16 Genomes Reveal the Cohesiveness of Bacterial Species Taxa And Provide a Path Towards Describing All of Bacterial Diversity

Frederick M. Cohan*
Department of Biology, Wesleyan University, Middletown, CT, USA

Introduction

Scientists in many disciplines rely on the systematics of bacteria – epidemiologists, biotechnologists, agriculturists and microbial ecologists, and evolutionary biologists – and they all place high demands on the classification of bacterial species. All these consumers of bacterial systematics demand a reasonably complete accounting and description of the world's species. I will address how far bacterial systematics has fallen behind in identifying and describing species that we now know exist (Yarza *et al.*, 2014; Garrity, 2016; Locey and Lennon, 2016). An existential problem for systematics is that our current approach to describing species applies only to cultivated bacteria, since it requires a cultivated type strain (Garrity, 2016). Moreover, the current 'polyphasic' approach to species taxonomy requires a labour-intensive characterization of any new species through laboratory testing of its properties (Vandamme and Peeters, 2014). I will discuss how genomic approaches give us a means to catch up, by potentially providing us the full metabolic capacity for any organism, and a means to demarcate organisms into species taxa (Garrity and Lyons, 2011; Thompson *et al.*, 2014; Vandamme and Peeters, 2014; Garrity, 2016; Konstantinidis *et al.*, 2017).

To satisfy evolutionary biologists, systematists have aspired to define species according to an evolutionary theory of species origination. Speciologists of animals and plants have proposed that a species should have certain dynamical properties. Most famously, Ernst Mayr and others have suggested that species should be defined by properties of genetic exchange (Mayr, 1963; Coyne and Orr, 2004), and some bacteriologists have suggested extending that property to bacteria (Dykhuizen and Green, 1991; Cadillo-Quiroz *et al.*, 2012; Bobay and Ochman, 2017). The motivation is that ecological diversification within a species is constrained by genetic exchange, while different species may diverge without limit (Mayr, 1963; Templeton, 1989; Cohan, 2017a).

We shall see that this widely celebrated property of species does not apply reliably to any group of organisms, even to the animals that inspired Mayr's work, and much less to bacteria (Mallet, 2008; Cohan, 2017a; see also Chapters 10 and 15). However, there is another universal property of species cohesion that applies to species of animals, plants and bacteria (and probably beyond) – that recombination prevents neutral

*fcohan@wesleyan.edu

 Trends in the Systematics of Bacteria and Fungi (eds. P. Bridge, D. Smith and E. Stackebrandt)

sequence divergence within a species (Cohan, 2011a, 2019). I will show that bacterial systematists of the mid-20th century fortuitously created a species-level systematics that actually fits an important universal theory of speciation.

Finally, I will discuss what may be the most exacting demand on species taxonomy. The animal ecologist G. Evelyn Hutchinson aspired that every species should be homogeneous in its biochemical, physiological, morphological and ecological characteristics. He argued that such a taxonomy would allow us to infer the important characteristics of any unknown organism once we classify it to species (Hutchinson, 1968). In the case of bacteria, many microbiologists may agree that we would benefit from a taxonomy that classifies bacteria to groups that are uniform in the characteristics we care most about. For example, we might want a pathogenic taxon to be uniform in its tissue tropisms, host range, disease aetiology and so on. However, microbiologists understand that the typical bacterial species taxon houses a substantial level of ecological diversity (Konstantinidis *et al.*, 2006; Cohan and Kopac, 2017). There are many good reasons not to move to a higher-resolution species taxonomy that would abide by Hutchinson's aspiration. However, I will discuss ways in which we can extend infraspecific systematics of close relatives to better abide by the Hutchinsonian aspiration of homogeneous taxa.

How Taxonomy Demarcates Bacterial Species

Each species experiences its world in its own unique way, and has evolved a special sense of its surroundings, known as an 'umwelt' (Yoon, 2009). In the case of our species, we inherited an umwelt from our hunter-gatherer past that enabled our ancestors to distinguish animal and plant species, driven by the need to distinguish beneficial and safe animal and plant species from their close relatives that were dangerous. Unfortunately, we have no umwelt for distinguishing bacteria because they are so new to us (Cohan, 2011a).

Lacking an umwelt for bacteria, mid-century microbiologists set the foundation for species systematics based on the metabolic and chemical traits that were then available. By borrowing the methods of numerical taxonomy developed by zoologists and botanists, bacterial systematists developed a new intuition for using phenotypes to estimate relatedness of organisms. Through numerical taxonomy, they sought to identify meaningful clusters of similar organisms (Sneath and Sokal, 1973), and these clusters became the species of bacteriology (Holmberg and Nord, 1984).

In principle, mid-century bacterial systematists could have defined their species narrowly, for example by distinguishing species by subtle, quantitative differences in metabolic capacities (Cohan, 2002). Instead, they made a pragmatic decision early on to include, within a species, strains that were hugely heterogeneous in the presence versus absence of many metabolic capabilities (Rosselló-Móra and Amann, 2001; Cohan, 2011a).

There was no reason to believe that these species should abide by any evolutionary theory about the properties of species, nor was there any desire to do this. However, we will see that these species fortuitously share a universal, species-like, dynamic property with species of animals and plants (Jain *et al.*, 2018; Cohan, 2019).

Beyond metabolic analyses, a series of molecular tools has contributed profoundly to bacterial systematics. DNA sequencing has revealed polyphyletic groups; that is, evolutionarily distant groups that were mistakenly identified as a single taxon on the basis of having converged on the same phenotype. For example, phylogenetic study of the 16S rRNA gene sequence revealed that the bacteria assigned to the *Caulobacter* genus (based on sharing the stalked phenotype) are really a set of divergent taxa that independently evolved the stalk (Stackebrandt *et al.*, 1988). Also, phylogenetic study of the 16S rRNA molecule was able to place all cellular organisms on a single tree (Woese, 1987). Universal trees have come a long way since then by utilizing a large set of universal, single-copy genes (Zhu *et al.*, 2019).

An early whole-genome approach to analyzing relatedness was able to corroborate the mid-century metabolic species clusters. That is, DNA-DNA hybridization (DDH) could estimate the percentage of the genome shared between a pair of strains (De Ley, 1970), and it turned out that sharing more than 70% of the genome gave about the same species demarcations as metabolic clustering (Wayne *et al.*, 1987). This was

the first in a series of molecular traits that was calibrated to yield the species previously demarcated by numerical taxonomy of metabolic traits.

Following DDH, systematists brought various molecular tools into the systematics of species demarcation, and they calibrated each to yield the earlier metabolic clusters (Thompson *et al.*, 2014; Jain *et al.*, 2018). Sequence identity at the 16S rRNA gene locus was calibrated to yield the metabolic species demarcations, first with a criterion of 97% sequence identity (Stackebrandt and Goebel, 1994) and then at 98.5% (Stackebrandt and Ebers, 2006). More recently, multilocus sequence analysis, with clusters based on sequence identity of seven or so shared genes, has also corroborated the species taxa based on metabolic sequence clusters (Gevers *et al.*, 2005). A multilocus approach yields an advantage over 16S rRNA in providing greater resolution for discovering significant within-species diversity (Gevers *et al.*, 2005). Moreover, multilocus approaches are less likely to misclassify when a given marker has recombined.

Whole-genome sequencing has given new opportunities for species-level systematics (see Chapter 13). First, the extent of gene content sharing, which DDH measures only indirectly, can be estimated directly by comparing whole-genome sequences (Auch *et al.*, 2010). While DDH was limited by requiring pairwise measures of genomic distance, and by requiring special expertize, whole-genome sequencing allows an incremental increase in the taxonomy of species. Each new species can be added to an existing database, without the need for comprehensive pairwise experiments every time a new species is added to the taxonomy (Garrity, 2016).

Additionally, whole-genome sequencing takes clustering by multiple loci to the limit, with the potential to take into account the sequence identity levels of thousands of shared genes. In 2005, with the prospect of extremely cheap, whole-genome sequencing on the horizon, Kostas Konstantinidis and his colleagues presciently developed whole-genome average nucleotide identity (ANI) as a measure of relatedness (Konstantinidis and Tiedje, 2005). Like each of the earlier molecular markers of relatedness, whole-genome sequencing was calibrated to yield the old metabolically defined species. This team found that an ANI value of 95% closely approximated the existing species demarcations of bacterial systematics.

A recent study by Matt Olm and colleagues found the optimal cutoff for yielding the species taxa with various genes and sets of genes, and it rated each marker for its ability to delineate the existing species taxa of bacterial systematics (Olm *et al.*, 2019). They found that the optimal level of 16S rRNA divergence for recalling the existing species taxa was 99% (similar to the suggestion by Stackebrandt and Ebers, 2006). Notably, 16S rRNA was the least discerning of all the single-gene markers that they studied, and the ribosomal protein L6 was the most discerning single-gene marker. However, an ANI value of 94.5% yielded the highest delineation of species of any molecular marker, far greater than for any single gene.

