Review
TRENDS in Ecology and Evolution Vol.21 No.9
Full text provided by www.sciencedirect.com
ScienceDirect

Microbial ecology

Population genomics in natural microbial communities

Rachel J. Whitaker[1,2] and Jillian F. Banfield[3]

[1] Ecosystem Sciences, 137 Mulford Hall, University of California, Berkeley, CA 94720–3114, USA
[2] Current address: Department of Microbiology, University of Illinois, 601 South Goodwin Ave, Urbana, IL 61801, USA
[3] Department of Earth and Planetary Sciences and Department of Environmental Science, Policy, and Management, 369 McCone Hall, University of California, Berkeley, CA 94720–4767, USA

Little is known about the evolutionary processes that structure and maintain microbial diversity because, until recently, it was difficult to explore individual-level patterns of variation at the microbial scale. Now, community-genomic sequence data enable such variation to be assessed across large segments of microbial genomes. Here, we discuss how population-genomic analysis of these data can be used to determine how selection and genetic exchange shape the evolution of new microbial lineages. We show that once independent lineages have been identified, such analyses enable the identification of genome changes that drive niche differentiation and promote the coexistence of closely related lineages within the same environment. We suggest that understanding the evolutionary ecology of natural microbial populations through population-genomic analyses will enhance our understanding of genome evolution across all domains of life.

Introduction

Over the past 30 years, the development of culture-independent, PCR-based molecular tools to amplify and sequence genes directly from environmental samples has unveiled an astounding array of microbial biodiversity [1,2]. As molecular tools become ubiquitous in environmental microbiology, nearly every environment examined has uncovered previously unrecognized genetic variants. The patterns of genetic variation found in these molecular surveys using single genes have a predictable structure: clusters of very closely related sequences at the tips of phylogenetic trees separated by relatively long branches [3] (Figure 1). In the microbial world, the ecological and evolutionary significance of this pattern of genetic diversity is not well understood. What evolutionary mechanisms drive divergence among clusters of sequences? Does every sequence cluster represent an ecologically unique species? What types of differential adaptations enable closely related clusters of organisms to coexist within a single environment? Answering these questions requires identifying the evolutionary and ecological mechanisms that structure diversity in microorganisms. These mechanisms can best be determined through population-genomic analysis of patterns of individual-level variation occurring within natural populations. Once these mechanisms are better understood, the ecological significance of microbial molecular diversity will be easier to integrate into models of ecosystem function.

Multilocus sequence typing (MLST) [4] of housekeeping genes (see Glossary) has provided some clues about the mechanisms that promote the patterns of diversity observed in molecular surveys of single genes. These studies suggest that the diversity of Bacteria and Archaea is structured [5,6] through mechanisms that are analogous to allopatric [7], ecological [8,9] and biological speciation [10,11] in macroorganisms. However, because MLST studies focus mainly on a few genome markers only, they describe the structure of diversity without identifying genome-level changes that drive ecological differentiation

Glossary

Allopatric speciation: speciation resulting from divergent evolution of populations that are geographically isolated from one another [20].
Biological speciation: speciation that results from intrinsic barriers to recombination. The definition is based on the biological species concept for sexual macroorganisms in which species are defined by reproductive isolation [20].
Community genomics: direct genome sequencing of communities of microorganisms from environmental samples. Many authors use metagenomics to refer to the same emerging field. We find this term imprecise and prefer the term 'community genomics' because we do not believe that communities have a (meta) genome. Reference to 'community' emphasizes the sequencing of an assemblage of genomes of different evolutionary and ecologically distinct organisms that co-occur.
Coverage: the average number of times each base pair in the genome is sequenced.
Ecological speciation: speciation caused by divergent natural selection on traits between populations (or disruptive selection between phenotypes of a single population) in different environments (including use of different resources) [20].
Housekeeping genes: core genes whose function is fundamental to the cell [4].
Individual-level variation: variation among individuals within a species. This type of variation is the focus of population genetics.
Microbial community: the sum of microbial populations co-occurring within a single environment.
Microbial population: a genetically cohesive assemblage of individuals within a single species.
Paired-end sequences: sequences from both ends of the same cloned genome fragment. The genetic distance between paired sequences is estimated as the average insert size for the clone library.
Population genomics: population genetic analysis of individual-level variation across the entire genome.

Corresponding author: Whitaker, R.J. (rwhitaker@life.uiuc.edu)
Available online 20 July 2006.

www.sciencedirect.com 0169-5347/$ – see front matter © 2006 Elsevier Ltd. All rights reserved. doi:10.1016/j.tree.2006.07.001

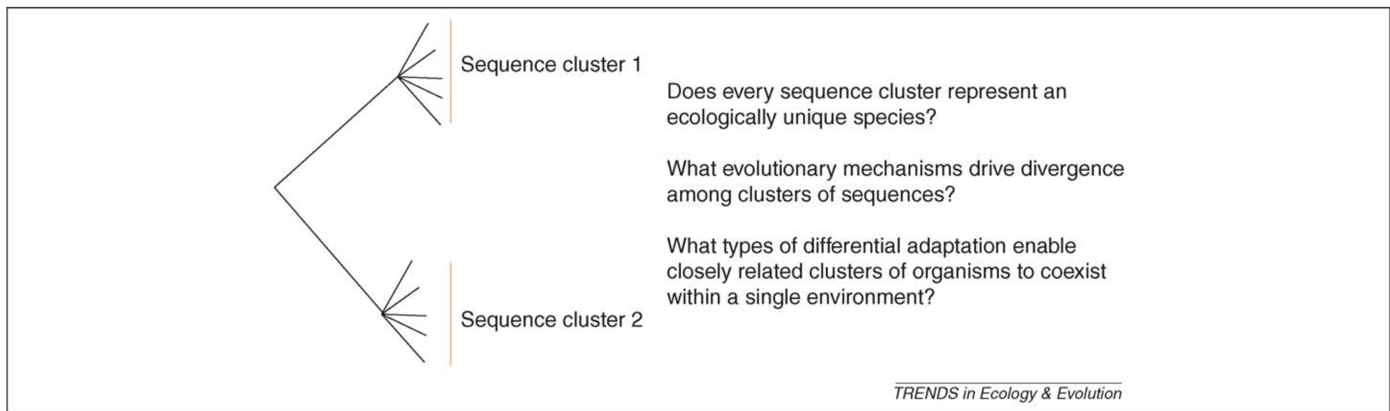

Figure 1. Typical pattern of phylogenetic relationships among sequences for a single gene amplified directly from a microbial community. Listed are fundamental questions in microbial evolutionary ecology that can be resolved through population-genomic analysis of community genome sequence data.

and lineage divergence. In addition, traditional MLST methods are not applicable to the >99% of microorganisms that cannot yet be grown in culture [12].

