varying selection acting on individual loci (and combinations of loci) but without subdivision into genome-wide subtypes. Such a population spread over a range of environmental conditions resembles a sexual population occupying a broad niche.

Studies of cyanobacteria in (*Synechococcus* sp.) the dense phototrophic biofilms (“microbial mats”) that develop along hot spring effluent channels in Yellowstone National Park establish that there are correlations between environmental gradients and genetic variation at several loci (10, 11). This is interpreted as evidence for the presence of ecotypes (12). Comparative genomics of two recently sequenced *Synechococcus* isolates (*Syn* OS-A and *Syn* OS-B’) from these biofilms reveals that they are closely related yet show extensive large and small genomic rearrangements (13). Metagenomic data corroborates this and also reveals evidence of recent horizontal gene transfer events (13). Furthermore, *Synechococcus* sp. are capable of natural transformation (14), have several active transposons (13, 15), and are host to cyanophages (16), all of which facilitate DNA exchange within closely related populations living in close proximity. This raises questions of whether the genome-scale linkage that should be associated with ecological subpopulations is detectable and whether genomic backbones can be identified in these populations.

Multilocus sequence typing (MLST) [the sequencing of a small number of loci from collections of (usually pathogenic) isolates] has been used extensively to identify variants (17) and, more recently, to analyze recombination (18). To examine the *Synechococcus* populations, we developed a variant of MLST in which we performed deep amplicon sequencing of multiple loci directly from the population (without isolating individuals) followed by correction of polymerase chain reaction (PCR) and sequencing errors (19). We focused on correlations within loci because they are more informative than the low-level linkage between loci when recombination rates are high. This protocol avoids the need for any potential biases associated with the isolation of multiple distinct clonal populations.

We used the previously sequenced *Syn* OS-A and *Syn* OS-B’ genomes and metagenomic data

to design primers optimized to maximally capture allelic diversity within the *Synechococcus* population at 90 genomic regions located along the genomes (20). Template DNA was isolated from

a microbial mat sample collected from a 50°C region of the effluent channel at Mushroom Spring (12). Both ends of the 90 amplicons (i.e., a total of 180 loci) were sequenced to an average depth of