Systematists are currently adopting ANI as a means for classifying new isolates and metagenome-assembled genomes to existing species and for discovering new species. For example, nearly 2500 published genomes in the *Lactobacillus* genus were recently classified by ANI (Wittouck *et al.*, 2019). Wittouck and colleagues demarcated the entire set based on 94% ANI, a value lower than that recommended (Jain *et al.*, 2018; Olm *et al.*, 2019). Their analysis merged several species, and allowed discovery of eight previously uncharacterized species. Similarly, Kanny Diallo and colleagues applied ANI to discover and classify the full extent of human-infecting *Neisseria* species (Diallo *et al.*, 2019).

There is also a pseudo-genomic approach similar to multilocus sequence analysis. Here, a small sample of genes from whole-sequence genomes is concatenated to yield sequence clusters. For example, a recent survey of *Acinetobacter* was based on a concatenation of 13 genes selected from whole-genome sequences (Mateo-Estrada *et al.*, 2019).

For decades now, and over generations of molecular techniques for demarcating species, systematists have applied a 'polyphasic' approach to identifying and describing species. This approach has sought to characterize novel species as fully as possible to reach a consensus among molecular and phenotypic traits (Vandamme *et al.*, 1996, Tindall *et al.*, 2010). Here, laboratory tests of physiology and metabolism are compared with molecular analyses with the aim of creating a stable taxonomy with a minimum of contradiction. While no one would argue that more information about a taxon is unhelpful,

there is a growing concern that polyphasic taxonomy places too high a standard on the quality of diagnostic information for a species taxon (Vandamme and Peeters, 2014).

The problem with polyphasic taxonomy is that bacterial systematics is almost hopelessly behind in bringing all the species we know exist into our taxonomy. Pablo Yarza and colleagues have pointed out that the number of species taxa that have been discovered and are not yet classified is increasing steadily. At the current rate of describing species, systematists would require thousands of years to classify just the species that we currently are aware of (Yarza *et al.*, 2014). High-throughput sequencing has made us aware of this problem but, as we will see, high-throughput sequencing can also solve the problem by giving us full genomes.

This is because genomes are more than a high-resolution method of distinguishing clusters of bacteria. They can, potentially, also provide the full metabolic capabilities and ultimately the physiology and ecology of bacteria. As a result, some systematists and microbial ecologists are eager to replace physiological testing with a genome-based estimate of a bacterium's capabilities (Garrity, 2016; Konstantinidis *et al.*, 2017). Let us next consider how a genome-based species taxonomy would serve the microbiological community.

A Genome-based Species Taxonomy

A genome-based species taxonomy will need to consider three points: (i) how to obviate the need for a cultivated type strain; (ii) how to demarcate genomes into new species; and (iii) how to describe and recurrently update the phenotype and diagnostic criteria for new species as new data become available.

Substituting a type genome sequence for a type strain

Recognizing the limitations of a culture-based systematics at the species level, microbiologists are increasingly demanding a genome-based route to characterizing new species (Garrity and Lyons, 2011; Thompson *et al.*, 2014; Vandamme and Peeters, 2014; Garrity, 2016; Konstantinidis *et al.*, 2017). One solution is to relax Rules 27 and 30 of the taxonomic code (Garrity, 2019), such that a genome sequence can be substituted for a type strain. Supporters of reform argue that this change will further the democratization of systematics, such that anyone who can sequence and analyse genomes will be able to demarcate novel species and characterize their metabolic features (Garrity, 2016).

Those genome sequences based on a single uncultivated cell would make the most reasonable substitute material for a type strain. Characterizing the metabolic and ecological features of the type sequence would be comparable to characterizing those features from study of an isolate.

However, I will argue that a metagenome-assembled genome should be considered as type material only for a candidate taxon. The problem is that a metagenome-assembled genome (MAG) is based on a concatenation across reads from multiple organisms. Therefore, a MAG could contain multiple ecologically distinct populations (ecotypes) (Nelson *et al.*, 2016), considering that ecotypes appear to have as little as 1% divergence in ANI (Konstantinidis and Tiedje, 2005), and possibly less (Cohan, 2016a).

Whether a genome sequence is based on a single cell or from a metagenome, any metabolism inferred from the genome should be considered tentative until assayed directly in laboratory tests. Taking into account both the limited opportunities for funding and the urgencies of the science, we should at least aspire to eventually confirm the inferred phenotypic features of a type genome sequence. George Garrity has anticipated that descriptions of species based on genomes will be in flux, and he and his coworkers have proposed and patented the Names for Life database (www.namesforlife.com/search, accessed 30 July, 2020), which allows continual updating of the data resources of a taxon. This includes a record of the original description as well as further taxonomic and nomenclatural events relating to the taxon (Garrity and Lyons, 2011).

Demarcating genomes into new species

Because ANI delineates genomes into the recognized species taxa more accurately than any

other molecular approach (Olm *et al.*, 2019), it is reasonable to demarcate genomes of uncultivated bacteria into species by ANI (Varghese *et al.*, 2015; Parks *et al.*, 2019). One protocol is to apply complete-linkage clustering of ANI values to demarcate the genomes. That is, a new species would be demarcated such that all pairwise distances yield > 95% ANI (Varghese *et al.*, 2015).

In an alternative classification, systematists would base the demarcation of each novel species by sequence identity with the type genome for the species (Parks *et al.*, 2019). That is, after the type genome for a novel species is chosen, other genomes would be added to the species based on having an ANI value > 95% of the type genome. This would capture the central importance of type strains in taxonomy.

Although ANI can now be calculated extremely quickly with the new FastANI algorithm (Jain *et al.*, 2018), there is still a need to improve the speed of classifying millions of novel species by their ANI values. One recent approach is to prescreen for close relatives by tetranucleotide composition, and to then apply ANI only to close relatives (Zhou *et al.*, 2020).

Describing the phenotype of novel species

Genomes provide a trove of information for describing the metabolic, chemical and ecological properties of species taxa. Most straightforwardly, the standard phenotypic traits of polyphasic studies can be estimated by analyzing gene content. For example, the capacity for the Voges–Proskauer reaction, indole production and utilization of any number of carbon sources can be tentatively determined through gene content analyses (Thompson *et al.*, 2014).

Using genomes to characterize novel bacteria for their tolerances of physical and chemical conditions is more challenging. However, microbial ecologists are making progress towards identifying genes that confer complex phenotypes. In genome-wide association studies, one correlates the genome content with phenotypes among close relatives, to identify genes responsible for the phenotypic differences. For example, the Traitar algorithm can predict 67 phenotypic traits from genomes from various phyla (Weimann *et al.*, 2016). The algorithm was based on correlating phenotypic data from the GIDEON database (Berger, 2005) with gene content from sequenced genomes. Others have predicted phenotypes from genome variation within a species, for example by predicting invasiveness and resistance among strains of *Neisseria meningitidis* (Collins and Didelot, 2018). Various approaches to identifying genes responsible for complex traits such as salt and pH tolerance in laboratory studies promise future predictions of phenotypes (Hahne *et al.*, 2010; Mirete *et al.*, 2015; Barberán *et al.*, 2017). Genes that promote successful interactions with other bacterial species have also been identified from laboratory studies (He *et al.*, 2017).

Genome-wide association studies (GWAS) can, in principle, be used more generally to discover the genes conferring the ability to live in various habitats (Dutilh *et al.*, 2013). For example, correlation of genes in *Novosphingium* with diverse habitats including rhizospheres, contaminated soils, freshwater and marine water, yielded the discovery of genes consistently associated with each habitat type (Kumar *et al.*, 2017). We should note that GWAS analyses can be performed retrospectively by future investigators, but only if microbiologists are careful to publish detailed accounts of the environments from which they isolate novel organisms. Following the MIxS protocols for describing habitats will become especially important as we try to characterize ecological abilities from GWAS (Cohan, 2011b; Yilmaz *et al.*, 2011). This open-ended discovery of genes responsible for ecological adaptations will contribute to estimating the phenotype from the genomes of unclassified strains.

In short, a genome-based systematics will allow us to demarcate novel, uncultivated species that are similar in their phylogenetic breadth (i.e. down to 95% ANI) to the traditional species, and we can in principle obtain a tentative outline of their phenotypic features from genomes. Moreover, I will discuss how the species we have classified up to now, as well as the species classified by their genomes, will abide by a universal theory of the dynamic evolutionary properties of species that a species should hold (Cohan, 2019).

Let us next consider the properties that evolutionary biologists expect for species.

Is There Something Real About Species?

Systematics begins with the observation that life's diversity is organized into clusters of related organisms that are similar in structure, function and genomic properties. These phenotypic and genetic clusters are found from the most complex organisms to the prokaryotes and at all levels of life's diversity, from the domains and phyla to species (Mallet, 1995; Caro-Quintero and Konstantinidis, 2012). Between these clusters are gaps that represent intermediate phenotypes, which we can imagine, but do not actually exist in the natural world (Wilkins, 2009; Cohan, 2013). This pattern of clusters and gaps reflects the genealogical continuity of all organisms, taking into account that some lineages have been extremely successful, while nearly every lineage that has ever existed has gone extinct.