Community genomics is a new molecular approach that has the potential to make great advances toward answering fundamental questions about the evolutionary ecology of microorganisms because patterns of variation can be assessed at the full genome scale. It involves cloning and sequencing DNA extracted directly from environmental samples, either as large (up to 200 kb) genome fragments [13] or as smaller shotgun libraries that are then used to reconstruct genome fragments or full genome sequences [14,15]. To date, most community-genomic studies have focused on microbial community ecology to identify (i) novel enzymes, antibiotics and signaling molecules through functional screens [16]; (ii) the community structure of viral populations [17]; (iii) the diversity of metabolisms that can be recognized from functional genes [18]; (iv) broad estimates of community-level diversity based on sampling statistics of genomic libraries [14]; and (v) coarse-scale assessments of the types of organisms present within a sample using phylogenetic 'anchors' that are well represented in the public sequence databases [14,18,19]. Here, we describe how population genomics can be applied to community-genomic data to resolve patterns of diversity on a finer scale. Appropriate data for population-genomic analyses (i.e. replicate sampling of the same locus from many individuals) can be obtained through directed approaches that select for similar genome fragments using target sequences or through extensive random sampling of shotgun libraries (Figure I in Box 1). As discussed in Table I, Box 1, the choice of appropriate sequencing method will depend on the questions being asked and the structure of the community under investigation. In Box 2, we describe how to calculate the amount of shotgun sequence necessary for a population-genomic analysis of a model community whose population structure is known. As these data become more available, some of the greatest unrealized contributions will lie in the potential to identify independent lineages and reveal patterns of variation within them simultaneously. Understanding these patterns will help to answer the fundamental questions about mechanisms that structure the microbial diversity described in Figure 1.

Population structure

Population genomics focuses on patterns of variation that occur within genetically cohesive, independently evolving groups of individuals (i.e. within species). For all organisms, the boundaries between independent lineages are defined where populations are isolated by barriers to gene flow [20]. These boundaries can be extrinsic (i.e. geographical or ecological isolation) or intrinsic (i.e. differentiation in resource use or genetic barriers to homologous recombination). We begin our discussion of population genomics in natural microbial populations by describing two models for the evolution of independent lineages within a single environment (see [21] for a discussion of geographical isolation mechanisms). These models differ in their assumptions about the relative importance of recombination in holding lineages together and selection in driving them apart. The two models provide slightly different answers to the first two questions posed in Figure 1.

Niche-specific clonal ecotypes

The periodic-selection model [22–24] describes lineage divergence in populations where recombination is rare and selection is strong. In purely clonal organisms, natural selection for an adaptive allele will increase the frequency of the entire linked genome and purge neutral diversity across all loci as the allele rises to fixation [22,23]. Adaptive alleles are introduced into the population by random mutations or through rare horizontal gene transfer events (HGT) [25]. Once acquired, adaptations make each clonal lineage ('ecotype') ecologically distinct and subject to its own independent periodic selection events, with the latter leading to cohesion within and divergence between ecotypes [24,26]. In a clonal model, each sequence cluster represents a single ecologically relevant unit of diversity (Figure 1a).

The periodic-selection model predicts that sequences from each locus in the genome of primarily clonal microorganisms will show the same relationship among individuals (Figure Ia in Box 3) [10]. Genetic variation within a single cluster of sequences represents neutral divergence since the last periodic selection event for that ecotype. The clonal population structure predicted by this model has been detected by MLST of several different

microbial species, including the pathogens *Staphylococcus aureus* and *Borrelia borgdorferi* [27,28].

Recombining populations

Although Bacteria and Archaea reproduce clonally, frequent recombination and weak selection can produce a recombining population structure [11,29–31]. In the recombining model, recombination counters the 'purifying' processes of clonal selection by unlinking regions of the genome so that selection acts on each locus independently [32]. Because microorganisms reproduce clonally, strong selection can still cause a genome-wide selective sweep even in a recombining microorganism. If selection is relatively weak, unlinking of loci through

Box 1. Two methods for collecting community genome sequence data for population-genomic analysis

Genomic DNA is extracted from all organisms in a single microbial community. The shotgun sequencing approach [15] is shown in Figure Ia. For this method, the genomic DNA is sheered into 3–5 kb pieces that are then cloned into standard *Escherichia coli* vectors to make a small-insert clone library. The clone library is then sequenced at random using primers anchored in either side of the vector. Sequencing from the vector primer extends 700–800 bp from either end of the 3-kb insert, creating paired-sequence ends (short colored bars) that are separated by a ~1500-bp intervening sequence (thin black lines). The shotgun sequence is then assembled into contiguous fragments (contigs) for different species by matching overlapping sequences.

The directed sequencing approach is shown in Figure Ib [70]. Here, the DNA is cut into large fragments using rare-cutting restriction enzymes. These fragments are then ligated into either the BAC (Bacterial Artificial Chromosome) or fosmid vectors designed to maintain large inserts, and are transformed into *E. coli* to make a large-insert clone library. Clones containing different fragments from the community are arrayed and screened using molecular probes to identify certain sequences (usually highly conserved 16S rRNA). Clones containing these target sequences are sequenced selectively, resulting in multiple genome fragments containing the marker for a specific taxon. Variations on these methods include random sequencing of libraries constructed with large fragments or a combination of large and smaller fragments to aid in assembly [19]. The two methods are compared in Table I.

