environmental microbiology

Environmental Microbiology (2012) **14**(2), 347–355 doi:10.1111/j.1462-2920.2011.02668.x

Minireview

Bacterial species may exist, metagenomics reveal

Alejandro Caro-Quintero[1] and Konstantinos T. Konstantinidis[1,2]*

Schools of [1]Biology and [2]Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Summary

Whether or not bacterial species exist remains an unresolved issue of paramount theoretical as well as practical consequences. Here we review and synthesize the findings emerging from metagenomic surveys of natural microbial populations and argue that microbial communities are predominantly organized in genetically and ecologically discernible populations, which possess the attributes expected for species. These sequence-discrete populations represent a major foundation for beginning high-resolution investigations on how populations are organized, interact, and evolve within communities. We also attempt to reconcile these findings with those of previous studies that reported indiscrete species and a genetic continuum within bacterial taxa and discuss the implications for the current bacterial species definition.

Introduction

A bacterial species is ideally considered to be 'a collection of strains that are genomically coherent (discrete) compared to strains of different species based on several independent features and characterized by at least one diagnostic phenotypic trait' (Wayne *et al*., 1987; Stackebrandt *et al*., 2002). Whether or not bacteria actually form such discrete units (species) has been put into question in light of recent sequence data that has frequently revealed a genetic continuum and indiscrete ('fuzzy') clades within several bacterial genera (Hanage *et al*., 2005; Konstantinidis *et al*., 2006; Rusch *et al*., 2007). Horizontal gene transfer is also so pervasive in the microbial world that it can potentially reduce genetic isolation between otherwise discrete bacterial populations (Lawrence, 2002; Sheppard *et al*., 2008). These findings have led several scientists to conclude that there is no intrinsic reason why the processes driving diversification and adaptation of bacteria must produce sufficiently coherent groups of individuals (species); consequently, no universally accepted species concept exists for bacteria. Despite the uncertainty about whether or not bacteria form discrete units, bacterial strains are conventionally classified into named species using an assortment of molecular and phenotypic methods and standards (species here denotes a manmade taxonomic category, contrasting with the preceding text that refers to species as a natural entity). This system and the corresponding species definition have been pragmatic, operational and universally applicable (Rossello-Mora and Amann, 2001; Stackebrandt *et al*., 2002). However, the methods and standards employed frequently result in too much genomic and phenotypic variation within the named species. Indeed, strains of the same named bacterial species frequently differ in up to 30–35% of their total genes in the genome, resulting in organisms with contrasting phenotypes and/or ecological strategies (e.g. highly invasive pathogens vs. commensal strains in the case of *Escherichia coli*) to be grouped together as the same species (Konstantinidis and Tiedje, 2005; Luo *et al*., 2011a). Efforts toward a more refined definition of a bacterial species are clearly needed.

A major limitation common in all previous studies that reported indiscrete species is that the strains analysed were frequently isolated from different populations and habitats (Hanage *et al*., 2005; Rusch *et al*., 2007). The extent of ecological success and genomic adaptation to the place of isolation remained typically inaccessible for almost every strain evaluated; therefore, the strains were likely too ecologically and phenotypically heterogeneous, severely confounding downstream population genetic analyses and comparisons. It is also important to point out that the great majority of microorganisms resists cultivation under laboratory conditions, i.e. 'the uncultivated majority' (Amann *et al*., 1995); hence, the majority of the previous studies, which were based on culture-based approaches, were limited in providing quantitative views

Received 5 September, 2011; revised 6 November, 2011; accepted 8 November, 2011. *For correspondence. E-mail kostas@ce.gatech.edu; Tel. (+1) 404 385 3628. Fax (+1) 404 894 8266.

of the naturally occurring diversity. To better understand the drivers of genetic variation in microbial populations, within-community genomic variation should be analysed, focusing ideally on abundant populations that are less likely to represent transient and/or allochthonous members of the community.

Microbial communities consist of sequence-discrete populations

Culture-independent approaches such as metagenomics are ideal to provide new insights into the previous issues because they can sample the entire community, bypassing the need to isolate and culture individual community members (Handelsman *et al*., 2007). The metagenomic studies that are most relevant for the thesis of this review are those that recovered a large number of whole-genome shotgun (WGS) sequences from community DNA (Tyson *et al*., 2004; Venter *et al*., 2004; Rusch *et al*., 2007; Konstantinidis *et al*., 2009a). In these studies, the WGS sequences typically originate from different cells in the sample and do not show any strong biases with respect to the part of the genome or the genotype (strain) they represent, thus providing random (unbiased) representations of the total naturally occurring diversity. Most of such studies have been conducted in marine environments; accordingly, the main focus of the remaining text is the diversity within marine microbial communities, with reference to terrestrial communities when appropriate.