The higher ranks of systematics take into account obvious gaps in phenotype among closely related genera, families and so on. Systematists generally agree that these higher ranks (above the species level) are simply a convenience for consumers of systematics (McDonald *et al.*, 2012). That is, systematists and evolutionary biologists have not hypothesized any dynamic force that would apply within a genus, for example, that would not apply across different genera (Cohan, 2017b).

Nevertheless, there has been recent interest in determining a universal criterion for how much diversity should be included within each taxonomic rank. To this end, Donovan Parks and colleagues have developed a universal method for reclassifying organisms so that every genus contains organisms that have diverged for the same amount of time (Parks *et al.*, 2018). In the case of genera, for example, each taxon is reclassified so that it contains organisms that have diverged up to 7% of the time since the last common ancestor of all of life.

Among the most contentious issues in systematics is whether there is some biological reality to species. Some systematists believe that species are no more biologically real than the higher taxa (Hey, 2001; Doolittle and Zhaxybayeva, 2009), while others hold that there is something special about species – that they hold certain dynamic properties that transcend human attempts at classification (Mayr, 1942; de Queiroz, 2005; Cohan and Perry, 2007). Among these proposed properties are that each species is ecologically distinct and irreversibly separate from other species, and that each species is cohesive in that some force constrains diversification within a species (de Queiroz, 2005; Kopac *et al.*, 2014). We will see that there has been recent progress in showing that bacterial species taxa (as well as plant and animal species) are cohesive, although not in the way that most evolutionary biologists had expected (Jain *et al.*, 2018; Cohan, 2019).

I will next consider how recombination and selection can act as forces of cohesion within and between bacterial populations.

Recombination Does Not Prevent Ecological Divergence Between Bacterial Populations

The first force of cohesion proposed for species was genetic exchange. Ernst Mayr and Theodosius Dobzhansky defined species such that populations within an animal or plant species could exchange genes at some high frequency, but that members of different species could not (Mayr, 1942; Dobzhansky, 1951). They argued that this pattern would limit the divergence among populations of the same species but not the divergence between different species. Thus, populations could diverge without bound only when they break free of their recurrent recombination, through evolving sexual isolation. The term 'Mayr's brake' was applied to the action of recombination in stifling the adaptive divergence between populations of the same species. The concept of Mayr's brake ruled with hegemony over the thinking of animal and plant speciologists throughout the 20th century, and its influence still rules to some extent over bacterial speciology (Bobay and Ochman, 2017; Cohan, 2017b).

However, nearly a century ago the population geneticist J.B.S. Haldane noted an essential problem with Mayr's brake (Haldane, 1932). He showed mathematically that a recurrent trickle of gene flow (exchange of genes) between populations adapted to different circumstances could have only a negligible effect on the abilities of the populations to maintain their unique adaptations.

That is, if c_b is the rate of recombination between populations (frequently called m, for migration) and s is the selection intensity against migrant alleles, then the equilibrium frequency of a maladaptive, migrant allele in a population is c_b/s, which would be tiny for any set of populations with limited recombination between them (Vos and Didelot, 2009; Cohan, 2011a). James Mallet recently argued that even adjacent populations of animals or plants that are adapted to different environments can diverge without hindrance from recombination (Mallet, 2008).

Recombination is exceedingly unlikely to hinder adaptive divergence between ecologically distinct populations of bacteria. This is because the rate of recombination in bacteria is extremely low, even within populations, hovering within an order of magnitude or two of the mutation rates, around 10^{-6} per gene per generation (Vos and Didelot, 2009). Thus, even if different populations were to recombine at the same rate as cells of the same population, the equilibrium frequency of a foreign allele would be negligible. This is to say that two populations could diverge even if they were in exactly the same place (e.g. two populations living on different soluble compounds in the same aquatic environment) and recombining at the same rate between as within populations. I have, therefore, argued that the evolution of sexual isolation is not a milestone in the ecological divergence of bacterial lineages (Cohan, 1994).

We should expect, then, that one bacterial population should be able to diverge into two ecologically distinct lineages, even without any geographic separation or any kind of reduction in their recombination rate. Thus, laboratory evolution experiments have repeatedly brought diversification of the founding lineage into multiple populations within a culture flask, owing to specialization on different soluble resources (Treves *et al.*, 1998; Blount *et al.*, 2012) or to specialization on different microhabitats within the same flask (Rainey and Travisano, 1998; Koeppel *et al.*, 2013). Likewise, surveys of ecological diversity within natural habitats have demonstrated ecological divergence among extremely closely related strains (Shapiro and Polz, 2014).

Some researchers (Cadillo-Quiroz *et al.*, 2012; Polz *et al.*, 2013; Kashtan *et al.*, 2014) have argued that recombination must be reduced before bacterial lineages can diverge into different species (Cohan, 2016a). Their evidence was that the closely related, ecologically distinct clades that were the focus of their studies showed reduced recombination between them, compared to recombination rates within them. However, these studies did not consider whether there was ecological diversification *within* each of the focus clades, where recombination may have been higher (Melendrez *et al.*, 2016). Our work with hot spring *Synechococcus* demonstrated ecological divergence among extremely close relatives, even those with the highest levels of recombination (Melendrez *et al.*, 2016). Given this evidence, as well as the theoretical expectation that reduced recombination is not necessary for ecological divergence, we may conclude that sexual isolation is not likely to be a necessary step for adaptive diversification of bacteria.

Let us next consider what might be the most significant forces of cohesion in the bacterial world.

Periodic Selection as a Force of Cohesion in Bacterial Species

One important force of cohesion for bacteria is periodic selection. Because recombination rates in bacteria are so low (Vos and Didelot, 2009), natural selection favouring an adaptive gene within an ecologically homogeneous population (or ecotype) can reduce the genome-wide genetic variation within the ecotype to near zero (Cohan, 1994). However, because different ecotypes are ecologically distinct, a periodic selection cannot purge the diversity genome-wide across ecotypes (Cohan, 2017a).

Genome-wide selective sweeps have been observed in the bacterial world, as expected for periodic selection events acting within a single ecotype. A metagenomic survey of diversity in a bog lake has yielded the first direct evidence of genome-wide sweeps in nature (Bendall *et al.*, 2016). However, more frequently genomic (Bhaya *et al.*, 2007; Shapiro *et al.*, 2012) and metagenomic (Bendall *et al.*, 2016) surveys have revealed evidence of single-gene sweeps (or sweeps over only a short segment of the chromosome) within a sequence cluster. These results appear at first to demonstrate that recombination

is sufficient to prevent genome-wide purges of diversity within any given population (Papke *et al.*, 2004; Shapiro and Polz, 2015). However, my colleagues and I have previously argued that single-gene sweeps do not occur within just a single population. Instead, single-gene sweeps are most likely involve the transfer of one generally adaptive gene segment across all the ecotypes within a sequence cluster (Majewski and Cohan, 1999a; Kopac and Cohan, 2012; Cohan, 2016a).

Ecotypes as Species-like Lineages

Bacterial ecotypes meet a diversity of aspirations of what a species should be (Ward, 1998; Koeppel *et al.*, 2008; Sikorski, 2008). Because ecotypes are each ecologically homogeneous, the ecotypes reach Hutchinson's call for a species taxonomy that allows a precise description of any unknown organism that is classified to an ecotype (Hutchinson, 1968). The ecotype also reaches the speciologists' aspiration that a species should hold species-like properties – that ecotypes are each ecologically homogeneous and cohesive, and that different ecotypes are ecologically distinct and irreversibly separate (Cohan, 2017a).

Ecotypes have been a target of study for microbial ecologists because they represent the most newly divergent, ecologically distinct populations of bacteria (Koeppel *et al.*, 2008; Martiny *et al.*, 2009; Becraft *et al.*, 2015; Chase *et al.*, 2019). Ecotypes may be tentatively discovered as closely related lineages that form distinct sequence-based clusters (Koeppel *et al.*, 2008; Martiny *et al.*, 2009; Wood *et al.*, 2020). The hypothesized clusters may then be confirmed to be ecologically distinct, most easily by finding that they are substantially different in their habitat associations (Cohan, 2017a). Microbial ecologists have found that closely related ecotypes can differ in the chemical and physical conditions to which they are adapted (Connor *et al.*, 2010; Denef *et al.*, 2010; Becraft *et al.*, 2015; Thompson and Kouba, 2019) or in the resources that they consume (Hunt *et al.*, 2008; Kopac *et al.*, 2014; Ramírez *et al.*, 2020). A sampling of the ecological dimensions along which infraspecific ecotypes have diverged include solar exposure and soil texture in desert *Bacillus* (Connor *et al.*, 2010), temperature and depth in hot spring *Synechococcus* (Becraft *et al.*, 2015), host specificity within *Agrobacterium tumefaciens* (Lassalle *et al.*, 2011) and adaptation in *Alteromonas macleodii* to marine environments with different levels of organic content (Koch *et al.*, 2020).