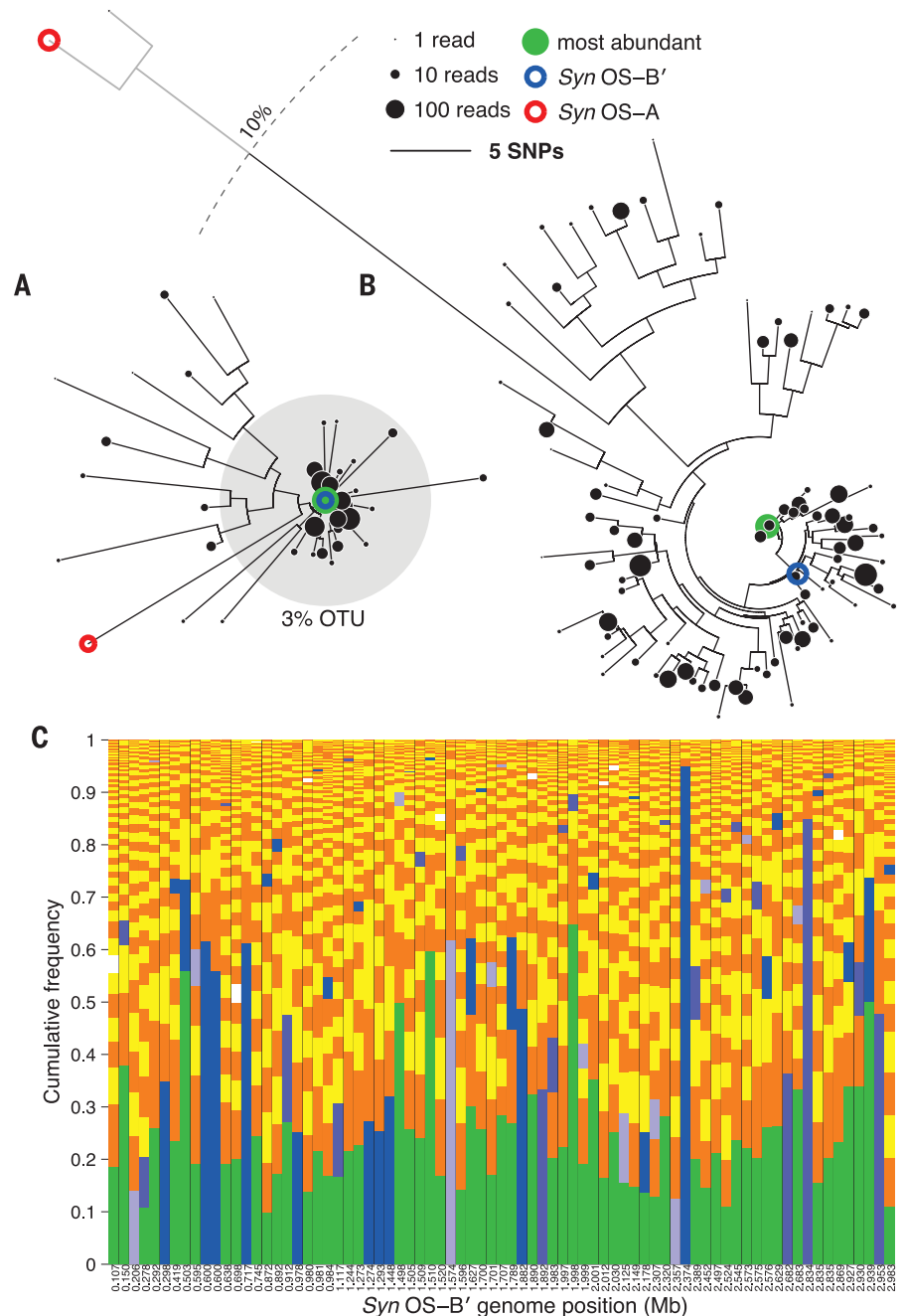


Fig. 1. Diversity of alleles. Neighbor-joining trees of alleles, with the leaf area being proportional to log of allele frequency. (A) 16S rRNA alleles. The gray circle highlights the 3% diameter cluster typically used to define an OTU designating a bacterial “species”. The genomes *Syn* OS-B’ (blue) and *Syn* OS-A (red) are shown. (B) A locus containing both genic and intergenic sequence (88 alleles derived from 2058 reads). *Syn* OS-B’ (blue) is identical to an allele with 11 reads, while *Syn* OS-A (red) is not among the alleles at this locus. The dashed gray line marks 10% sequence divergence from the most abundant allele (green). (C) The spectrum of allele frequencies for 77 loci is ordered along the *Syn* OS-B’ genome (left to right) and from most abundant to least abundant at each locus (bottom to top). The allele closest to *Syn* OS-B’ is blue when identical, white when >5 SNPs away, and intermediate shades of blue when 1 to 2 or 3 to 5 SNPs away. When not closest to *Syn* OS-B’, the most abundant allele is green, and all other alleles are colored as alternating yellow and orange.

1000x and trimmed to a uniform 400-nucleotide (nt) length (a total of ~65.4 Mb, the equivalent of >20x coverage of the 3.1 Mb *Syn* OS-B' genome). To reliably distinguish true allelic variation from PCR-generated errors (21), we applied an algorithm [DADA (19)] that exploits variation in sequence abundances and the distances between them to parametrize a simple error model and correct errors in the raw reads (see fig. S1 for a schematic overview). Consistency checks, particularly using the contrast between the statistics of nonsynonymous and synonymous single-nucleotide polymorphisms (SNPs) in the sample (the ratio d_n/d_s of the numbers of differences between alleles averages ~0.14) and those of PCR errors, which are agnostic to this difference, validated that the vast majority of inferred alleles, including many with only a single read, are real and cannot be ascribed to errors (20). Furthermore, comparison of the inferred alleles to sequenced bacterial genomes indicates that even these rare alleles almost all have best matches to *Syn* OS-A or *Syn* OS-B' (supplementary materials, text 1).

At the 16S ribosomal RNA (rRNA) locus (V1 to V3 segment, 438 nt), the most abundant allele was identical to the *Syn* OS-B' isolate, consistent with previous findings that *Syn* OS-B' has the dominant V7 to V8 16S allele at 56°C (12). Within the 3% sequence-divergence diameter operational taxonomic unit (OTU) centered on this allele, there were 20 other alleles, which comprised 93% of the *Synechococcus*-like reads. Farther away, we identified another 17 *Synechococcus* alleles, including a few reads identical to *Syn* OS-A, which is rare at 50°C but common at ~65°C (Fig. 1A) (13).