In all large-scale metagenomic studies performed in the ocean environment to date, the microbial communities sampled were found to predominantly consist of discrete, sequence-defined populations. The populations became evident when all WGS reads in a sample were queried against a reference genome sequence from the same sample: the WGS reads representing members of the same population as the reference genome were highly related among themselves, i.e. they typically showed higher than 90% nucleotide identity, and divergent relative to WGS reads representing any co-occurring relatives in the sample, showing $< 80\%$ nucleotide identity. A few WGS reads that showed intermediate nucleotide identities (e.g. 80–90%) to the reference sequence were shown to represent sequencing errors and artifacts, fast-evolving genes or populations that have differentiated genetically and ecologically compared to the reference population since under the same conditions (co-occurring in the same sample) the former populations were less abundant based on the number of WGS reads representing each population, which serves as a proxy of *in situ* abundance. Thus, a genetic discontinuity between a population and the rest of the community was observed in the 80–90% nucleotide identity range [Fig. 1 and in Konstantinidis and DeLong (2008)]. Furthermore, the number of WGS reads originating from individual genotypes of a population were similar, indicating similar *in situ* abundances among the genotypes (Konstantinidis and DeLong, 2008). The genetic distinctiveness and the matching sympatric abundances imply that the genotypes of a population share similar ecophysiological properties, and therefore their population may represent the important unit of microbial diversity. In other words, a population is defined in this case as the natural entity present in a community/sample that comprises genotypes, which are clearly distinguishable from their closest co-occurring relatives (if any) based on their high genetic relatedness and comparable relative abundance *in situ*. Such sequence-discrete populations were observed for phylogenetically diverse microbial groups, including, but not limited to, *Crenarchaea*, *Cyanobacteria* and *Proteobacteria*, independent of the lifestyle of the organism considered (e.g. autotrophic vs. heterotrophic); albeit the extent of intra-population divergence ranged from ~1% to ~10% nucleotide sequence dissimilarity based on single WGS reads, depending on the population considered (Konstantinidis and DeLong, 2008).

Although these findings were initially (Rusch *et al*., 2007) based on short, error-prone WGS sequences, similar results were obtained with longer fosmids clone sequences (Konstantinidis and DeLong, 2008) and, more recently, with whole genomes (Caro-Quintero *et al*., 2011). The latter studies also revealed that the intra-population genetic diversity was, in reality, smaller than that observed based on WGS reads, i.e. it ranged from $< 1\%$ to ~5% genome-aggregate average nucleotide sequence dissimilarity; the difference is attributable to the increased sequencing error rate of individual WGS reads and the larger distribution around the genome average identity for short vs. long sequences. These studies also revealed that the gene content differences among genotypes of a population were small, typically less than 5% of the total genes in the genome (Konstantinidis and DeLong, 2008), which contrasts with up to 30% of the genes in the genome to differ in strains of many commonly defined bacterial species (Konstantinidis and Tiedje, 2005). These findings reveal that when the strains compared are ecologically homogeneous, they show small genomic (and presumably phenotypic) differentiation. It is important to point out, however, that the sequence-discrete populations are not clonal but contain significant intra-population sequence and gene content diversity, albeit smaller diversity relative to what observed within several named bacterial species. This intra-population diversity is presumably important for evolution and adaptation, consistent with the core-genome hypothesis (Medigue *et al*., 1991; Lan and Reeves, 2000). Furthermore, the 1% to 5% genome-aggregate nucleotide