There are generally many ecotypes within a bacterial species taxon that are recognized by bacterial systematics (Staley, 2006; Hunt *et al.*, 2008; Connor *et al.*, 2010; Cohan, 2016b). For example, the marine species *Vibrio splendidus* was found to have 15 ecotypes, which were confirmed to differ by the size of the particle on which they were sampled and by their seasonal abundance (Hunt *et al.*, 2008).

One may argue that ecotypes are the true species of the bacterial world, for being ecologically homogeneous and having periodic selection as a force of cohesion that limits their diversity. However, this would severely disrupt the stability of bacterial taxonomy. I will, instead, discuss the prospects for enriching bacterial systematics by including ecotypes as infraspecific taxa.

Enriching Bacterial Systematics with Ecotypes

Within species taxa, both bacteria and higher organisms have diverged to form ecotypes that are adapted to different conditions and resources. Botanists developed the concept of 'ecotype' to represent populations in different locations that have diverged in their local adaptations (Turesson, 1922; Clausen *et al.*, 1947). Whenever botanists study a single location of plants belonging to one species, they are not confused by the exuberance of ecotypes within the species, because the ecotypes tend to be in different places. When we study a collection within an animal or plant community, we know for example that all the fruit flies of *Drosophila melanogaster* from one site represent one evolutionary unit. The existence of ecotypes is not generally a confusion for botanists and zoologists.

On the other hand, the proliferation of ecotypes is much more confusing for bacteriology. When we study *Escherichia coli* strains isolated from one habitat (even from one microhabitat), there can be any number of ecotypes subsumed

within the collection (Cohan and Kopac, 2011; Luo *et al.*, 2011). Because bacterial ecotypes can differ quantitatively in their habitat preferences, any one microhabitat could include various ecotypes, some specialized to different resources in the same microhabitat and others specialized to different microhabitats (Hunt *et al.*, 2008).

There are significant pitfalls in reifying a pool of ecotypes into one bacterial species. One problem is that population geneticists may incorrectly estimate population sizes and migration rates from sequence data when they focus on an entire bacterial species taxon. Population genetic estimates work well for a typical animal species, when we can assume that all individuals from a region are members of the same evolutionary unit (Ho and Shapiro, 2011; Volz, 2012). The principle of estimating the size of a single population is that, as population size increases, genetic drift will have lower potential to reduce the sequence diversity of the population. This works well for the *Drosophila melanogaster* fruit flies of a region, where we can reasonably figure that the sequence diversity of the species is limited by genetic drift. However, the sequence diversity within a bacterial species taxon containing multiple ecotypes is determined only very little by the population size of any one ecotype. Instead, the sequence diversity of the whole species taxon is determined mostly by the time that ecotypes have diverged from their common ancestor (Cohan and Kopac, 2017). Mistaking a bacterial species taxon for a single evolutionary unit has recurrently introduced errors into population genetic estimates (Roberts and Cohan, 1995; Bobay and Ochmann, 2018).

My colleagues and I have previously discussed how classifying organisms to ecotypes may bring practical benefits for biotechnologists who are looking for close relatives of a useful strain that may differ in its optimal conditions; the search for a vaccine may also benefit from an ecotype-based taxonomy (Cohan and Kopac, 2017). Most generally, when we define a species taxon so broadly as to include many ecotypes, we reduce the opportunity for a full exploration of the metabolic, physiological, ecological and genomic diversity within the species. Systematics would, therefore, benefit from official recognition of the ecotypic diversity within a species taxon.

The *International Code of Nomenclature of Prokaryotes* allows for infraspecific classification (Parker *et al.*, 2019), and gives a path for inclusion of ecotypes in taxonomy. Rule 13a–d regulates subspecies taxa, and Rule 14a states that taxa below the subspecies level are not regulated (Garrity, 2016; Parker *et al.*, 2019). Systematists have regularly applied infrasubspecific labels to describe diversity within a species or subspecies. These labels include pathovar, serovar, phagovar, biovar, chemovar and morphovar (among others) (Parker *et al.*, 2019). While each of these labels applies specifically to certain kinds of variants (e.g. biovar applies to symbionts of different plant species), ecovar could apply more generally to closely related phylogenetic groups within a species that are ecologically distinct in any way (Cohan, 2006).

I will next consider possible criteria for introducing an ecovar to the taxonomy. The proposed ecovar should be identifiable as a sequence cluster, ideally demarcated by an algorithm intended to discover ecologically distinct groups, such ecotype simulation (Koeppel *et al.*, 2008; Wood *et al.*, 2020), AdaptML (Hunt *et al.*, 2008), or Generalized Mixed Yule Coalescent (GMYC) (Barraclough *et al.*, 2009). Alternatively, ecotypes could be revealed by whole-genome clustering at an ANI level of about 99.5% (Konstantinidis and Tiedje, 2005). Ecotypes hypothesized from sequence analysis should then be confirmed, at least tentatively, by quantitative differences in habitat associations. We need to keep in mind that very closely related ecotypes are frequently not totally divergent in their habitat preferences, and so we should not discount habitat differences that are not complete (Hunt *et al.*, 2008). It is worth noting that if a survey of diversity within a species taxon provides the habitat provenance of each strain, future researchers can later add ecovar descriptions.

Also, analysis of genomes can more fully confirm the ecological distinctness of ecotypes hypothesized by sequence clustering analyses. Michiel Vos has argued for using genome-wide analyses of positive selection to identify ecotypes (Vos, 2011). That is, when two closely related lineages show different histories of adaptive evolution in their shared genes, we may conclude that they are ecotypes (Kopac *et al.*, 2014). Also, differences in genome content can predict ecological differences among hypothesized ecotypes (Lassalle *et al.*, 2011).

Bacterial systematics will be enriched by including ecotypes where it is convenient and

useful to describe them. Perhaps most appealing to microbial evolutionary biologists is that the ecotypes would represent taxa with the species-like properties of each being cohesive and ecologically homogeneous. Also, ecotypes are the most newly divergent lineages that can diverge without bound in their adaptations (Cohan, 2017a).

However, I will argue that ecotypes are not the only bacterial groups to abide by a universal species concept. Recent developments in comparative genomics have indicated that the more inclusive and recognized species taxa, originally based on metabolic and chemical features alone, surprisingly follow a different criterion of species dynamics.

Recombination as a Force of Cohesion Among Ecologically Distinct Lineages

While recombination does not act to hinder adaptive divergence among bacterial lineages, recombination can nevertheless act as a force of cohesion in a different way. In the world of animals and plants, recombination can act to spread an adaptive gene from one population to another, within a species or between extremely closely related species that can interbreed. In the process of 'adaptive introgression', a generally useful adaptation can spread between populations that have adapted to different environments while the populations maintain their adaptive divergence. For example, in the case of humans, an adaptive allele for lactase persistence (allowing lifelong consumption of fresh cow milk in pastoral societies) spread between North Africa and Scandinavia, while allowing each population to maintain its adaptations to its particular environment (in this case, maintaining adaptive differences in skin colour) (Gerbault *et al.*, 2011). Also, in our species, Tibetans acquired some of their adaptation to high elevations through adaptive introgression from Denisovans, an extinct sister species with whom humans once interbred (Huerta-Sánchez *et al.*, 2014).

This process of adaptive introgression takes many generations in animals and plants because it involves a meiotic recombination of whole genomes, followed by natural selection of the hybrids and backcrosses that contain the adaptive recombinants. However, in bacteria, transferring a single adaptive gene or a set of chromosomally linked adaptive genes can be immediate. This is because the size of recombined segments is usually much smaller than the full genome, and can contain just the generally adaptive genes (i.e. that are adaptive across different populations) without bringing along the whole genetic baggage of the donor cell (Zawadzki and Cohan, 1995; Wiedenbeck and Cohan, 2011).

We can view recombination among close relatives as a cohesive force because of its ability to spread generally adaptive genes across ecologically distinct populations. Thus, closely related but ecologically distinct populations that recombine frequently have the greatest opportunity to share generally adaptive genes, as proposed in the Adapt Globally Act Locally model (Majewski and Cohan, 1999a, 1999b; Cohan, 2016a). In whole-genome comparisons, Martin Polz and his colleagues found many examples of a single gene (or closely linked set of genes) that swept across ecologically distinct populations within a *Vibrio* species taxon (Shapiro *et al.*, 2012; Arevalo *et al.*, 2019). Likewise, a metagenomic survey of diversity has shown many cases of such single-gene sweeps across ecologically distinct populations (Bendall *et al.*, 2016; Cohan, 2016).

To sum up, in both the bacteria and higher organisms, recombination can act as a force of cohesion that allows the sharing of adaptations among close relatives. We can add that ongoing, recurrent recombination can also act as a force of cohesion for neutral sequence variation. That is, recurrent transfer of ecologically interchangeable, neutral variants across ecotypes may homogenize the ecotypes' sequences genome-wide (with the exception of genes involved in adaptive divergence). Ecotypes will homogenize nearly genome-wide when homologous recombination of neutral variants within a species taxon occurs faster than sequence divergence by mutations, causing a limit on the accumulation of neutral diversity (Palmer *et al.*, 2019).