Other loci were more diverse. Focusing on the 136 loci with >300 reads (of the 180 sequenced), we found an average per site heterozygosity (the probability that a random pair of reads differ at a random site) of $\pi = 3.5\%$ (compare $\pi_{16S} = 0.5\%$) and an average of 44 alleles (range 7 to 121). Figure 1B shows a typical locus with 88 inferred nucleotide alleles. Each locus had a spectrum of allele abundances, from tens of percent to <0.1%. Figure 1C depicts these allele frequencies at 77 of the loci (one end of each amplicon with >300

reads) ordered along the *Syn* OS-B' genome. This graphical view shows how variable the patterns of diversity were across the genome. Furthermore, there is no obvious correlation between nearby loci, which provides a first hint that linkage may be very limited.

Recent recombination events between well-diverged alleles produce alleles with a characteristic "chimeric" structure that is easy to recognize (fig. S2). After finding the most likely pair of parents for each allele, even with conservative criteria, we observed >700 examples of chimeric alleles (table S1). An additional d_n/d_s consistency check (table S2) showed that most cannot be chimeras caused by PCR; for alleles one or two SNPs away from being a chimeric mixture of two parent alleles, these SNPs had a strong synonymous skew relative to the expectation from errors (22). The large number of chimeric alleles suggested that homologous recombination is common, but we could not estimate the rate at which it occurs because recombination events between highly similar parents, or far