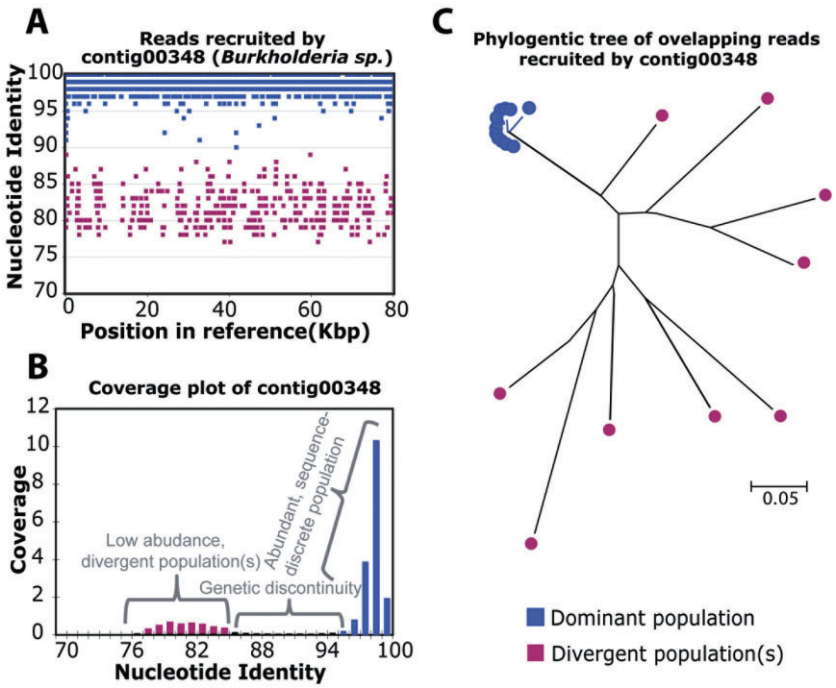

Fig. 1. Sequence-discrete populations.
A. Fragment recruitment plot of the Lake Lanier (Atlanta, GA) metagenome (Oh *et al*., 2011), performed essentially as described previously (Rusch *et al*., 2007), using as reference a large contig (100 kb) of a *Burkholderia* sp. isolate (heterotrophic, *β-Proteobacterium*). All WGS reads of the metagenome were searched against the reference sequence and the results shown are based on all reads that showed at least 70% nucleotide identity and fully aligned to the reference.
B. Coverage plot of the same data as in A, performed as described previously (Konstantinidis and DeLong, 2008). The coverage (*y*-axis) is calculated by summing the length of all reads matching the reference sequence with a given nucleotide identity (*x*-axis) divided by the total length of the contig. Thus, coverage is normalized to the length of the reference and representative of the relative *in situ* abundance of the corresponding population. Note that in the case of *Burkholderia* a lower abundance (compared to the reference population) closely related (showing 75–85% nucleotide identity to the reference) population was detectable.
C. Neighbour-joining phylogenetic tree of all fully overlapping reads of the metagenome that mapped on the single-copy transcription termination factor Rho encoded on the contig. Note that a sequence-discrete populations is also evident by the phylogenetic approach. Similar sequence-discrete populations were observed for other microbial groups and habitats. If sequence-discrete populations did not typify the microbial communities sampled then WGS reads of intermediate nucleotide sequence identity (i.e. a genetic continuum) and/or coverage plots that did not look like normal distributions around the genome average nucleotide sequence identity (e.g. uneven coverage plots) would have been observed [for some examples see Konstantinidis and DeLong (2008), Luo *et al*. (2011a)].

sequence divergence observed among genotypes of the same population corresponds to thousands of generations from the last common ancestor (Lawrence and Ochman, 1998). Thus, the sequence-discrete populations discussed above likely represent long-lived entities, shaped presumably by long-term ecological and genetic selection pressures. Similar patterns of diversity to those described above for oceanic communities have been observed in human gut-associated (Qin *et al*., 2010), iron-reducing biofilm (Tyson *et al*., 2004), phosphorus removal bioreactor (Garcia Martin *et al*., 2006), hot springs (Bhaya *et al*., 2007) and freshwater microbial communities (Oh *et al*., 2011), revealing that sequence-discrete populations typify natural microbial communities.

It is possible that the identified sequence-discrete populations encompass further ecologically distinct subpopulations. Indeed, if such subpopulations emerged only recently from within a population they would not have been differentiated enough genetically to be detectable by the metagenomic approaches described above [e.g. to be represented by uneven coverage plots (Konstantinidis and DeLong, 2008; Luo *et al*., 2011b)]. To detect such subpopulations, time series data combined with whole-genome sequences and population genetic analyses will be required. Application of the latter techniques in a marine *Vibrio splendidus* population identified recently emerged, ecologically distinct subpopulations (Hunt *et al*., 2008). Even though the existence of ecologically distinct subpopulations within the identified sequence-discrete populations cannot be ruled out, the identification of sequence-discrete populations for almost every known abundant microbial group and in diverse habitats indicate that subpopulations are quantitatively infrequent. It should be also mentioned that rare members of the communities were not assessed in the previous studies due to the low sequence coverage achieved for these organisms. It is