A Force of Cohesion That is Limited to Species Taxa Across Much of Life

Fortuitously, the systematics laid down by the mid-century microbiologists appears to have

delineated bacterial species taxa that are cohesive in the same way as animal and plant species taxa. Adaptive divergence within any species taxon, whether bacterial, animal or plant, does not appear hindered by recurrent recombination, yet recombination within animal or plant species appears to homogenize variation within species for genes not involved in adaptive divergence. In the case of bacteria, comprehensive comparisons of whole bacterial genomes have revealed that there is a similar cohesion acting within the broadly defined species taxa of bacteria, even though each species taxon is ecologically heterogeneous with a great diversity of ecotypes (Jain *et al.*, 2018; Cohan, 2019).

Early metagenomic studies indicated a genome-wide cohesion among closely related but ecologically distinct bacteria. This was suggested by early 'tiling' studies, in which the genome sequence of one isolate from an environment was compared to random sequence reads from the same environment. For example, when my colleagues and I compared the genomes of each of two *Synechococcus* isolates from a Yellowstone hot spring to the metagenome reads from the same hot spring, the environmental reads were either very closely similar to the isolate (> 95% identical) or they were much more divergent (usually < 80%) (Bhaya *et al.*, 2007); other tiling experiments have yielded similar results (Caro-Quintero and Konstantinidis, 2012). These results suggested that some force acted cohesively to bacteria within 95% identity, perhaps slowing down their rate of divergence. Beyond 95% identity, it would seem that lineages are able to diverge freely at a faster rate (Cohan, 2019).

The first study of ANI, based on a limited set of genomes in 2005, showed that 95% sequence identity has a special significance for bacterial taxonomy (Konstantinidis and Tiedje, 2005). An ANI value of 95% appeared to be a universal molecular criterion of relatedness that yielded the metabolically delimited species taxa of mid-century systematics.

More extensive studies of genomes and metagenomes corroborated that there was a special cohesion of some kind that extends down to 95% sequence identity. Konstantinidis and colleagues analysed the ANI between all 90,000 pairs of sequenced genomes that were available at the time (Jain *et al.*, 2018). A more recent study, using both cultivated and uncultivated genomes, has confirmed a universal gap in ANI, where many close relatives span values from 95% to nearly 100%, and extremely few pairs of organisms are found between 83% and 95% ANI (Olm *et al.*, 2019). The species taxa of bacterial systematics appear universally to be cohesive groups, with some force limiting divergence within them but not (or to a much lesser extent) between them (Cohan, 2019).

It should surprise microbial ecologists that bacterial species taxa would be cohesive, given that they are each heterogeneous in their physiology, ecology and genome content (Cohan, 2016c). The species taxa each appear to be rife with ecotypes (Connor *et al.*, 2010; Becraft *et al.*, 2015), and each ecotype has its own force of cohesion through periodic selection events, which constrains the diversity within each of them (Bendall *et al.*, 2016; Cohan, 2016a).

Homologous recombination may be a force hindering divergence among relatives down to 95% ANI. One possibility is that homologous recombination is passing interchangeable sequence variants between the various ecotypes within a given species taxon. A recent study by Bobay and Ochman (2017) searched widely for recombination events between genome pairs from the same and different species taxa, and found that the rate of recombination was much higher within than between. While their study was not framed in terms of ANI, we can note that recombination rates decreased by orders of magnitude in groups with less than 95% ANI. Olm and colleagues extended the study of recombination to metagenome-assembled genomes from diverse habitats, and found a very similar result. That is, recombination decreased by orders of magnitude down to 95% ANI, at which point recombination rates became negligible (Olm *et al.*, 2019).

Why should divergence within any species taxon, from any walk of life, be constrained to the same extent by homologous recombination? We should first note that 95% ANI represents only the lower bound of within-taxon identity (Jain *et al.*, 2018). While pairs of *Escherichia coli* strains extend down to 95% ANI, pairs of strains within *Mycobacterium* species taxa are rarely less than 99% identical. We can allow that homogenization of genomes is occurring across different levels of relatedness in different groups.

One possibility is that homogenization of genomes with high ANI may be mostly passive, with recombination bringing functionally interchangeable sequence variants between the various ecotypes of a species taxon (Marttinen *et al.*, 2015; Iranzo *et al.*, 2019). Various forces can cause the rate of recombination to decrease with increasing divergence. It is challenging to see why all taxa should be subject to the same constraints on homologous recombination, such that homogenization of sequences would fade out by 95% ANI.

First, consider that there is a molecular requirement of sequence identity between donor and recipient at one or both ends of a recombining segment (Shen and Huang, 1986). This yields an exponential decay of recombination with increasing sequence divergence (Majewski and Cohan, 1999a,b; Majewski, 2001; Costechareyre *et al.*, 2009). Of all constraints on recombination, one can most easily imagine how this molecular constraint on recombination may act more or less uniformly across the conserved recombination apparatus in different bacterial phyla.

On the other hand, it is difficult to see why ecological constraints on recombination would increase uniformly with sequence divergence across all bacterial groups. Recombination between lineages requires that they live in the same microhabitat (Matte-Tailliez *et al.*, 2002), and it is not clear why the probability of inhabiting the same environment would decline uniformly with sequence divergence. Other constraints on homologous recombination include differences in restriction endonuclease systems and in the plasmids and phage that could carry host genes (Wiedenbeck and Cohan, 2011; Hanage, 2016). It is not clear why these aspects of sexual isolation would increase uniformly with sequence divergence, especially when sexual isolation does not always increase monotonically with evolutionary divergence (Stefanic *et al.*, 2019).

An alternative explanation for homogenization of genomes with greater than 95% ANI (within a species taxon) could be that, above a particular level of relatedness, generally adaptive genes can pass from one ecotype to another and could thereby provide their adaptation to all the ecotypes within a species taxon (Cohan, 2016a, 2019). Each such sweep of an adaptive gene would homogenize the entire taxon for that gene. However, it does not appear that this model could explain homogeneity at more than a few tens of thousands of base pairs (Arevalo *et al.*, 2019). Instead, a passive homogenization of neutral sequence variants by homologous recombination is somehow responsible for blocking divergence within species taxa, down to a level of around 95% ANI, over all groups of bacteria.

So we are faced with the convenient but not totally understood phenomenon that homologous recombination declines to near zero by 95% ANI in all bacterial species taxa. The resulting gap between 95% and 83% ANI, across all groups, can enable systematists to demarcate species taxa even when we know very little about a prospective species taxon. Demarcating bacterial species taxa by 95% ANI not only yields the familiar species of bacterial taxonomy. It also yields species taxa that have the species-like property of cohesion in limiting the genome-wide divergence among a set of ecologically differentiated ecotypes.

Conclusion

It turns out, unexpectedly, that bacterial species taxa share a species-like property with the species taxa of zoology and botany. While recombination within species taxa of all these groups fails to prevent diversification within species, recombination nevertheless appears to act universally as a force of cohesion within species taxa. That is, recurrent recombination within species limits neutral sequence divergence within species taxa of plants, animals, and bacteria; recombination also allows a sharing of generally adaptive genes across a species range. The 95% ANI criterion that demarcates the traditionally defined species taxa of bacteria fortuitously also yields groups of bacteria that are subject to the species-like property of cohesion, where recombination prevents neutral sequence divergence among ecotypes within a species. Use of the ANI criterion, then, not only provides an easily used algorithm for demarcating bacterial species; it also places bacterial demarcation on the same theory-based foundation as the species taxonomy of animals and plants.