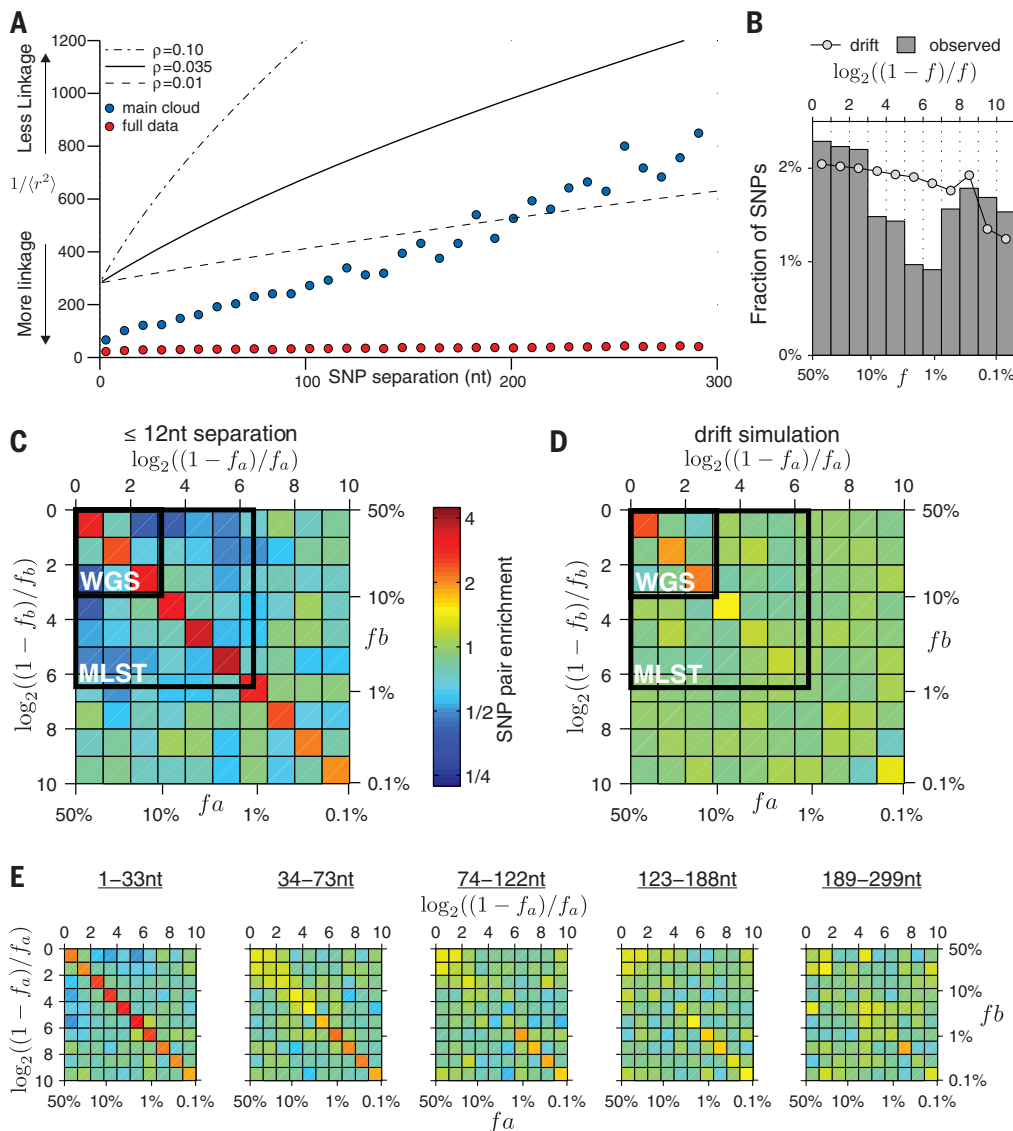


Fig. 2. Frequency statistics and linkage correlations between pairs of SNPs. (A) The inverse of the average linkage correlations, $\langle r^2 \rangle = \langle (f_{ab} - f_a f_b)^2 / [f_a(1-f_a)f_b(1-f_b)] \rangle$ (20), are shown as a function of SNP separation. Red, full data; blue, the main cloud, including only alleles within $\leq 10\%$ of the most abundant allele. The drift expectations, with an observed heterozygosity of $\pi = 0.03$ and three values of $\rho = 0.01$ (a rough best fit), 0.035 (giving a similar slope), and 0.1 (giving a similar fractional increase), generated numerically with ER2, are shown for comparison (26). (B) SFS for the synonymous SNPs in the main cloud. The roughly log-spaced frequency bins are chosen so that the neutral-drift expectation (dots) is flat (cutoff by the read depth). (C to E) Two-dimensional histograms of the frequencies of pairs of SNPs from alleles in the main cloud; each bin shows on a logarithmic color scale the ratio of the number of SNP pairs with minority allele frequencies f_a and f_b , and the expected number if the frequencies were drawn independently from the SFS. (C) Tightly linked SNP pairs are separated by at most 12 nt, with (D) an asexual drift simulation for comparison; (E) data for several ranges of separations between the SNPs (with ~6400 pairs in each window). The frequency ranges that could be studied with the 10x and 100x depths typical of WGS and MLST, respectively, are indicated in (C) and (D).

in the past (and hence obscured by subsequent mutations and recombinations), are not possible to identify directly. We therefore turned to statistical methods to probe recombination quantitatively.

If recombination rates are comparable to mutation rates, inferring recombination events relative to an asexual phylogeny, as often done for MLST and whole-genome sequencing (WGS) data (23, 24), is not feasible. Likelihood-based methods (25) can be used to infer recombination parameters without an underlying phylogeny. However, these methods are based on the assumptions of a well-mixed population with neutral drift and unrestricted recombination and thus can obscure potentially informative deviations from model assumptions. Extensive deep sequencing data allowed us to avoid the pitfalls of fitting to particular models.

The >7500 synonymous SNPs that we observed serve as convenient neutral markers. Within loci were >50,000 pairs of these SNPs. We studied the dependence of the frequency statistics on the within-pair separation, x . Each pair has four haplotypes, AB (the most abundant), Ab , aB , and ab , with frequencies f_{AB} , f_{Ab} , f_{aB} , and f_{ab} . If x is large, and thus recombination frequent enough to unlink the two sites, then the frequencies will become independent—i.e., $f_{ab} \approx f_a f_b$ (with $f_a = f_{ab} + f_{aB}$ the frequency of SNP a). But if the sites are close together, systematic deviations from this linkage equilibrium are likely.