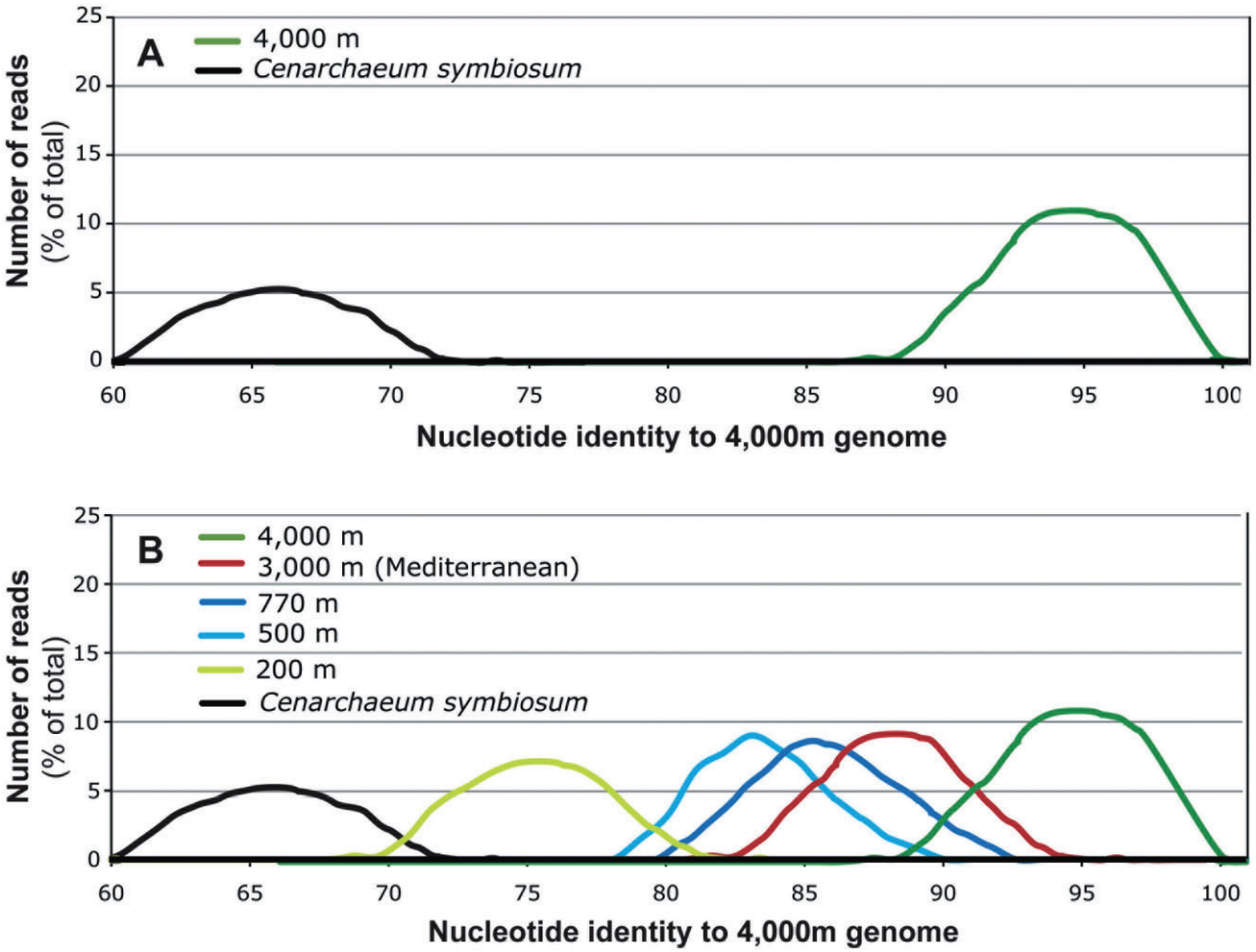

Fig. 2. Genetic relatedness among oceanic Group I *Crenarchaea* populations. The graph shows, similar to Fig. 1B, the coverage of the crenarchaeal genome from 4000 m depth reported previously (Konstantinidis and DeLong, 2008) by selected crenarchaeal populations from the same site but different depths (Station Aloha, Pacific Ocean) or the Mediterranean Sea (figure key). *Cenarchaeum symbiosum* represents the closest relative of the oceanic crenarchaeal populations with a complete genome available. Note that the populations from different depths cover the sequence space between *C. symbiosum* and the 4000 m population and that the Mediterranean population from 3000 m is more closely related to the 4000 m population compared to populations from shallower waters from the same site as the 4000 m population. Populations from 500 m (above the dissolved oxygen minimum layer) and 770 m (in the middle of the dissolved oxygen minimum layer) show about 85% ANI, i.e. they are roughly as divergent from each other as each population is from the 4000 m population, while the 500 m population is more related (~ 90% ANI) to the population from 200 m (data not shown), which is consistent with the vertical zonation of the water column. The graph is drawn based on the data in Konstantinidis (2011) and Konstantinidis and DeLong (2008).

likely, however, that such low-abundance organisms form homogeneous sequence-discrete populations when favourable environmental conditions for their growth prevail (Campbell *et al*., 2011).