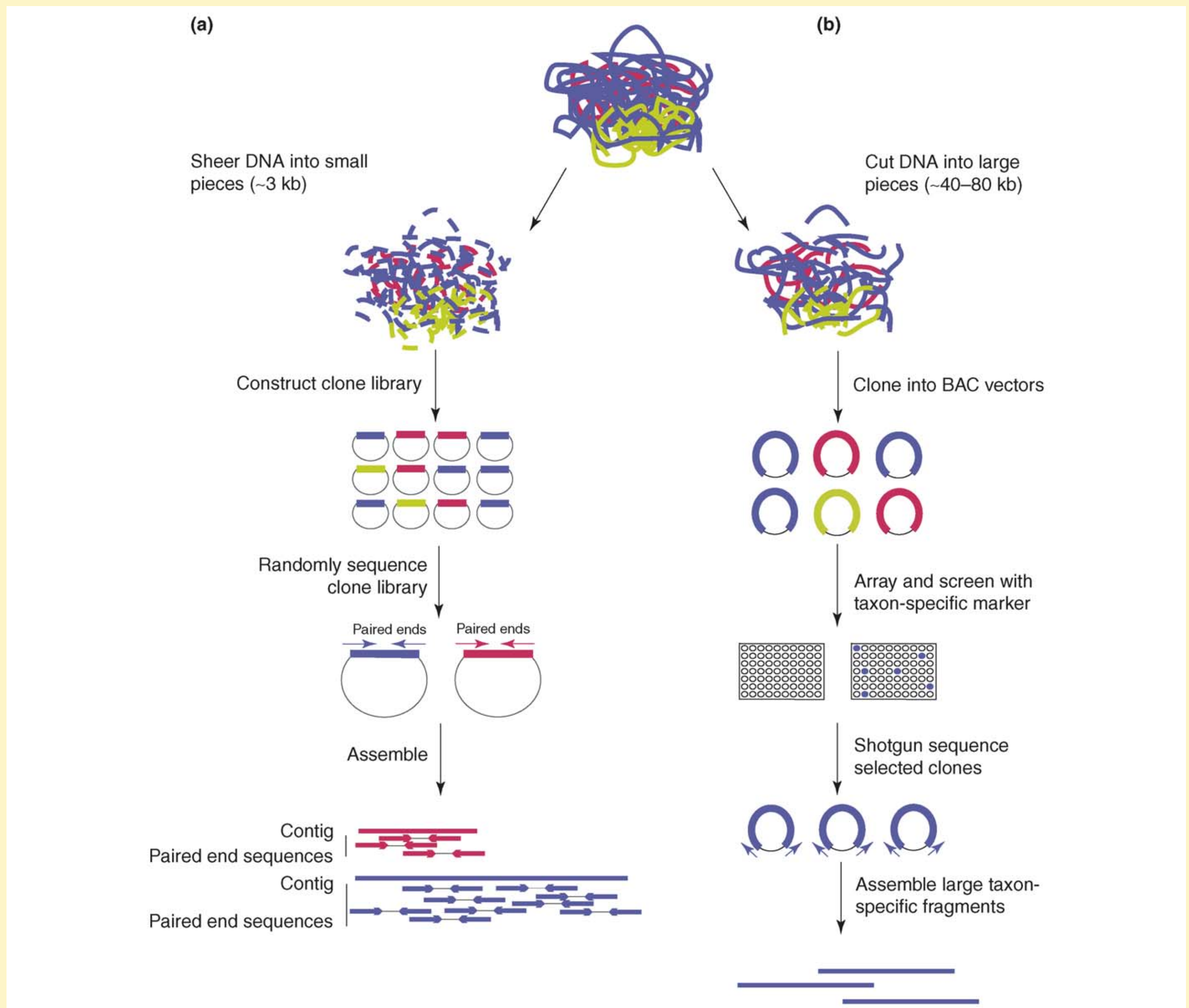

Figure I.

Box 1 (*Continued*)

Table I. Comparison of two methods for population-genomic analysis of community-genomic data

Method	Advantages	Disadvantages
Random shotgun sequencing	• All coexisting genomes within a community can be sampled included, enabling the identification of novel types that might be missed by directed methods • With a sufficient level of sequencing, full genome sequences of the dominant types can be assembled • Best applied to microbial communities with low species richness or evenness (for which one population is highly abundant relative to the others)	• Only the dominant organisms are well represented • Although there is genetic linkage between the end sequences from the same clone (usually spanning 3–5 kbp), linkage between genome regions can only be inferred • Assembly tools often co-assemble independent lineages. Careful manual attention is necessary to dissect very closely related lineages
Directed sequencing	• Sequencing is focused on a specific taxon. Sequence data can be acquired from organisms that are not dominant in the population • Linkage across many loci from a single individual is known • Best applied to microbial communities with high diversity	• The use of specific markers excludes recovery of novel types • Large genome fragments rather than the entire genome are sampled • Sequence data focus on a single group rather than on whole communities

Box 2. Calculating how much shotgun sequencing is needed for population genomics

Figure Ia shows a rank-abundance curve with the relative distribution of different species in a model community. This model community has a structure that characteristic of highly diverse communities, and which has been observed for several microbial communities [38,71], where a few species are abundant but most are rare. Assuming that there is no bias in cloning, the proportion of the clone library with sequence from each species (P_i) can be calculated as $[G_i\,(A_i)]/\Sigma\,[G_i\,(A_i)]$, where G_i is the average genome size of each species i and A_i is the relative abundance of that species in the community.

Figure Ib shows the average depth of coverage for each species from the model community, assuming that every species has same genome size (2 Mb) and a total of 100 Mbp is sequenced from the library. The average number of times each base pair will be sequenced (depth of coverage, C) equals $[P_i\,(T)]/G_i$, where T is the total number of base pairs sequenced for the library. The thick colored bars represent composite sequence, whereas the thin colored bars represent individual paired-end sequence reads. Numbers above thick bars represent estimated average coverage of each base pair (i.e. 10X means that, on average, each base pair in the genome was sequenced ten times). Thin black connections between colored bars represent paired-end sequences from the same clone (as in Figure I of Box 1). Figure Ic shows the proportion of the genome of each species that is sequenced estimated as $1-(e^{-C})$, where e^{-C} is an estimate of the proportion of the genome that will not be sampled given a Poisson distribution for sampling each base-pair position in the genome.