References

Arevalo, P., VanInsberghe, D., Elsherbini, J., Gore, J. and Polz, M.F. (2019) A reverse ecology approach based on a biological definition of microbial populations. *Cell* 178, 820. https://doi.org/10.1016/j.cell.2019.06.033

Auch, A.F., von Jan, M., Klenk, H.P. and Goker, M. (2010) Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences* 2, 117–134. https://doi.org/10.4056/sigs.531120

Barberán, A., Caceres Velazquez, H., Jones, S. and Fierer, N. (2017) Hiding in plain sight: mining bacterial species records for phenotypic trait information. *mSphere* 2, e00237. https://doi.org/10.1128/mSphere.00237-17

Barraclough, T.G., Hughes, M., Ashford-Hodges, N. and Fujisawa, T. (2009) Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biology Letters* 5, 425–428. https://doi.org/10.1098/rsbl.2009.0091

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M.C., Bateson, M.M., Ward, D.M. and Heidelberg, J.F. (2007) Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME Journal* 1, 703–713. https://doi.org/10.1038/ismej.2007.46

Bobay, L.-M. and Ochman, H. (2017) Biological species are universal across Life's Domains. *Genome Biology and Evolution* 9, 491–501. https://doi.org/10.1093/gbe/evx026

Bobay, L.-M. and Ochman, H. (2018) Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evolutionary Biology* 18, 153–153. https://doi.org/10.1186/s12862-018-1272-4

Becraft, E.D., Wood, J.M., Rusch, D.B., Kühl, M., Jensen, S.I., Bryant, D.A., Roberts, D.W., Cohan, F.M. and Ward, D.M. (2015) The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. *Frontiers in Microbiology* 6, 590. https://doi.org/10.3389/fmicb.2015.00590

Bendall, M.L., Stevens, S.L., Chan, L.K., Malfatti, S., Schwientek, P., Tremblay, J., Schackwitz, W., Martin, J., Pati, A., Bushnell, B. *et al.* (2016) Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. *ISME Journal.* https://doi.org/10.1038/ismej.2015.241

Berger, S.A. (2005) GIDEON: a comprehensive Web-based resource for geographic medicine. *International Journal of Health Geographics* 4, 10. https://doi.org/10.1186/1476-072X-4-10

Blount, Z.D., Barrick, J.E., Davidson, C.J. and Lenski, R.E. (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature Advance Online Publication.* https://doi.org/10.1038/nature11514

Cadillo-Quiroz, H., Didelot, X., Held, N.L., Herrera, A., Darling, A., Reno, M.L., Krause, D.J. and Whitaker, R.J. (2012) Patterns of gene flow define species of thermophilic Archaea. *PLoS Biology* 10, e1001265. https://doi.org/10.1371/journal.pbio.1001265

Caro-Quintero, A. and Konstantinidis, K.T. (2012) Bacterial species may exist, metagenomics reveal. *Environmental Microbiology* 14, 347–355. https://doi.org/10.1111/j.1462-2920.2011.02668.x

Clausen, J., Keck, D.D. and Hiesey, W.M. (1947) Heredity of geographically and ecologically isolated races. *The American Naturalist* 81, 114–133. https://doi.org/10.1086/281507

Cohan, F.M. (1994) The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *The American Naturalist* 143, 965–986. https://doi.org/10.1086/285644

Cohan, F.M. (2002) What are bacterial species? *Annual Review of Microbiology* 56, 457–487. https://doi.org/10.1146/annurev.micro.56.012302.160634

Cohan, F.M. (2006) Toward a conceptual and operational union of bacterial systematics, ecology, and evolution. *Proceedings of the Royal Society of London Series B* 361, 1985–1996. https://doi.org/10.1098/rstb.2006.1918

Cohan, F.M. (2011a) Are species cohesive?--A view from bacteriology. In: Walk, S. and Feng, P. (eds) *Bacterial Population Genetics: A Tribute to Thomas S. Whittam*. American Society for Microbiology Press, Washington, DC, pp. 43–65. https://doi.org/10.1128/9781555817114.ch5

Cohan, F.M. (2011b) Koufax's perfect game-the tale of the data. In: *Los Angeles Times*. Tribune Newspapers, Los Angeles, California.

Cohan, F.M. (2013) Species. In: Maloy, S. and Hughes, K. (eds). *Brenner's Encyclopedia of Genetics*, 2nd edn. Elsevier, Amsterdam, pp. 506–511. https://doi.org/10.1016/B978-0-12-374984-0.01454-6

Cohan, F.M. (2016a) Bacterial speciation: genetic sweeps in bacterial species. *Current Biology* 26, R112–R115. https://doi.org/10.1016/j.cub.2015.10.022

Cohan, F.M. (2016b) Prokaryotic species concepts. In: Kliman, R.M. (ed.) *Encyclopedia of Evolutionary Biology*. Elsevier, Amsterdam. https://doi.org/10.1016/B978-0-12-800049-6.00230-4

Cohan, F.M. (2016c) Bacterial species concepts. In: Kliman, R.M. (ed.) *Encyclopedia of Evolutionary Biology, Volume 1*. Academic Press, Oxford, UK, pp. 119–129. https://doi.org/10.1016/B978-0-12-800049-6.00230-4

Cohan, F.M. (2017a) Transmission in the origins of bacterial diversity, from ecotypes to phyla. *Microbiology Spectrum* 5. https://doi.org/10.1128/microbiolspec.MTBP-0014-2016

Cohan, F.M. (2017b) Species. In: Osmanaj, B. and Escalante-Santos, L. (eds). *Reference Module in Life Sciences*. Elsevier, Oxford, UK. https://doi.org/10.1016/B978-0-12-809633-8.07184-3

Cohan, F.M. (2019) Systematics: The cohesive nature of bacterial species taxa. *Current Biology* 29, R169–R172. https://doi.org/10.1016/j.cub.2019.01.033

Cohan, F.M. and Kopac, S.M. (2011) Microbial genomics: *E. coli* relatives out of doors and out of body. *Current Biology* 21, R587–589. https://doi.org/10.1016/j.cub.2011.06.011

Cohan, F.M. and Kopac, S.M. (2017) A theory-based pragmatism for discovering and classifying newly divergent species of bacterial pathogens. In: Tibayrenc, M. (ed.) *Genetics and Evolution of Infectious Diseases*, 2nd edn. Elsevier Inc., Oxford, UK, pp. 25–49. https://doi.org/10.1016/B978-0-12-799942-5.00002-0

Cohan, F.M. and Perry, E.B. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Current Biology* 17, R373–R386. https://doi.org/10.1016/j.cub.2007.03.032

Collins, C. and Didelot, X. (2018) A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1005958

Connor, N., Sikorski, J., Rooney, A.P., Kopac, S., Koeppel, A.F., Burger, A., Cole, S.G., Perry, E.B., Krizanc, D., Field, N.C. *et al.* (2010) The ecology of speciation in *Bacillus*. *Applied and Environmental Microbiology* 76, 1349–1358. https://doi.org/10.1128/AEM.01988-09

Costechareyre, D., Bertolla, F. and Nesme, X. (2009) Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. *Molecular Biology and Evolution* 26, 167–176. https://doi.org/10.1093/molbev/msn236

Coyne, J.A. and Orr, H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Chase, A.B., Arevalo, P., Brodie, E.L., Polz, M.F., Karaoz, U. and Martiny, J.B.H. (2019) Maintenance of sympatric and allopatric populations in free-living terrestrial bacteria. *mBio* 10, e02361. https://doi.org/10.1128/mBio.02361-19

Denef, V.J., Kalnejais, L.H., Mueller, R.S., Wilmes, P., Baker, B.J., Thomas, B.C., VerBerkmoes, N.C., Hettich, R.L. and Banfield, J.F. (2010) Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* 107, 2383–2390. https://doi.org/10.1073/pnas.0907041107

De Ley, J. (1970) Reexamination of the association between melting point, buoyant density, and chemical base composition of deoxyribonucleic acid. *Journal of Bacteriology* 101, 738–754. https://doi.org/10.1128/JB.101.3.738-754.1970

de Queiroz, K. (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America* 102 Suppl 1, 6600–6607. https://doi.org/10.1073/pnas.0502030102

Diallo, K., MacLennan, J., Harrison, O.B., Msefula, C., Sow, S.O., Daugla, D.M., Johnson, E., Trotter, C., MacLennan, C.A., Parkhill, J. *et al.* (2019) Genomic characterization of novel *Neisseria* species. *Scientific reports* 9, 13742–13742. https://doi.org/10.1038/s41598-019-50203-2

Dobzhansky, T. (1951) *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.

Doolittle, W.F., and Zhaxybayeva, O. (2009) On the origin of prokaryotic species. *Genome Research* 19, 744–756. https://doi.org/10.1101/gr.086645.108

Dutilh, B.E., Backus, L., Edwards, R.A., Wels, M., Bayjanov, J.R. and van Hijum, S.A.F.T. (2013) Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Briefings in Functional Genomics* 12, 366–380. https://doi.org/10.1093/bfgp/elt008

Dykhuizen, D.E. and Green, L. (1991) Recombination in *Escherichia coli* and the definition of biological species. *Journal of Bacteriology* 173, 7257–7268. https://doi.org/10.1128/JB.173.22.7257-7268.1991

Garrity, G.M. (2016) A new genomics-driven taxonomy of bacteria and archaea: are we there yet? *Journal of Clinical Microbiology* 54 (8), 1956–1963. doi: 10.1128/JCM.00200-16

Garrity, G. (2019) International Code of Nomenclature of Prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* 69, S1–S111.

Garrity, G. and Lyons, C. (2011) *Systems and Methods For Resolving Ambiguity Between Names and Entities*. Volume US 7.925,444 B2. (Board of Trustees of Michigan State University).

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D.M. and Thomas, M.G. (2011) Evolution of lactase persistence: an example of human niche construction. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 366, 863–877. https://doi.org/10.1098/rstb.2010.0268

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L. *et al.* (2005) Opinion: Re-evaluating prokaryotic species. *Nature Reviews Microbiology* 3, 733–739. https://doi.org/10.1038/nrmicro1236

Hahne, H., Mader, U., Otto, A., Bonn, F., Steil, L., Bremer, E., Hecker, M. and Becher, D. (2010) A comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation. *Journal of Bacteriology* 192, 870–882. https://doi.org/10.1128/JB.01106-09

Haldane, J.B.S. (1932) *The Causes of Evolution*. Longmans, Green, and Co., London.