A measure of the linkage disequilibrium is r^2 , with r normalized to be unity with full linkage—

i.e., when only AB and ab are observed. Figure 2A shows the inverse of r^2 as a function of separation x . For the complete data set, linkage decayed only slightly with distance (Fig. 2A, red circles). However, this masked the behavior of the large majority of the population because outliers dominated the average linkage. Considering only alleles within some cutoff distance from the most abundant, a more rapid decay of r^2 with separation was observed as the cutoff was decreased. Including only the alleles within 10% of the most abundant (corresponding to 92% of the reads, coincidentally the same fraction as the 3% diameter OTU from the 16S rRNA) characterized well the behavior within an ad hoc “main cloud” (fig. S3A). We also analyzed metagenomic data from a higher temperature sample (13), finding similar behavior (fig. S3B): linkage decayed comparably rapidly, although within a narrower (5% cutoff) cloud. These metagenomic data were much shallower, but because the DNA was not amplified and thus not subject to potential PCR-generated recombination artifacts, they provided an independent confirmation.

Within the main cloud of the amplicon data, linkage dropped by a factor of 10 over 300 nt (Fig. 2A) (σ_d^2 , a related measure of linkage, behaves similarly, as defined and shown in fig. S4). As $1/(r^2)$ is roughly proportional to the probability that a recombination event has unlinked the sites since the SNPs occurred (26), the slope of this curve is related to $\rho = RT$, where R is the rate of creation of a recombination break point between neighboring sites and T is the average time

since the most recent common ancestor. The data thus suggest that ρ is in the range 10^{-2} to 10^{-1} , which is at the very high end of values inferred for free-living bacteria (27) and similar to some pathogens (28, 29). Because the synonymous heterozygosity, π_s , is determined by the product of T and the mutation rate per site, μ , and was found to be ~3% in the main cloud, this implies that the mutation rate and recombination rates are about the same, i.e., $R \sim \mu$. At larger separations than the typical length δ of the DNA fragments exchanged, the decay of linkage should saturate. That it did not saturate by $x = 300$ nt suggests that δ is considerably longer than this, consistent with inferences for other bacteria (28, 30).

With abundant data extending to low frequencies, more can be gleaned from the SNP pair statistics than ρ . Nearby pairs [with $\delta \gg x$] within the main cloud (Fig. 2A) were more linked than expected for a population driven solely by neutral drift and random recombination (26). To explore this excess linkage, we studied the joint distribution of the SNP frequencies f_a and f_b . For the main cloud, the single-site SNP frequency spectrum (SFS) (Fig. 2B) had a roughly similar form to the drift expectation, albeit with a dip in the 1 to 10% frequency range, (see fig. S5A for the full data), but the frequency correlations of nearby SNP pairs differed strikingly from drift. Over a broad range of frequency f_a , in the histogram of Fig. 2C, there was a strong excess of SNPs pairs along the diagonal (i.e., with $f_a \sim f_b$), relative to the expectation under the independence of the two SNPs. This contrasted with asexual drift (Fig. 2D), for which only a small excess would occur and only at high frequencies. Much of the excess correlation, as well as the additional linkage in Fig. 2A, was caused by an abundance of AB - ab pairs in the absence of other combinations, so that $f_{ab} = f_a = f_b$.

Over time, correlations between allele frequencies are, like linkage, broken up by recombination, although only indirectly (from AB and ab pairs, recombination must first produce Ab and aB before fluctuations in allele frequencies can break up the correlation between f_a and f_b). Thus, the frequency correlations should decay with separation. In Fig. 2E, the joint-frequency histograms for increasing separation ranges exhibited this decay, but the correlations persisted over larger separations at lower frequencies, consistent with low-frequency SNPs being, on average, younger (see fig. S5, C and D, for a comparison with full data and drift simulations). Together, the data in Fig. 2, A and E, show that by 300 base pairs (bp) the four pair-haplotype frequencies were close to being independent. This is strong evidence against the presence of subpopulations within the main cloud, as these would manifest as a persistence of frequency correlations.