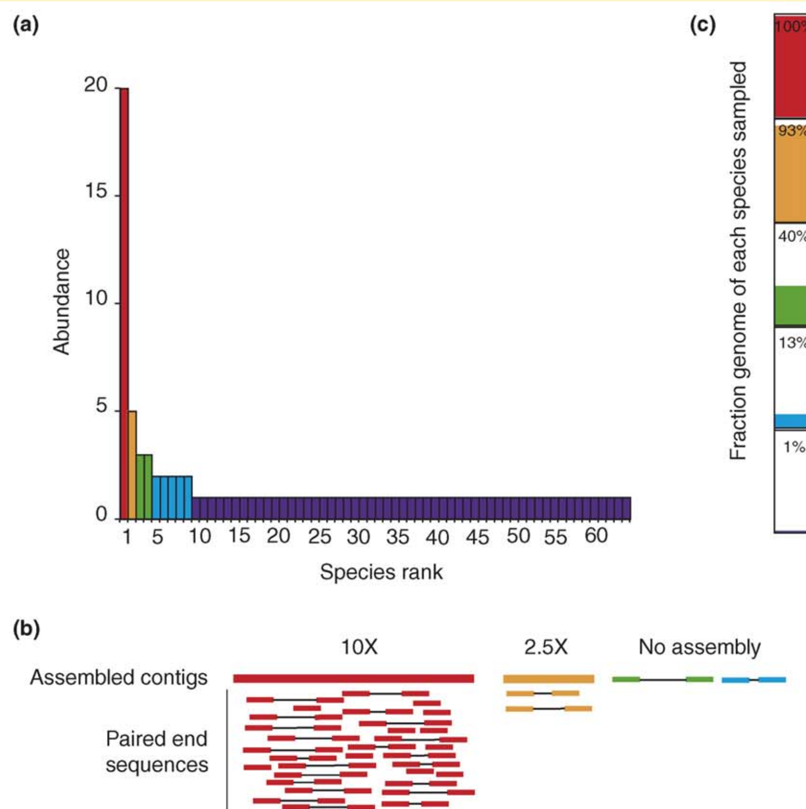

Figure I.

Box 3. Clonal and recombining models for defining boundaries of a population using multilocus genealogies

Figures Ia and Ib show simplified examples of multilocus gene genealogies for clonal and recombining species, respectively. In both Ia and Ib, branches at the end of each tree represent sequences that have diverged through random mutation. Figure Ia shows that neutral diversity is purged (dotted lines) after a periodic selection event as a single sequence type rises to fixation. If a lineage acquires an adaptive mutation that enables it to occupy a new niche, the two lineages (A and B) diverge. Each lineage is subject to its own independent periodic selection event that causes cohesion within an ecotype. Because recombination is rare, events at one locus affect the entire genome; therefore, as shown in the final frame of Figure Ia, the resulting organismal tree matches the locus tree.

In Figure Ib, the same processes of neutral mutation as shown in Figure Ia cause divergence among four lineages A–D. Unlike in the clonal model, recombination redistributes alleles between lineages so that each locus has its own history of selection and neutral divergence, and each history can be different. As shown in red, the allele from lineage B is transferred to D through homologous recombination. Therefore, in the second frame of Figure Ib, B and D are closest relatives for the red locus. The same process occurs at the green locus, this time transferring an allele from B to C. When the different histories for each locus are combined into the organismal tree, independent lineages are not resolved. Instead, the combined tree represents a single recombining lineage. Statements to the right of (a) and (b) answer the first two questions posed earlier in Figure 1 (main text) for the clonal and recombining models.