Biogeography of sequence-discrete populations and cases of indiscrete populations

Studying the distribution of sequence-discrete populations across geographically distant sites (biogeography) revealed further insights into the mechanisms maintaining bacterial diversity. When the sites were characterized by similar physicochemical conditions, were interconnected (such as the oceans, which are connected through the global water circulation), and the organisms occupied the same ecological niche(s) within each site, the populations showed no biogeography. For instance, photosynthetic *Prochlorococcus* (*Cyanobacteria*) populations were indistinguishable both in terms of nucleotide sequence relatedness as well as gene content between the surface waters (20–25 m depth) of the Atlantic Ocean (Station Bermuda) and the Pacific Ocean (Station Aloha) (Luo and Konstantinidis, 2011). Similarly, populations of Group I *Crenarchaea*, an abundant ammonia-oxidizing marine archaeal group, were genetically very similar between the Mediterranean Sea and the Pacific Ocean at similar depths (3000 m and 4000 m respectively) [Fig. 2 and in Konstantinidis (2011)]. In contrast, populations of *Prochlo-*

rococcus (Coleman and Chisholm, 2010) and *Crenarchaea* (Konstantinidis, 2011) were distinct at different depths within the water column, resulting from genomic adaptations to the unique physicochemical conditions that characterize each depth (e.g. different hydrostatic pressure, light and resource availability). In fact, when organisms from different depths were analysed together, a genetic continuum, which included organisms showing the whole range of nucleotide identity relatedness from 70% to 100%, as opposed to sequence-discrete clusters, was observed [Fig. 2 and in Konstantinidis (2011)]. These results reveal that the sequence-discrete populations are not limited to a specific site or represent an artefact of a few samples analysed; rather, they are distributed worldwide in interconnected habitats that are characterized by similar environmental conditions, which is consistent with the notion that they represent the important units of diversity. The results also highlight that when organisms, which have adapted to living within habitats that exercise different selection pressures such as different depths of the water column, are compared, indiscrete populations (species) may emerge, confirming the limitations of the culture-based studies mentioned above and the importance of population data for robust conclusions.

What are the forces of population cohesion?

A major question remaining is how the genotypes of a population cohere together, i.e. what the forces of population cohesion are. Several theories based on recombination frequency (Feil *et al*., 2001; Fraser *et al*., 2007) or population sweeps caused by periodic natural selection (Cohan, 2001; Acinas *et al*., 2004) have been advanced to explain the emergence and maintenance of sequence-discrete populations. Alternative explanations such as population bottlenecks and random birth/extinction are less favourable and probably applicable to more restricted habitats and microbial groups compared to the groups discussed above, such as the vertically transmitted microbial pathogens (Moran, 2007). An increasing body of evidence indicates that both recombination and periodic selection likely play a role in maintaining the populations recovered in the metagenomic datasets, albeit the relative importance of the two processes remains to the determined. For instance, several studies have reported detectable levels of homologous recombination within marine planktonic (Rusch *et al*., 2007; Vergin *et al*., 2007; Konstantinidis and DeLong, 2008) and terrestrial biofilm natural populations (Tyson *et al*., 2004; Eppley *et al*., 2007). In most of these cases, recombination did not appear to be restricted to a few loci under positive selection but affected all genes in the genome and thus, could serve us a force of homogenization within the population (Fig. 3). Further, the efficiency of homologous recombination is known to decrease with increasing divergence of the recombined DNA segments (Vulic *et al*., 1999; Majewski *et al*., 2000). Thus, recombination could have accounted for the sequence-discrete populations if it was rampant among co-occurring cells and there was a dramatic decrease in recombination efficiency around the 90–95% nucleotide identity. Nonetheless, a precise estimation of the recombination rate and its relative importance compared to point mutation is basically out of reach for current algorithms due to the complications associated with detecting recombination among genetically very similar organisms (Eppley *et al*., 2007; Konstantinidis and DeLong, 2008). Thus, the extent to which recombination drives the diversity patterns observed remains essentially speculative. We recently documented, based on whole-genome sequencing and an approach that did not suffer the previous limitations in reliably detecting recombination, that *Shewanella baltica* strains (heterotrophic, γ-*Proteobacteria*) recovered from the same sample from the Baltic Sea are evolving sexually, i.e. genome-wide recombination is purging more single nucleotide differences than those generated by point mutation (Caro-Quintero *et al*., 2011). These results reveal that recombination could indeed serve as the force of population cohesion under natural settings. However, it remains unclear whether the *S. baltica* case represents a rare event or the norm. More populations and habitats must be analysed before a more complete understanding of the influence of the environment on evolutionary processes such as recombination can emerge.