Hanage, W.P. (2016) Not so simple after all: Bacteria, their population genetics, and recombination. *Cold Spring Harbor Perspectives in Biology* 8, a018069. https://doi.org/10.1101/cshperspect.a018069

He, X., Jin, Y., Ye, M., Chen, N., Zhu, J., Wang, J., Jiang, L. and Wu, R. (2017) Bacterial genetic architecture of ecological interactions in co-culture by GWAS-taking *Escherichia coli* and *Staphylococcus aureus* as an example. *Frontiers in Microbiology* 8, 2332–2332. https://doi.org/10.3389/fmicb.2017.02332

Hey, J. (2001) *Genes, Categories, and Species: The Evolutionary and Cognitive Cause of the Species Problem*. Oxford University Press, Oxford, UK.

Ho, S.Y. and Shapiro, B. (2011) Skyline-plot methods for estimating demographic history from nucleotide sequences. *Molecular Ecology Resources* 11, 423–434. https://doi.org/10.1111/j.1755-0998.2011.02988.x

Holmberg, K. and Nord, C.E. (1984) 14 Application of numerical taxonomy to the classification and identification of microaerophilic *Actinomycetes*. In: Bergan, T. (ed.) *Methods in Microbiology*, Volume 16. Academic Press, pp. 341–360. https://doi.org/10.1016/S0580-9517(08)70399-4

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M. *et al*. (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512, 194–197. https://doi.org/10.1038/nature13408

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. https://doi.org/10.1126/science.1157890

Hutchinson, G.E. (1968) When are species necessary? In: Lewontin, R.C. (ed.) *Population Biology and Evolution*. Syracuse University Press, Syracuse, New York, pp. 177–186.

Iranzo, J., Wolf, Y.I., Koonin, E.V. and Sela, I. (2019) Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nature Communications* 10, 5376. https://doi.org/10.1038/s41467-019-13429-2

Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T. and Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications* 9, 5114. https://doi.org/10.1038/s41467-018-07641-9

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R. *et al.* (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. *Science* 344, 416–420. https://doi.org/10.1126/science.1248575

Koch, H., Germscheid, N., Freese, H.M., Noriega-Ortega, B., Lücking, D., Berger, M., Qiu, G., Marzinelli, E.M., Campbell, A.H., Steinberg, P.D. *et al.* (2020) Genomic, metabolic and phenotypic variability shapes ecological differentiation and intraspecies interactions of *Alteromonas macleodii*. *Scientific Reports* 10, 809. https://doi.org/10.1038/s41598-020-57526-5

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, W.A., Ward, D.M., Rooney, A.P., Brambilla, E., Connor, N., Ratcliff, R.M. *et al.* (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences* 105, 2504–2509. https://doi.org/10.1073/pnas.0712205105

Koeppel, A.F., Wertheim, J.O., Barone, L., Gentile, N., Krizanc, D. and Cohan, F.M. (2013) Speedy speciation in a bacterial microcosm: New species can arise as frequently as adaptations within a species. *ISME Journal* 7, 1080–1091. https://doi.org/10.1038/ismej.2013.3

Konstantinidis, K.T. and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 102, 2567–2572. https://doi.org/10.1073/pnas.0409727102

Konstantinidis, K.T., Ramette, A. and Tiedje, J.M. (2006) The bacterial species definition in the genomic era. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 361, 1929–1940. https://doi.org/10.1098/rstb.2006.1920

Konstantinidis, K.T., Rosselló-Móra, R. and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME Journal* 11, 2399–2406. https://doi.org/10.1038/ismej.2017.113

Kopac, S.M. and Cohan, F.M. (2012) Comment on 'Population genomics of early events in the ecological differentiation of bacteria'. *Science* 336.

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M. and Cohan, F.M. (2014) Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. *Applied and Environmental Microbiology* 80, 4842–4853. https://doi.org/10.1128/AEM.00576-14

Kumar, R., Verma, H., Haider, S., Bajaj, A., Sood, U., Ponnusamy, K., Nagar, S., Shakarad, M.N., Negi, R.K., Singh, Y. *et al.* (2017) Comparative genomic analysis reveals habitat-specific genes and regulatory hubs within the genus *novosphingobium*. *mSystems* 2, e00020-00017. https://doi.org/10.1128/mSystems.00020-17

Lassalle, F., Campillo, T., Vial, L., Baude, J., Costechareyre, D., Chapulliot, D., Shams, M., Abrouk, D., Lavire, C., Oger-Desfeux, C. *et al.* (2011) Genomic species Are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. *Genome Biology and Evolution* 3, 762–781. https://doi.org/10.1093/gbe/evr070

Locey, K.J. and Lennon, J.T. (2016) Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences* 113, 5970–5975. https://doi.org/10.1073/pnas.1521291113

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M. and Konstantinidis, K.T. (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences of the United States of America* 108, 7200–7205. https://doi.org/10.1073/pnas.1015622108

Majewski, J. (2001) Sexual isolation in bacteria. *FEMS Microbiology Letters* 199, 161–169. https://doi.org/10.1111/j.1574-6968.2001.tb10668.x

Majewski, J. and Cohan, F.M. (1999a) Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152, 1459–1474.

Majewski, J. and Cohan, F.M. (1999b) DNA sequence similarity requirements for interspecific recombination in Bacillus. *Genetics* 153, 1525–1533.

Mallet, J. (1995) A species definition for the modern synthesis. *Trends in Ecology and Evolution* 10, 294–299. https://doi.org/10.1016/0169-5347(95)90031-4

Mallet, J. (2008) Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363, 2971–2986. https://doi.org/10.1098/rstb.2008.0081

Martiny, A.C., Tai, A.P., Veneziano, D., Primeau, F. and Chisholm, S.W. (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environmental Microbiology* 11, 823–832. https://doi.org/10.1111/j.1462-2920.2008.01803.x

Marttinen, P., Croucher, N.J., Gutmann, M.U., Corander, J. and Hanage, W.P. (2015) Recombination produces coherent bacterial species clusters in both core and accessory genomes. *Microbial Genomics* 1, e000038. https://doi.org/10.1099/mgen.0.000038

Mateo-Estrada, V., Graña, L., López-Leal, G. and Castillo-Ramírez, S. (2019) Phylogenomics reveals clear cases of misclassification and genus-wide phylogenetic markers for *Acinetobacter*. *Genome Biology and Evolution.* https://doi.org/10.1093/gbe/evz178

Matte-Tailliez, O., Brochier, C., Forterre, P. and Philippe, H. (2002) Archaeal phylogeny based on ribosomal proteins. *Molecular Biology and Evolution* 19, 631–639. https://doi.org/10.1093/oxfordjournals.molbev.a004122

Mayr, E. (1942) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Columbia University, New York.

Mayr, E. (1963) *Animal Species and Evolution*. Belknap Press of Harvard University Press, Cambridge. https://doi.org/10.4159/harvard.9780674865327

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R. and Hugenholtz, P. (2012) An improved Greengenes taxonomy with explicit ranks for

ecological and evolutionary analyses of bacteria and archaea. *ISME Journal* 6, 610–618. https://doi.org/10.1038/ismej.2011.139

Melendrez, M.C., Becraft, E.D., Wood, J.M., Olsen, M.T., Bryant, D.A., Heidelberg, J.F., Rusch, D.B., Cohan, F.M. and Ward, D.M. (2016) Recombination does not hinder formation or detection of ecological species of *Synechococcus* inhabiting a hot spring cyanobacterial mat. *Frontiers in Microbiology* 6, 1540. https://doi.org/10.3389/fmicb.2015.01540

Mirete, S., Mora-Ruiz, M.R., Lamprecht-Grandío, M., de Figueras, C.G., Rosselló-Móra, R. and González-Pastor, J.E. (2015) Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. *Frontiers in Microbiology* 6, 1121. https://doi.org/10.3389/fmicb.2015.01121

Nelson, W.C., Maezato, Y., Wu, Y.-W., Romine, M.F. and Lindemann, S.R. (2016) Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. *Applied and Environmental Microbiology* 82, 255–267. https://doi.org/10.1128/AEM.02274-15

Olm, M.R., Crits-Christoph, A., Diamond, S., Lavy, A., Matheus Carnevali, P.B. and Banfield, J.F. (2019) Consistent metagenome-derived metrics verify and define bacterial species boundaries. *bioRxiv* 647511. https://doi.org/10.1101/647511

Palmer, M., Venter, S.N., Coetzee, M.P.A. and Steenkamp, E.T. (2019) Prokaryotic species are sui generis evolutionary units. *Systematic and Applied Microbiology* 42, 145–158. https://doi.org/10.1016/j.syapm.2018.10.002

Papke, R.T., Koenig, J.E., Rodriguez-Valera, F. and Doolittle, W.F. (2004) Frequent recombination in a saltern population of *Halorubrum*. *Science* 306, 1928–1929.