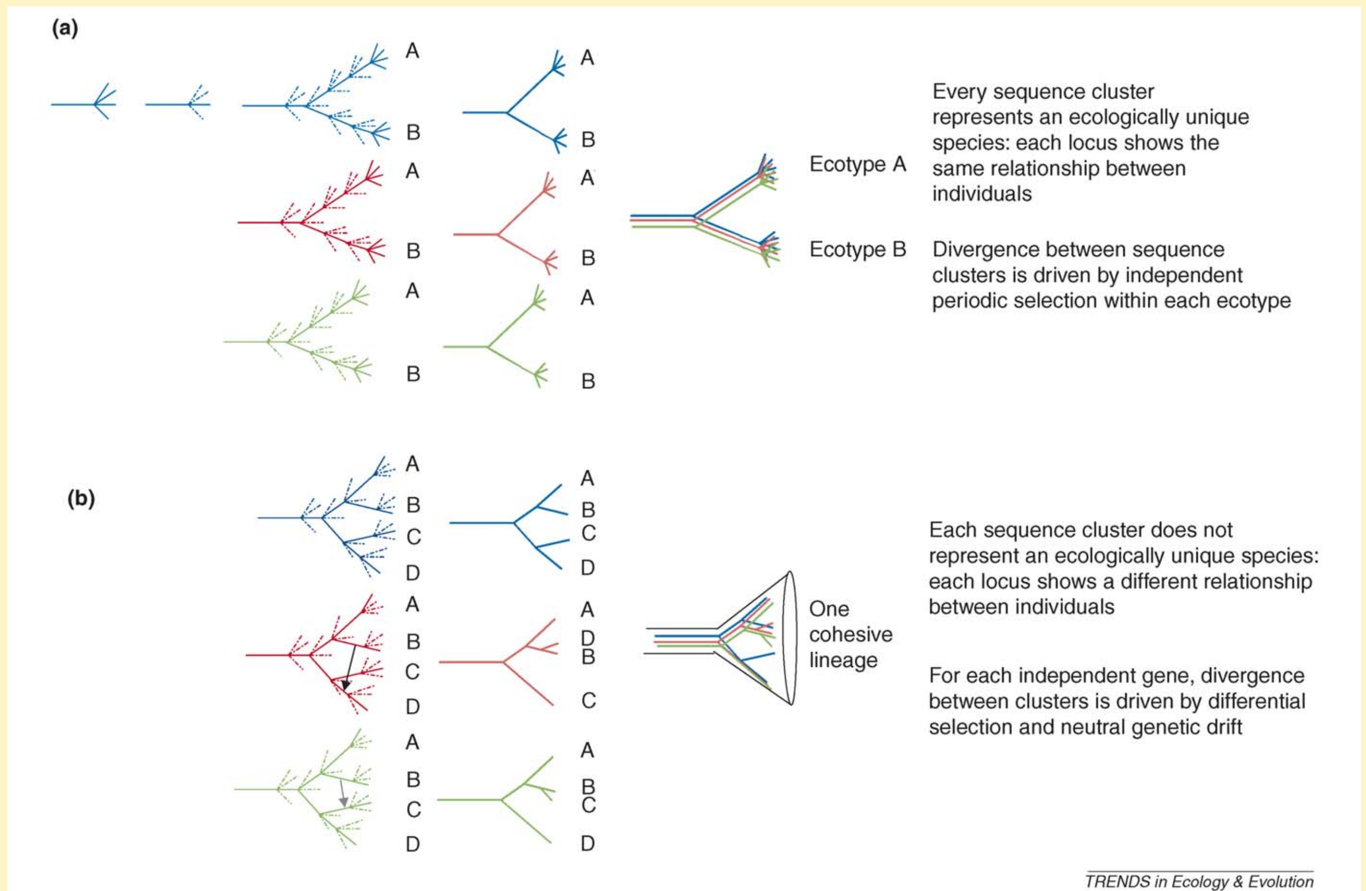

Figure I.

recombination can result in an accumulation of genetic diversity at loci that are not under selection [33]. The distribution of alleles into new combinations also results in greater variance in phenotype and fitness within a cohesive lineage [34,35].

In a recombining population, different loci in the genome define different phylogenetic relationships among individuals (Figure Ib in Box 3) [10]. Unlike the periodic-selection model, clusters of individuals described by consensus phylogenies from several genes represent a single ecological unit. Divergence between sequence clusters at each locus represents either neutral divergence or diversifying selection within a species rather than between ecologically distinct ecotypes as in the clonal model.

Recombining population structures have been observed by MLST in a variety of microorganisms from bacterial gut pathogens [30] to thermoacidophilic Archaea [11]. However, the barriers to gene flow that reinforce divergence among recombining lineages have yet to be defined. It has been suggested that genome-level changes such as rearrangements or insertions of novel genes [36], or an increase in neutral divergence following ecological or physical isolation could cause a breakdown of recombination between lineages [37].

Recognizing population structure in community-genomic data

Because different patterns of variation are predicted by each evolutionary model, examining community genome data for

such patterns will reveal which model best describes the structure of diversity for different organisms within a microbial community. Figure 2 illustrates the ways in which community-genomic data can be used to identify independent lineages from natural microbial populations.

A clonal population structure is revealed when variants with co-segregating single nucleotide polymorphisms (SNPs) are linked across contiguous genome fragments, so that each position describes the same relationship between genotypes (Figure 2a). Because the probability of random mutation at the same nucleotide is low, SNPs that are common between variants can be assumed to have a common ancestry. In shotgun sequence data, genotypes of different clonal ecotypes can be reconstructed through tiling sequence fragments by matching variants and bridging between variants by pairing of sequencing ends. Because each sequence fragment in a shotgun library is derived from a different individual, linkage between matched sequence variants in reconstructed genotypes can only be inferred.

Potential clonal ecotypes have been recognized in environmental genomic data. For example, Strous *et al.* [38] uncovered a single clonal ecotype of the bacteria *Kuenenia stuttgartiensis* with an average of less than one SNP in 15000 bases from a complex bioreactor community. Venter *et al.* [14] assembled the complete genomes of two distinct clonal populations of *Shewanella*-like organisms from environmental data generated from the Sargasso Sea. As mentioned in Table I, Box 1, genome assembly software can mistakenly co-assemble clonal ecotypes into a single contig. It seems likely that this has occurred in the *Prochlorococcus* sequences assembled from the Sargasso Sea, which were significantly more diverse than other organisms recovered from these samples. The divergence of *Prochlorococcus* ecotypes adapted to light conditions at different depths has been well established in similar marine environments [39]. Based on these data, we infer that the relatively high levels of diversity within the *Prochlorococcus* reflect coincident sampling and assembly of several clonal ecotypes from the Sargasso Sea [39].

In recombining populations, contiguous loci within genome fragments can describe different relationships between individual genotypes. Using directed methods with large DNA fragments, the linkage between loci on a single fragment is known. Because linkage is difficult to determine in shotgun sequencing data, recombination can be identified where transitions between two variants are observed within a single sequencing read (Figure 2b). Recombination in a population is also indicated by differences in relationships between variants described by the paired ends of a single cloned fragment. Tyson *et al.* [15] used these methods to identify two discrete recombining archaeal lineages coexisting in a single acid mine drainage biofilm community.