On the other hand, periodic selection, caused by environmental perturbations, most likely, accounts for the fact that the sequence diversity within the identified sequence-discrete populations varied, depending on the organism considered. For instance, the crenarchaeal population in the surface waters of the Sargasso Sea showed substantially less intra-population diversity compared to its counterpart at 4000 m depth (98–100% vs. 90–100% nucleotide identity based on short WGS read data) (Konstantinidis and DeLong, 2008). *Crenarchaea* primarily reside below the euphotic zone of the oceans ($>$ 150 m) (Karner *et al*., 2001); hence, their presence in the surface waters is presumably linked to the winter upwelling that brings to the surface populations from deeper waters. Thus, the smaller intra-population diversity in the surface vs. the deep populations is best explained by lack of or infrequent strong periodic selection in the remarkably stable environment at 4000 m depth and, in contrast, frequent periodic selection in the surface waters due to continuous environmental fluctuations. Regardless, however, of what the actual mechanism(s) of population cohesion are or what their relative importance is, the members of the populations recovered in the available metagenomic

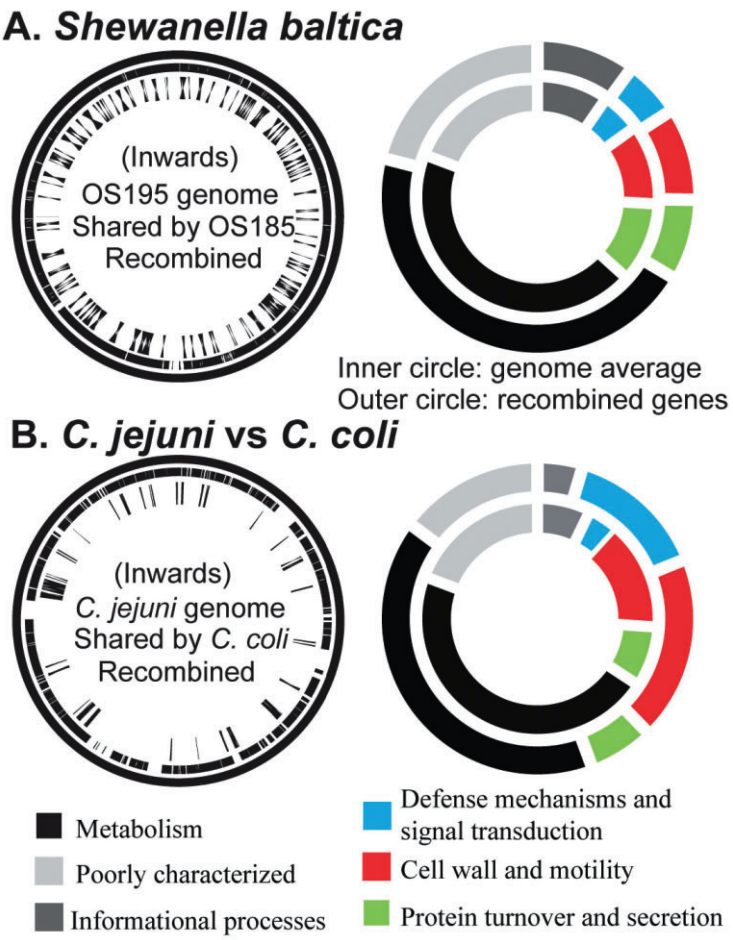

Fig. 3. Patterns of genetic exchange leading to speciation. Genetic exchange among co-occurring *Shewanella baltica* strains, mediated by homologous recombination, occurs throughout the genome (A, left) and does not appear to be biased in terms of the function of the genes recombined (A, right). In such cases, recombination could lead to genome, and hence, population cohesion. In contrast, the closely related (~ 85% ANI) pathogenic *Campylobacter jejuni* and *Campylobacter coli* genomes appear to exchange a limited number of genes that are mostly related to defence mechanisms, antibiotic resistance and membrane proteins for evading host immune response; genes with no obvious ecological advantage are only rarely exchanged in hitchhiking events. In such cases where recombination is apparently driven by positive selection, recombination cannot lead to population cohesion or species convergence but could lead to niche invasion (speciation). Data in A and B are from Caro-Quintero *et al.* (2011) and Caro-Quintero *et al.* (2009), respectively.

data sets appear to somehow cohere together and form sequence-discrete populations.

Implications for the current species definition

The sequence-discrete populations discussed above are characterized by several properties expected for species; namely, they are genetically discrete, tractable based on sequence data, and members of the population show similar ecophysiological properties based on their similar abundances *in situ* and small gene content differences. These results contrast with what was reported previously (Gevers *et al.*, 2005; Hanage *et al.*, 2005); the discrepancy appears to be due, at least in part, to the biases introduced by cultivation and the fact that most of the previous studies analysed collection of isolates recovered from different populations and habitats. Thus, it appears that when genetically homogenous organisms are maintained within a stable niche for sometime, in other words, share a similar ecological trajectory, like some of the abundant marine organisms discussed above presumably do, a sequence-discrete species-like population is expected to (and does) emerge.