Despite the weak decay of linkage in the full data, the probability of observing recently formed chimeric alleles with one parent inside the main cloud and the other outside is consistent with recombination occurring at similar rates inside and outside the main cloud (table S3). The observed persistence of linkage to longer distances

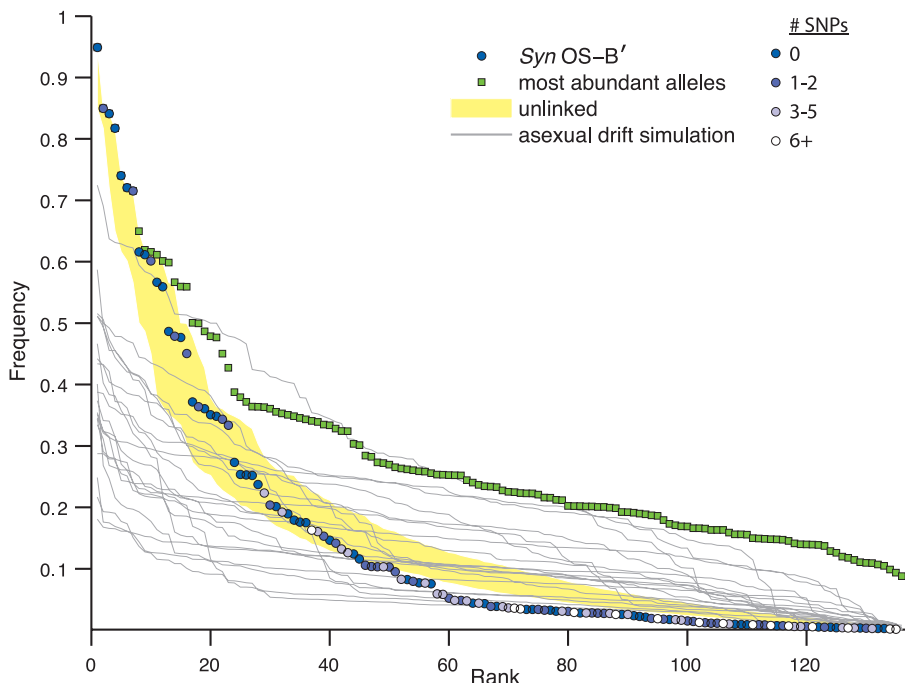


Fig. 3. Syn OS-B' as a random mixture of the allelic diversity. The rank-frequency spectrum of the alleles most similar to the Syn OS-B' genome at each locus (blue-white scale from Fig. 1). For comparison, two simple models of genomes are shown: an unlinked, highly recombinant model [the yellow band contains the central 95% of the frequencies at each rank across 10^5 simulations (20)] and 20 samples from a neutral asexual model with $\theta = 0.02$ chosen to match the per-site heterozygosity of the data [gray lines (20)]. The green points show the frequencies of the most abundant alleles.

could instead result from selection against recombinants with the outliers because of epistatic incompatibilities (31) suppressing older recombination events and increasing linkage. However, we cannot rule out some spurious linkage being produced by biased sampling from the choice of PCR primers or other effects.

At least within the main cloud, the precipitous decay of linkage within loci, without evidence of saturation, indicated that there might be very little residual linkage at genomic distance scales. Genomic-scale linkage has been studied by sequencing whole genomes (1, 3, 4), which requires multiple isolates and can be technically challenging. We chose an alternate approach in which allele frequency information from the amplicon data were combined with the extended linkage information of the *Syn* OS-B' genome.

We asked what scenario for the population best describes the relationship of the *Syn* OS-B' genome to the observed allelic diversity. If the population were predominantly clonal, a typical genome of the dominant 16S ribotype, such as *Syn* OS-B', would tend to be similar to high-frequency alleles in other genomic regions; however, at 23 loci, *Syn* OS-B' is not even in the main cloud. More broadly, the alleles closest to *Syn* OS-B' exhibited a wide range of distances, indicated by the shades of blue in Fig. 1C and abundances, which are shown rank-ordered in Fig. 3. Thus, the *Syn* OS-B' genome does not represent a dominant genotype (see the distribution of top allele abundances shown in green), but that does not exclude the possibility of a dominant genomic backbone as seen in *Prochlorococcus* (1). To assess this, we compared two scenarios. At one extreme, we considered the simplest asexual model: a neutrally drifting population with *Syn* OS-B' a random member. Such rank-frequency spectra are widely variable (gray lines in Fig. 3) but generally rather flat because of the correlations between loci arising from a common phylogeny. At the other extreme, we considered a sexual population that is completely unlinked from one locus to the next, with genomes in which each locus consists of a population-frequency-weighted random allele, as shown by the yellow region in Fig. 3. In this latter case, there was little variability in the rank-frequency spectra. This extreme sexual model was a far better fit for *Syn* OS-B', particularly at high frequencies (this is also true for *Syn* OS-A) (fig. S6). We concluded that this cyanobacterial population had undergone sufficiently extensive homologous recombination among close relatives (comprising at least 90% of the population) to make it similar to a sexual population, in which genomes consist of random mixtures of alleles.