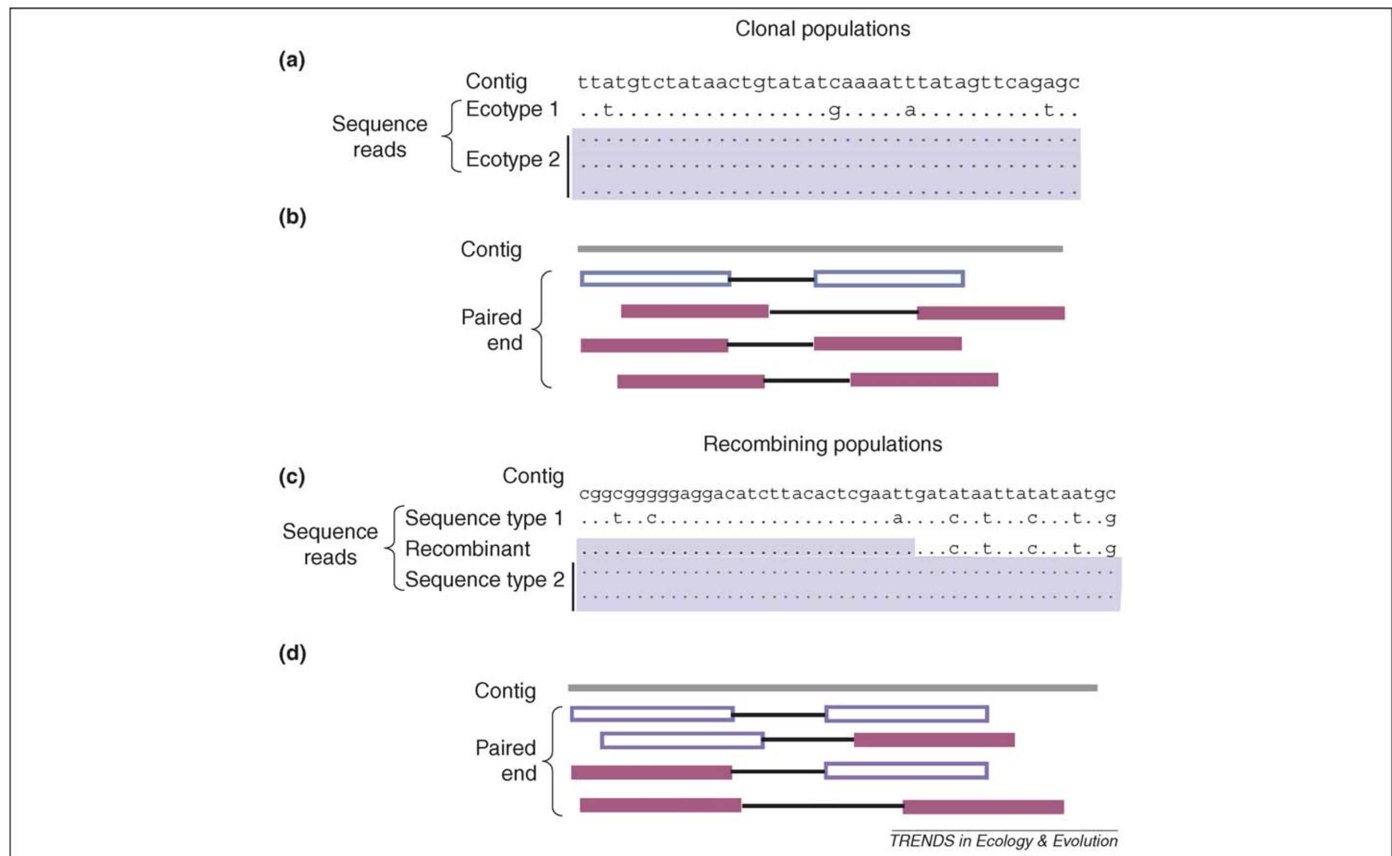

Figure 2. Signatures of clonal and recombining population structures in population-genomic data. **(a)** and **(c)** show assembled sequencing reads from a random shotgun library, whereas **(b)** and **(d)** show paired-end sequence reads from the same clone with an assembled contig. For a clonal population, two ecotypes are shown distinguished by (a) a pattern of co-segregating SNPs and (b) consistent co-segregating patterns between paired sequences from the same cloned fragment. (c) and (d) show the signature of a recombining population, where in (c) transitions between sequence patterns are observed within a single sequence read (blue to red in the second read labeled recombinant) or in (d) paired-end sequences are found in different combinations.

We have discussed population-genomic approaches to the first two questions posed in Figure 1 by identifying the structure of the genetic diversity within a microbial community. Now, we approach the third question, namely how population-genomic data can be used to identify loci that are under natural selection in microbial communities.

Genomic sources of niche differentiation

The long-standing 'paradox of plankton' asks how a diversity of lineages is maintained in an apparently limited number of ecological niches [40], examples of which abound in the microbial world [41–44]. For most microbial communities, it seems probable that one solution is that the number of ecological niches cannot be resolved accurately at the microbial scale. High diversity is maintained by differential selection associated with distinct 'microniches' that cannot be resolved with tools available currently to microbial ecologists. Here, we describe how the process of adapting to a niche should leave a signature in the pattern of genomic variation within and between independent lineages. Once independent lineages have been defined, the function of these genome changes provides hints of the niche-specific adaptations that promote ecological divergence and enable lineages to coexist.

Recognizing adaptive loci in recombining populations

By sampling many individual examples of homologous loci, population genomics can identify variation across large regions of the genome. This provides the unique opportunity to identify loci under selection without relying on *a priori* assumptions about the types of genes that are selected for. Many methods have been developed for identifying selection in sexual eukaryotes when loci are unlinked [45]. These methods are useful only in recombining microbial populations because the extensive genome-wide hitchhiking associated with periodic selection in clonal populations obscures the effect of selection on different loci. Distinguishing selection at a given locus from other population dynamics with similar sequence signatures requires the assessment of variation across many loci that is made possible through population genomics.

Methods for detecting selection based on patterns of sequence variation rely on differences in the frequency distribution and types of substitutions from those expected under neutral evolution [46–48]. Loci under positive selection are identified by comparing the relative proportion of nonsynonymous and synonymous substitutions among alleles within a species or between two closely related species that coexist in a community. This type of analysis has been used to identify the amino-acid residues responsible for the tuning of light-harvesting proteins by marine bacteria found at different ocean depths [49]. In community-genomic data sets with sufficient coverage, estimation of the frequency distribution of different types of nucleotide substitutions is possible. In these data sets, loci under positive selection are identifiable even if they have not completely risen to fixation. Finally, genome regions with relatively low levels of neutral diversity can also be used to identify linked loci under selection [50]. Again, this method is only viable for samples with sufficient coverage of a single lineage.

Adaptive changes on the genome scale

Comparative genomics among closely related, isolate genomes suggests that dramatic changes are the primary driving force of adaptive evolution in microorganisms rather than the gradual acquisition of adaptive mutations. These changes include the insertion of a block of adaptive genes known as genomic islands [51], the acquisition of novel genes from divergent organisms through HGT [52], and the insertion or deletion of autonomous mobile genomic elements such as viruses or transposons [53]. Careful reconstruction of genotypes from community-genomic data can be used to identify differences either in the position of mobile elements or in the gene content of coexisting clonal or recombining lineages [54]. Such analyses will reveal the importance of these mechanisms to adaptation and divergence in microbial species.