Although the findings reported here are primarily based on short sequence fragments and do not encompass gene expression and phenotypic similarity, it is strongly anticipated that if phenotypic data were available, it would have been consistent with the results from the sequence comparisons. In other words, it is expected that the genotypes of a population are highly similar among themselves in several phenotypic properties. Our expectation is based on the long period of coexistence as well as the relatively small sequence and gene content diversity among the genotypes of a population (discussed above); evidently, much smaller compared to strains of several named species. Related to this, recent studies of isolates have revealed that gene content and proteomic profiles are more similar among ecologically more uniform strains compared to strains of the same named species and genetic relatedness that have differentiated ecologically and adapted to leaving in different environments (Konstantinidis *et al.*, 2009b). Thus, the sequence-discrete populations would likely represent distinct species based on the current species definition, which requires discreteness in several independent features in addition to the sequence identity and *in situ* abundance assessed by

metagenomics (Rossello-Mora and Amann, 2001; Stackebrandt *et al*., 2002), as well. Nonetheless, it should be noted that no population phenotypic data is currently available for any of the natural populations discussed above.

Identifying long-lived populations and the exact ecologically trajectory of each genotype in a sample remains challenging, however, primarily due to the very efficient dispersal and slow decay kinetics of microbial cells and the high speed with which microorganisms adapt to new perturbations, especially in highly heterogeneous and dynamic environments such as soils. Therefore, the current practice of defining bacterial species based on genetic similarity, commonly assessed either based on 16S rRNA (or other) gene sequence identity or the relatedness of DNA molecules, and phenotypic similarity, assessed using various techniques under laboratory conditions (Rossello-Mora and Amann, 2001; Stackebrandt *et al*., 2002), represents a pragmatic approach. The findings from metagenomic studies also suggest that the current practice is somewhat reliable but more stringent and ecological standards are both more appropriate and achievable. For instance, species are frequently demarcated using a 70% DNA–DNA hybridization standard (DDH) and it was shown that the 70% DDH standard translates to about 95% genome-aggregate average nucleotide identity, or ANI (Goris *et al*., 2007). Genetic discontinuity among the sequence-discrete populations composing a community is typically observed in the 80–95% ANI range, revealing that the traditional 70% DDH encompasses well the metagenomic findings.

However, the current methods for species demarcation do not take into account the extent of ecological and genomic adaptations of the strain within its habitat of isolation and can not easily detect discernible species among strains sharing more than 95% ANI due to lack of resolving power [as is the case for 16S rRNA gene sequence analysis; see Konstantinidis and Tiedje (2007)]. Metagenomic approaches can not only assess the *in situ* abundance of each organism evaluated and frequently discern closely related incipient species but also reveal the genetic differences among differentially adapted populations as was the case, for instance, for the depth-stratified populations discussed above. It is also important to point out that traditional taxonomic techniques and 16S rRNA gene sequencing can not differentiate robustly *Prochlorococcus* or *Crenarchaea* isolates from different depths of the water column; yet the depth-stratified populations of such strains are clearly distinguishable from each other based on metagenomic data (e.g. Fig. 2). These findings further corroborate the notion that it is time to start supplementing traditional laboratory-based approaches for defining diagnostic phenotypes of new species with omics-based procedures. We would also argue that more emphasis should be given to data from metagenomic and related techniques because such data reflect better *in situ* processes and the evolutionary history of populations compared to data derived from laboratory, batch-culture conditions, which are frequently artificial compared to natural conditions.

Our synthesis of the findings emerging from metagenomic studies suggests that the sequence-discrete populations discussed above represent apparently the key level of diversity to begin high-resolution investigations on how populations interact and evolve within communities, beyond the limitations and caveats of traditional techniques such as 16S rRNA gene sequencing. Incorporating the sequence-discrete populations into the current species definition should also lead to a more reliable and predictive classification system relative to the current practice. Related to the latter, it will be important to perform a full taxonomic analysis of members of several sequence-discrete populations to more knowledgably guide how to incorporate the findings from metagenomics into the species definition and how to best define and identify bacterial species.

Acknowledgements

This research was supported, in part, by the by the National Science Foundation under Grant no. 0919251 and by the U.S. Department of Energy under Grant No. DE – SC0004601.

References

Acinas, S.G., Klepac-Ceraj, V., Hunt, D.E., Pharino, C., Ceraj, I., Distel, D.L., and Polz, M.F. (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430:** 551–554.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143–169.

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., *et al*. (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* **1:** 703–713.

Campbell, B.J., Yu, L., Heidelberg, J.F., and Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci USA* **108:** 12776–12781.

Caro-Quintero, A., Rodriguez-Castano, G.P., and Konstantinidis, K.T. (2009) Genomic insights into the convergence and pathogenicity factors of *Campylobacter jejuni* and *Campylobacter coli* species. *J Bacteriol* **191:** 5824–5831.

Caro-Quintero, A., Deng, J., Auchtung, J., Brettar, I., Hofle, M.G., Klappenbach, J., and Konstantinidis, K.T. (2011) Unprecedented levels of horizontal gene transfer among spatially co-occurring *Shewanella bacteria* from the Baltic Sea. *ISME J* **5:** 131–140.