The observation of rapidly decaying linkage is a feature that has often been used to infer high rates of recombination via fits to a drift model (32). Our statistical analysis of pairs of neutral SNPs simultaneously evaluated the drift model itself, revealing considerable departures that suggested distinctive dynamics. Selection on other parts of the genome, triggered by environmental or ecological changes, new beneficial alleles, or

other factors, can cause large changes in SNP frequencies before recombination unlinks them from the selected region. Recurrent hitchhiking events with a range of magnitudes can lead to single-SNP frequency spectra that are similar to drift, such as that observed. But the consequences for joint-frequency spectra of nearby pairs of SNPs can be very different. For example, in a population strongly dominated by *AB*, a single hitchhiking event could rapidly raise the *ab* frequency while leaving *AB* and *Ab* rare, in contrast to drift. Such events can give rise to the striking diagonal feature in the joint-frequency histogram evident in Fig. 2C and the large excess linkage at short distances shown in Fig. 2A.

Without the depth of sequencing and error correction used in our study that enabled low-frequency variants to be identified, the striking departures from neutral drift could not have been identified. In Fig. 2, C and D, black squares show the more limited range of frequencies typical of MLST and even more of current WGS. Going forward, for deep microbial population sequencing to achieve its potential for understanding underlying evolutionary dynamics, a range of scenarios, including recombination and recurrent hitchhiking with migration between local populations, needs to be explored. Theoretical predictions of correlations in multiple joint distributions and an understanding of which of these can best distinguish between scenarios are sorely needed.

The existence and persistence of extensive fine-scale microbial diversity is particularly puzzling in a localized population in which multiple variants have persisted for $T > 10^6$ generations (corresponding to the observed π and assuming a typical bacterial mutation rate of $<10^{-8}$). Based on theoretical modeling (8) and sequencing of a small number of loci (12), this diversity has previously been attributed to the existence of multiple discrete microniches occupied by ecotypes, with typical diameters $<2\%$ (11). But our results are inconsistent with the presence of multiple ecotypes, at least within the 20% diameter main cloud. This cyanobacterial population is thus a striking contrast to *Prochlorococcus*, for which multiple whole-genome ecotypes have been found each with $\sim 1\%$ diameter (1). We conjecture that in this dense biofilm with multiple local environmental gradients and ubiquitous recombination, the microbial population behaves as an effectively sexual species in which there are no discrete ecotypes or barriers to gene flow. Rather, it is a single population occupying a broad environmental niche, with spatially varying adaptation maintaining diversity at the scale of genes. For instance, *Syn* OS-A and *Syn* OS-B' contain different suites of genes encoding proteins required for the use of urea (13) and for phosphonate metabolism (33). These differences could have arisen as a result of varying selection pressures along the gradients that exist within the biofilms. Temporally varying selection at some loci could drive the dynamics of the rest of the genome by hitchhiking, with high rates of recombination preventing genomic sweeps that would purge the diversity.

The implications of such a scenario should motivate future studies of fine-scale microbial diversity in biofilms and other localized populations where the effects of nutrient, gas, and light gradients, fluctuations such as phage infections, and migration can be explored.