Community genome sequencing has already shed light on the process of HGT in microbial communities. Rampant gene transfer would result in every individual in the community having a different gene complement and preclude genome assembly from shotgun sequence data. Assembly of nearly complete genomes from shotgun sequence data shows that, at least on recent evolutionary time-scales, HGT is not rampant in natural microbial communities [17]. Rare horizontal transfer of novel genes might confer adaptive advantage and drive diversification, but the genomes of individual species appear to be largely cohesive [55].

Insertions, deletions or genome rearrangements resulting from the movement of mobile elements can either inactivate or up-regulate nearby genes [56,57]. Although such elements are often considered 'genetic parasites', several experimental evolution studies and population surveys have shown their positioning to be adaptive [53,58–60]. Differences in the position or content of genetic elements have also been noted in several community-genomic data sets [15,61,62]. Because genotypes in which the positions of the mobile elements are adaptive will increase within populations, a conserved location of these elements within independent lineages indicates selective fixation. Differential positioning among genomes within a population indicates either that the elements move within a genome at a higher frequency than do selection events or that their location is essentially neutral. A conserved genomic location within a lineage, but differential locations between lineages, might indicate adaptive differences between closely related species. Hypotheses about the ecological function of conserved mobile elements can be developed by examining the function of neighboring genes on assembled genome fragments.

It has been suggested that the acquisition of novel laterally derived genes can confer adaptive function, enabling the colonization of new ecological niches and divergence between lineages [63,64]. The assembly of genomes from community-genomic data has identified differences in gene content among genomes of closely related microorganisms within a community [14,43,54, 62,65,66]. As with mobile elements, the adaptive significance of novel genes can be ascertained from their conservation within populations or differential distribution between them. It might also be possible to determine the adaptive

significance of novel genes to the different lineages by using homology searches of the public databases to infer their function.

Testing ecological hypotheses derived from population genomics

Although the analyses described here provide clues about the genes that are under selection in natural populations, proof that these genomic changes lead to divergence and coexistence requires more work. For example, to test for the importance of adaptations to niche separation, common garden competition experiments are needed to prove coexistence [67]. Additionally, the adaptive significance of newly acquired genes can be proved through experimental genetic knockouts and physiological assays of culturable organisms. Approaches that do not involve culturing include comparative environmental genomics among replicates sampled through space and time to link hypotheses about the adaptive significance of genome changes to their native environments. Furthermore, changes in gene expression can be tested using expression microarrays or quantitative proteomics on natural samples [68].

Conclusions

Molecular tools involving single genes enabled microbiologists to begin to describe the extent and distribution of microbial diversity across the planet. These surveys laid the groundwork for the analysis of the mechanisms of evolutionary ecology through which this biodiversity evolved. The crucial next step as environmental microbiology matures will be to put the data from single genes into a genomic context. We have discussed here how population genomics enables the simultaneous assessment of the structure of genetic diversity into independent evolutionary lineages and the identification of genome changes within these lineages that can maintain diversity within microbial communities.

We have focused primarily on population genomics because we feel that identifying lineages is the essential first step toward the accurate reconstruction of the genetic architecture of microbial communities. However, perhaps the greatest promise of this emerging field is its ability to put these lineages back together into communities; that is, to describe patterns of community and species-level diversity simultaneously. Whereas the fundamental processes that structure diversity at the community and population levels might be similar (e.g. environmental heterogeneity, rates of dispersal, cooperation, competition, or genetic exchange, among others) [69], understanding the relationships between community- and population-level diversity represents a new frontier for combining ecological and evolutionary paradigms. Applying these comparisons to microbial communities begins with identifying the boundaries of ecologically relevant independent lineages through population genomics. Once such boundaries are better understood, and their ecological relevance identified, community-genomic analysis should enable the simultaneous observation of community- and population-level dynamics.

Whereas evolutionary biologists have been studying the natural history of macroorganisms for centuries, microbiologists are just beginning to explore the evolutionary ecology of the microbial world rigorously. However, questions about the mechanisms of speciation, adaptation, niche partitioning and community structure are not unique to this fledgling field. We have described the potential of population-genomic analysis of microbial community-genomic data to answer some of these fundamental questions in natural microbial communities. Because microorganisms have small genomes, population-genomic analysis is possible on a larger scale (i.e. reconstructing more genomes from more individuals and sampling more sites) than is possible for organisms with larger genomes. Therefore, as this field develops, we anticipate that microbial population genomics will contribute to understanding the principles of ecology and evolution that apply across the tree of life.

Acknowledgements

We thank B. Bohannan, C. Horner-Devine, E. Turner, S. Wald and three anonymous reviewers for helpful discussion and comments on this article. Funding was provided by a grant from the NSF Biocomplexity Program, EAR0221768.

References

1 Olsen, G.J. *et al.* (1986) Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* 40, 337–365
2 Schmidt, T.M. *et al.* (1991) Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J. Bacteriol.* 173, 4371–4378
3 Acinas, S.G. *et al.* (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* 430, 551–554
4 Maiden, M.C.J. *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145
5 Hanage, W.P. *et al.* (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol.* 3, 6
6 Stackebrandt, E. *et al.* (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047
7 Whitaker, R.J. *et al.* (2003) Geographic barriers isolated endemic populations of hyperthermophilic Archaea. *Science* 301, 976–978
8 Palys, T. *et al.* (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
9 Cohan, F.M. (2001) Bacterial species and speciation. *Syst. Biol.* 50, 513–524
10 Dykhuizen, D.E. and Green, L. (1991) Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* 173, 7257–7268
11 Whitaker, R.J. *et al.* (2005) Recombination shapes the natural population structure of the hyperthermophilic archaeon *Sulfolobus 'islandicus'*. *Mol. Biol. Evol.* 22, 2354–2361
12 Amann, R.I. *et al.* (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169
13 Beja, O. *et al.* (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ. Microbiol.* 2, 516–529
14 Venter, J.C. *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66–74
15 Tyson, G.W. *et al.* (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37–43
16 Riesenfeld, C.S. *et al.* (2004) Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38, 525–552
17 Breitbart, M. *et al.* (2002) Genomic analysis of uncultured marine viral communities. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14250–14255
18 Tringe, S.G. *et al.* (2005) Comparative metagenomics of microbial communities. *Science* 308, 554–557
19 DeLong, E.F. *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 311, 496–503