Parker, C.T., Tindall, B.J. and Garrity, G.M. (2019) International Code of Nomenclature of Prokaryotes. *International Journal of Systematic and Evolutionary Microbiology* 69, S1–S111. https://doi.org/10.1099/ijsem.0.000778

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A. and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* 36, 996. https://doi.org/10.1038/nbt.4229

Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J. and Hugenholtz, P. (2019) Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv* 771964. https://doi.org/10.1101/771964

Polz, M.F., Alm, E.J. and Hanage, W.P. (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics* 29, 170–175. https://doi.org/10.1016/j.tig.2012.12.006

Rainey, P.B. and Travisano, M. (1998) Adaptive radiation in a heterogeneous environment. *Nature* 394, 69–72. https://doi.org/10.1038/27900

Ramírez, M., Moncada, R.N., Villegas-Escobar, V., Jackson, R.W. and Ramírez, C.A. (2020) Phylogenetic and pathogenic variability of strains of *Ralstonia solanacearum* causing moko disease in Colombia. *Plant Pathology* 69, 360–369. https://doi.org/10.1111/ppa.13121

Roberts, M.S. and Cohan, F.M. (1995) Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution* 49, 1081–1094. https://doi.org/10.1111/j.1558-5646.1995.tb04435.x

Rosselló-Móra, R. and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiology Reviews* 25, 39–67. https://doi.org/10.1111/j.1574-6976.2001.tb00571.x

Shapiro, B.J., Friedman, J., Cordero, O.X., Preheim, S.P., Timberlake, S.C., Szabo, G., Polz, M.F. and Alm, E.J. (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science* 336, 48–51. https://doi.org/10.1126/science.1218198

Shapiro, B.J. and Polz, M.F. (2014) Ordering microbial diversity into ecologically and genetically cohesive units. *Trends in Microbiology* 22, 235–247. https://doi.org/10.1016/j.tim.2014.02.006

Shapiro, B.J. and Polz, M.F. (2015) Microbial speciation. *Cold Spring Harbor Perspectives in Biology* 7, a018143. https://doi.org/10.1101/cshperspect.a018143

Shen, P. and Huang, H.V. (1986) Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics* 112, 441–457.

Sikorski, J. (2008) Populations under microevolutionary scrutiny: what will we gain? *Archives of Microbiology* 189, 1–5. https://doi.org/10.1007/s00203-007-0294-x

Sneath, P. and Sokal, R. (1973) *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco, California.

Stackebrandt, E. and Ebers, J. (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiology Today* 33, 152–155.

Stackebrandt, E. and Goebel, B.M. (1994) Taxonomic note: a place for DNA:DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology* 44, 846–849. https://doi.org/10.1099/00207713-44-4-846

Stackebrandt, E., Fischer, A., Roggentin, T., Wehmeyer, U., Bomar, D. and Smida, J. (1988) A phylogenetic survey of budding, and/or prosthecate, non-phototrophic eubacteria: membership of *Hyphomicrobium*, *Hyphomonas*, *Pedomicrobium*, *Filomicrobium*, *Caulobacter* and '*Dichotomicrobium*' to the alpha-subdivision of purple non-sulfur bacteria. *Archives of Microbiology* 149, 547–556. https://doi.org/10.1007/BF00446759

Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 361, 1899–1909. https://doi.org/10.1098/rstb.2006.1914

Stefanic, P., Belcijan, K., Kraigher, B., Kostanjšek, R., Nesme, J., Madsen, J., Kovac, J., Sørensen, S., Vos, M. and Mandic-Mulec, I. (2019) Intra-species DNA exchange: *Bacillus subtilis* prefers sex with less related strains. *bioRxiv* 756569. https://doi.org/10.1101/756569

Templeton, A. (1989) The meaning of species and speciation: a genetic perspective. In: Otte, D. and Endler, J. (eds) *Speciation and its Consequences*, Sinauer Associates, Sunderland, Massachusetts, pp. 3–27.

Thompson, A.W. and Kouba, K. (2019) Differential activity of coexisting *Prochlorococcus* ecotypes. *Frontiers in Marine Science* 6. https://doi.org/10.3389/fmars.2019.00701

Thompson, C.C., Amaral, G.R., Campeao, M., Edwards, R.A., Polz, M.F., Dutilh, B.E., Ussery, D.W., Sawabe, T., Swings, J. and Thompson, F.L. (2014) Microbial taxonomy in the post-genomic era: Rebuilding from scratch? *Archives of Microbiology* 197, 359–370. https://doi.org/10.1007/s00203-014-1071-2

Tindall, B.J., Rosselló-Móra, R., Busse, H.J., Ludwig, W. and Kampfer, P. (2010) Notes on the characterization of prokaryote strains for taxonomic purposes. *International Journal of Systematic and Evolutionary Microbiology* 60, 249–266. https://doi.org/10.1099/ijs.0.016949-0

Treves, D.S., Manning, S. and Adams, J. (1998) Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* 15, 789–797. https://doi.org/10.1093/oxfordjournals.molbev.a025984

Turesson, G. (1922) The species and the variety as ecological units. *Hereditas* 3, 100–113. https://doi.org/10.1111/j.1601-5223.1922.tb02727.x

Vandamme, P. and Peeters, C. (2014) Time to revisit polyphasic taxonomy. *Antonie Van Leeuwenhoek* 106, 57–65. https://doi.org/10.1007/s10482-014-0148-x

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K. and Swings, J. (1996) Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological Reviews* 60, 407–438. https://doi.org/10.1128/MMBR.60.2.407-438.1996

Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C. and Pati, A. (2015) Microbial species delineation using whole genome sequences. *Nucleic Acids Research* 43, 6761–6771. https://doi.org/10.1093/nar/gkv657

Volz, E.M. (2012) Complex population dynamics and the coalescent under neutrality. *Genetics* 190, 187–201. https://doi.org/10.1534/genetics.111.134627

Vos, M. (2011) A species concept for bacteria based on adaptive divergence. *Trends in Microbiology* 19, 1–7. https://doi.org/10.1016/j.tim.2010.10.003

Vos, M. and Didelot, X. (2009) A comparison of homologous recombination rates in bacteria and archaea. *ISME Journal* 3, 199–208. https://doi.org/10.1038/ismej.2008.93

Ward, D.M. (1998) A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1, 271–277. https://doi.org/10.1016/S1369-5274(98)80029-5

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P. *et al.* (1987) Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology* 37, 463–464. https://doi.org/10.1099/00207713-37-4-463

Weimann, A., Mooren, K., Frank, J., Pope, P.B., Bremges, A. and McHardy, A.C. (2016) From genomes to phenotypes: Traitar, the microbial trait analyzer. *mSystems* 1, e00101-00116. https://doi.org/10.1128/mSystems.00101-16

Wiedenbeck, J. and Cohan, F.M. (2011) Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews* 35, 957–976. https://doi.org/10.1111/j.1574-6976.2011.00292.x

Wilkins, J.S. (2009) *Species: A History of the Idea*. University of California, Berkeley, California.

Wittouck, S., Wuyts, S., Meehan, C.J., van Noort, V. and Lebeer, S. (2019) A genome-based species taxonomy of the *Lactobacillus* genus complex. *mSystems* 4, e00264-00219. https://doi.org/10.1128/mSystems.00264-19

Woese, C.R. (1987) Bacterial evolution. *Microbiology Reviews* 51, 221–271. https://doi.org/10.1128/MMBR.51.2.221-271.1987

Wood, J.M., Becraft, E.M., Cohan, F., Krizanc, D. and Ward, D.M. (2020) Ecotype Simulation 2: An improved algorithm for efficiently demarcating microbial species from large sequence datasets. *bioRxiv* 2020.2002.2010.940734. https://doi.org/10.1101/2020.02.10.940734

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzeby, J., Amann, R. and Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* 12, 635–645. https://doi.org/10.1038/nrmicro3330

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G. *et al.* (2011) Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nature Biotechnology* 29, 415–420. https://doi.org/10.1038/nbt.1823

Yoon, C.K. (2009) *Naming Nature: The Clash between Instinct and Science*. Norton, New York.

Zawadzki, P. and Cohan, F.M. (1995) The size and continuity of DNA segments integrated in Bacillus transformation. *Genetics* 141, 1231–1243.

Zhou, Y., Zheng, J., Wu, Y., Zhang, W. and Jin, J. (2020) A completeness-independent method for preselection of closely related genomes for species delineation in prokaryotes. *BMC Genomics* 21, 183. https://doi.org/10.1186/s12864-020-6597-x

Zhu, Q., Mai, U., Pfeiffer, W., Janssen, S., Asnicar, F., Sanders, J.G., Belda-Ferre, P., Al-Ghalith, G.A., Kopylova, E., McDonald, D. *et al.* (2019) Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nature Communications* 10, 5477. https://doi.org/10.1038/s41467-019-13443-4