REFERENCES AND NOTES

- N. Kashtan et al., *Science* **344**, 416–420 (2014).
- Z. I. Johnson et al., *Science* **311**, 1737–1740 (2006).
- B. J. Shapiro et al., *Science* **336**, 48–51 (2012).
- H. Cadillo-Quiroz et al., *PLOS Biol.* **10**, e1001265 (2012).
- O. X. Cordero, M. F. Polz, *Nat. Rev. Microbiol.* **12**, 263–273 (2014).
- W. P. Hanage, C. Fraser, B. G. Spratt, *Phil. Trans. R. Soc. B* **361**, 1917–1927 (2006).
- C. Fraser, W. P. Hanage, B. G. Spratt, *Science* **315**, 476–480 (2007).
- F. M. Cohan, E. B. Perry, *Curr. Biol.* **17**, R373–R386 (2007).
- M. F. Polz, E. J. Alm, W. P. Hanage, *Trends Genet.* **29**, 170–175 (2013).
- E. D. Becraft, F. M. Cohan, M. Köhl, S. I. Jensen, D. M. Ward, *Appl. Environ. Microbiol.* **77**, 7689–7697 (2011).
- M. C. Melendrez, R. K. Lange, F. M. Cohan, D. M. Ward, *Appl. Environ. Microbiol.* **77**, 1359–1367 (2011).
- D. M. Ward et al., *Phil. Trans. R. Soc. B* **361**, 1997–2008 (2006).
- D. Bhaya et al., *ISME J.* **1**, 703–713 (2007).
- M. R. Gomez-Garcia, F. Fazeli, A. Grote, A. R. Grossman, D. Bhaya, *J. Bacteriol.* **195**, 3309–3319 (2013).
- W. C. Nelson, L. Wollerman, D. Bhaya, J. F. Heidelberg, *Appl. Environ. Microbiol.* **77**, 5458–5466 (2011).
- J. F. Heidelberg, W. C. Nelson, T. Schoenfeld, D. Bhaya, *PLOS One* **4**, e4169 (2009).
- D. Gevers et al., *Nat. Rev. Microbiol.* **3**, 733–739 (2005).
- X. Didelot, M. C. J. Maiden, *Trends Microbiol.* **18**, 315–322 (2010).
- M. J. Rosen, B. J. Callahan, D. S. Fisher, S. P. Holmes, *BMC Bioinformatics* **13**, 283 (2012).
- Materials and methods are available as supplementary materials on Science Online.
- V. Kunin, A. Engelbrektson, H. Ochman, P. Hugenholtz, *Environ. Microbiol.* **12**, 118–123 (2010).
- R. P. Smyth et al., *Gene* **469**, 45–51 (2010).
- X. Didelot, D. Falush, *Genetics* **175**, 1251–1266 (2007).
- M. Vos, X. Didelot, *ISME J.* **3**, 199–208 (2009).
- G. McVean, P. Awadalla, P. Fearnhead, *Genetics* **160**, 1231–1241 (2002).
- Y. S. Song, J. S. Song, *Theor. Popul. Biol.* **71**, 49–60 (2007).
- J. R. Doroghazi, D. H. Buckley, *ISME J.* **4**, 1136–1143 (2010).
- K. A. Jolley, D. J. Wilson, P. Kriz, G. McVean, M. C. J. Maiden, *Mol. Biol. Evol.* **22**, 562–569 (2005).
- X. Didelot, G. Méric, D. Falush, A. E. Darling, *BMC Genomics* **13**, 256 (2012).
- G. Morelli et al., *PLOS Genet.* **6**, e1001036 (2010).
- J. Majewski, *FEMS Microbiol. Lett.* **199**, 161–169 (2001).
- S. Takuno, T. Kado, R. P. Sugino, L. Nakhleh, H. Innan, *Mol. Biol. Evol.* **29**, 797–809 (2012).
- M. R. Gomez-Garcia, M. Davison, M. Blain-Hartnug, A. R. Grossman, D. Bhaya, *ISME J.* **5**, 141–149 (2011).

ACKNOWLEDGMENTS

We thank J. Blundell and B. Callahan for valuable suggestions and comments. Funding for this work was provided by NSF grants DMS-1120699 (D.S.F.), MCB-1024155 (D.B.), and FIBR EF-0328698 (D.B.), and by the Carnegie Institution for Science (D.B.). M.J.R. was partially supported by a Stanford Graduate Fellowship and an IBM Fellowship. Raw sequences have been deposited in National Center for Biotechnology Information (BioProject PRJNA268121).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/348/6238/1019/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S4
References (34–46)

8 December 2014; accepted 16 April 2015
10.1126/science.aaa4456

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of June 9, 2015):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/348/6238/1019.full.html>

Supporting Online Material can be found at:

<http://www.sciencemag.org/content/suppl/2015/05/27/348.6238.1019.DC1.html>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/content/348/6238/1019.full.html#related>

This article **cites 45 articles**, 23 of which can be accessed free:

<http://www.sciencemag.org/content/348/6238/1019.full.html#ref-list-1>

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/348/6238/1019.full.html#related-urls>

This article appears in the following **subject collections**:

Microbiology

<http://www.sciencemag.org/cgi/collection/microbio>