20 Coyne, J.A. and Orr, H.A. (2004) *Speciation,* Sinauer Associates
21 Green, J. and Bohannan, B.J. (2006) Spatial scaling of microbial biodiversity. *Trend. Ecol. Evol.* 21, 501–507
22 Atwood, K.C. *et al.* (1951) Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 37, 146–155
23 Levin, B.R. (1981) Periodic selection, infectious gene exchange and the genetic structure of *Escherichia coli* populations. *Genetics* 99, 1–23
24 Cohan, F.M. (2002) What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487
25 Cohan, F.M. (2005) Periodic selection and ecological diversity in bacteria. In *Selective Sweep* (Nurminsky, D., ed.), pp. 78–93, Kluwer Academic
26 Rundle, H.D. and Nosil, P. (2005) Ecological speciation. *Ecol. Lett.* 8, 336–352
27 Feil, E.J. *et al.* (2003) How clonal is *Staphylococcus aureus*? *J. Bacteriol.* 185, 3307–3316
28 Smith, J.M. *et al.* (2000) Population structure and evolutionary dynamics of pathogenic bacteria. *BioEssays* 22, 1115–1122
29 Feil, E.J. *et al.* (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc. Natl. Acad. Sci. U. S. A.* 98, 182–187
30 Falush, D. *et al.* (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. U. S. A.* 98, 15056–15061
31 Holmes, E.C. *et al.* (1999) The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.* 16, 741–749
32 Muller, H.J. (1964) The relation of recombination to mutational advance. *Mutat. Res.* 1, 2–9
33 Smith, J.M. *et al.* (1993) How clonal are bacteria? *Proc. Natl. Acad. Sci. U. S. A.* 90, 4384–4388
34 Goddard, M.R. *et al.* (2005) Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* 434, 636–640
35 Burt, A. (2000) Perspective. Sex, recombination, and the efficacy of selection–was Weismann right? *Evolution* 54, 337–351
36 Vetsigian, K. and Goldenfeld, N. (2005) Global divergence of microbial genome sequences mediated by propagating fronts. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7332–7337
37 Majewski, J. and Cohan, F.M. (1998) The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* 148, 13–18
38 Strous, M. *et al.* (2006) Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* 440, 790–794
39 Rocap, G. *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424, 1042–1047
40 Hutchinson, G.E. (1961) The paradox of plankton. *Am. Nat.* 95, 137–145
41 Thompson, J.R. *et al.* (2005) Genotypic diversity within a natural coastal bacterioplankton population. *Science* 307, 1311–1313
42 Dykhuizen, D.E. (1998) Santa Rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek Int. J. Gen. Mol. Microbiol.* 73, 25–33
43 Schleper, C. *et al.* (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *J. Bacteriol.* 180, 5003–5009
44 Quaiser, A. *et al.* (2003) Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* 50, 563–575
45 Bamshad, M. and Wooding, S.P. (2003) Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4, 99–111
46 Bustamante, C.D. *et al.* (2001) Directional selection and the site-frequency spectrum. *Genetics* 159, 1779–1788
47 Sawyer, S.A. and Hartl, D.L. (1992) Population genetics of polymorphism and divergence. *Genetics* 132, 1161–1176
48 McDonald, H. and Kreitman, M. (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351, 652–654
49 Bielawski, J.P. *et al.* (2004) Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14824–14829
50 Guttman, D.S. and Dykhuizen, D.E. (1994) Detecting selective sweeps in naturally occurring *Escherichia coli*. *Genetics* 138, 993–1003
51 Hsiao, W.W. *et al.* (2005) Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genet.* 1, e62
52 Daubin, V. *et al.* (2003) Phylogenetics and the cohesion of bacterial genomes. *Science* 301, 829–832
53 Schneider, D. and Lenski, R.E. (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. *Res. Microbiol.* 155, 319–327
54 Allen, E.E. and Banfield, J.F. (2005) Community genomics in microbial ecology and evolution. *Nat. Rev. Microbiol.* 3, 489–498
55 Lerat, E. *et al.* (2005) Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* 3, e130
56 Saedler, H. *et al.* (1974) IS2, a genetic element for turn-off and turn-on of gene activity in *E. coli*. *Mol. Gen. Genet.* 132, 265–289
57 Reynolds, A.E. *et al.* (1981) Insertion of DNA activates the cryptic bgl operon in *E. coli* k12. *Nature* 293, 625–629
58 Blot, M. (1994) Transposable elements and adaptation of host bacteria. *Genetica* 93, 5–12
59 Cooper, V.S. *et al.* (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *J. Bacteriol.* 183, 2834–2841
60 Aminetzach, Y.T. *et al.* (2005) Pesticide resistance via transposition-mediated adaptive gene truncation in *Drosophila*. *Science* 309, 764–767
61 Schirmer, A. *et al.* (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl. Environ. Microbiol.* 71, 4840–4849
62 Hallam, S.J. *et al.* (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* 305, 1457–1462
63 Lawrence, J.G. (2001) Catalysing bacterial speciation: correlating lateral transfer with genetic headroom. *Syst. Biol.* 50, 479–496
64 Gogarten, J.P. *et al.* (2002) Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* 19, 2226–2238
65 de la Torre, J.R. *et al.* (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12830–12835
66 Sabehi, G. *et al.* (2004) Different sar86 subgroups harbour divergent proteorhodopsins novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ. Microbiol.* 6, 903–910
67 Rainey, P.B. and Rainey, K. (2003) Evolution of cooperation and conflict in experimental bacterial populations. *Nature* 425, 72–74
68 Ram, R.J. *et al.* (2005) Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920
69 Vellend, M. (2005) Species diversity and genetic diversity: parallel processes and correlated patterns. *Am. Nat.* 166, 199–215
70 Beja, O. *et al.* (2002) Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Appl. Environ. Microbiol.* 68, 335–345
71 Hughes, J.B. *et al.* (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67, 4399–4406

Reproduction of material from Elsevier articles

Interested in reproducing part or all of an article published by Elsevier, or one of our article figures? If so, please contact our *Global Rights Department* with details of how and where the requested material will be used. To submit a permission request online, please visit:

www.elsevier.com/locate/permissions