Cohan, F.M. (2001) Bacterial species and speciation. *Syst Biol* **50:** 513–524.

Coleman, M.L., and Chisholm, S.W. (2010) Ecosystem-specific selection pressures revealed through comparative population genomics. *Proc Natl Acad Sci USA* **107:** 18634–18639.

Eppley, J.M., Tyson, G.W., Getz, W.M., and Banfield, J.F. (2007) Genetic exchange across a species boundary in the archaeal genus ferroplasma. *Genetics* **177:** 407–416.

Feil, E.J., Holmes, E.C., Bessen, D.E., Chan, M.S., Day, N.P., Enright, M.C., *et al*. (2001) Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* **98:** 182–187.

Fraser, C., Hanage, W.P., and Spratt, B.G. (2007) Recombination and the nature of bacterial speciation. *Science* **315:** 476–480.

Garcia Martin, H., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., *et al*. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* **24:** 1263–1269.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., *et al*. (2005) Opinion: re-evaluating prokaryotic species. *Nat Rev Microbiol* **3:** 733–739.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., and Tiedje, J.M. (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57:** 81–91.

Hanage, W.P., Fraser, C., and Spratt, B.G. (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol* **3:** 6.

Handelsman, J., Tiedje, J., Alvarez-Cohen, L., Ashburner, M., Cann, I., Delong, E., *et al*. (2007) *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*. Washington, DC, USA: The National Academies Press.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., and Polz, M.F. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320:** 1081–1085.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001) Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409:** 507–510.

Konstantinidis, K.T. (2011) *Metagenomic Insights into Bacterial Species*. Hoboken, NJ, USA: John Wiley & Sons.

Konstantinidis, K.T., and DeLong, E.F. (2008) Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. *ISME J* **2:** 1052–1065.

Konstantinidis, K.T., and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* **102:** 2567–2572.

Konstantinidis, K.T., and Tiedje, J.M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10:** 504–509.

Konstantinidis, K.T., Ramette, A., and Tiedje, J.M. (2006) The bacterial species definition in the genomic era. *Philos Trans R Soc Lond B Biol Sci* **361:** 1929–1940.

Konstantinidis, K.T., Braff, J., Karl, D.M., and DeLong, E.F. (2009a) Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl Environ Microbiol* **75:** 5345–5355.

Konstantinidis, K.T., Serres, M.H., Romine, M.F., Rodrigues, J.L., Auchtung, J., McCue, L.A., *et al*. (2009b) Comparative systems biology across an evolutionary gradient within the *Shewanella* genus. *Proc Natl Acad Sci USA* **106:** 15909–15914.

Lan, R., and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol* **8:** 396–401.

Lawrence, J.G. (2002) Gene transfer in bacteria: speciation without species? *Theor Popul Biol* **61:** 449–460.

Lawrence, J.G., and Ochman, H. (1998) Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA* **95:** 9413–9417.

Luo, C., and Konstantinidis, K.T. (2011) Phosphorus-related gene content is similar in *Prochlorococcus* populations from the North Pacific and North Atlantic Oceans. *Proc Natl Acad Sci USA* **108:** E62–E63.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011a) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* **108:** 7200–7205.

Luo, C., Tsementzi, D., Kyrpides, N.C., and Konstantinidis, K.T. (2011b) Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* **77:** 6000–6011.

Majewski, J., Zawadzki, P., Pickerill, P., Cohan, F.M., and Dowson, C.G. (2000) Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *J Bacteriol* **182:** 1016–1023.

Medigue, C., Rouxel, T., Vigier, P., Henaut, A., and Danchin, A. (1991) Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* **222:** 851–856.

Moran, N.A. (2007) Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA* **104** (Suppl. 1)**:** 8627–8633.

Oh, S., Caro-Quintero, A., Tsementzi, D., Deleon-Rodriguez, N., Luo, C., Poretsky, R., and Konstantinidis, K.T. (2011) Metagenomic insights into the evolution, function and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl Environ Microbiol*. doi:10.1038/ismej.2011.147.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., *et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59–65.

Rossello-Mora, R., and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25:** 39–67.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., *et al*. (2007) The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Sheppard, S.K., McCarthy, N.D., Falush, D., and Maiden, M.C. (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* **320:** 237–239.

Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C., *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.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., *et al*. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37–43.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al*. (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.

Vergin, K.L., Tripp, H.J., Wilhelm, L.J., Denver, D.R., Rappe, M.S., and Giovannoni, S.J. (2007) High intraspecific recombination rate in a native population of Candidatus pelagibacter ubique (SAR11). *Environ Microbiol* **9:** 2430–2440.

Vulic, M., Lenski, R.E., and Radman, M. (1999) Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc Natl Acad Sci USA* **96:** 7348–7351.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., *et al*. (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37:** 463–464.