

# Microbial diversity and the genetic nature of microbial species

Mark Achtman\* and Michael Wagner†

**Abstract** | The earth contains a huge number of largely uncharacterized Bacteria and Archaea. Microbiologists are struggling to summarize their genetic diversity and classify them, which has resulted in heated debates on methods for defining species, mechanisms that lead to speciation and whether microbial species even exist. This Review proposes that decisions on the existence of species and methods to define them should be guided by a method-free species concept that is based on cohesive evolutionary forces. It summarizes current approaches to defining species and the problems of these approaches, and presents selected examples of the population genetic patterns at and below the species level.

## Homologous recombination

The acquisition of DNA segments from closely related microbes via recombination at flanking homologous sequence stretches.

## Lateral gene transfer

The acquisition of novel DNA segments from unrelated sources.

## Operational taxonomic unit

Usually defined as the number of distinct 16S ribosomal RNA sequences at a certain cut-off level of sequence diversity.

The classification system developed by Carl Linnaeus extended to animals, plants and rocks. Linnaeus did not classify microbes, used in this Review to refer collectively and exclusively to Bacteria and Archaea<sup>1</sup>, but since the mid-nineteenth century, binomial Linnaean names have been used by microbiologists to designate microbial species. The species level is where multiple disciplines intersect, including microbial systematics, ecology, population genetics, evolution and genomics. The explosion of data during the genomic era has been accompanied by debates that threaten the existence of any cohesive overview of the genetic nature of microbial species. Furthermore, the newly appreciated importance of the frequency of recombination, either in the form of homologous recombination<sup>2</sup> or lateral gene transfer (LGT)<sup>3</sup>, has challenged our concepts in each field and has raised many questions. It is difficult to integrate these diverse sources of information, because microbiologists lack a widely accepted theoretical species concept that is comparable to the biological species concept proposed by Ernst Mayr<sup>4</sup>. According to the biological species concept, the existence of many animal species and some plant species is ensured by the cohesive evolutionary forces that result from pre-zygotic and post-zygotic barriers between eukaryotic species<sup>5</sup>. Cohesive evolutionary forces are also necessary to generate microbial species, because in their absence the accumulation of genetic variation would probably result in a genetic continuum and microbial species would represent an artificial classification rather than a natural, circumscribed biological grouping<sup>3</sup>.

Microbial species are currently defined by a pragmatic, polyphasic approach that is based on clear rules for both genotypic and phenotypic properties<sup>6</sup> (BOX 1).

This pragmatic approach has served the community well, resulting in more than 7,031 accepted microbial species (G.M. Garrity, personal communication), and is being adapted to the genomic era. However, as currently practised, this approach faces serious problems, because a primary criterion for distinguishing species is a certain cut-off level for pairwise genomic DNA–DNA hybridization levels. This cut-off level is not based on any particular theoretical justification, but instead was chosen 20 years ago to match pre-existing species definitions. In addition, similar to the situation with eukaryotes<sup>7,8</sup>, most subsequent debates among microbiologists on the topic of microbial species have focused on methodologies for their definition<sup>9–13</sup>, rather than on explanations for their existence. Surveys of microbial diversity have equated species with operational taxonomic units (OTUs), based on 16S ribosomal RNA (rRNA) sequences<sup>14</sup>. However, 16S rRNA possesses insufficient genetic resolution for the reliable binning of microbes into species, and it might be preferable to use the average sequence diversity between all orthologous genes in pairs of genome sequences<sup>12,15</sup> (BOX 1). Recommending whole-genome comparisons for the definition of species may seem premature, considering that only approximately 600 completed microbial genomes are currently available and that the diversity of uncultured microbes in the environment is high. However, more than 1,400 bacterial genomes are currently being sequenced; genomic sequences have been obtained from uncultured bacteria by metagenomic studies<sup>16–19</sup> and after whole-genome amplification of single cells<sup>20,21</sup>; and the rapid development of re-sequencing technologies will soon increase the number of available genome sequences by several orders of magnitude<sup>22</sup>.

\*Environmental Research Institute, University College Cork, Ireland and Department of Molecular Biology, Max-Planck Institut für Infektionsbiologie, 10117 Berlin, Germany.  
†Department of Microbial Ecology, Faculty of Life Sciences, Vienna Ecology Centre, University of Vienna, A-1090 Vienna, Austria.  
Correspondence to M.A.  
e-mail: [m.achtman@ucc.ie](mailto:m.achtman@ucc.ie)  
doi: 10.1038/nrmicro1872  
Published online 7 May 2008

## Box 1 | Species definitions and concepts in microbiology

**Definitions**

Microbes are currently assigned to a common species if their reciprocal, pairwise DNA re-association values are  $\geq 70\%$  in DNA–DNA hybridization experiments under standardized conditions and their  $\Delta T_m$  (melting temperature) is  $\leq 5^\circ\text{C}$ <sup>79</sup>. In addition, all strains within a species must possess a certain degree of phenotypic consistency, and species descriptions should be based on more than one type strain<sup>11</sup>. A species name is only assigned if its members can be distinguished from other species by at least one diagnostic phenotypic trait<sup>79</sup>. Microbes with 16S ribosomal RNAs (rRNAs) that are  $\leq 98.7\%$  identical are always members of different species, because such strong differences in rRNA correlate with  $< 70\%$  DNA–DNA similarity<sup>80</sup>. However, the opposite is not necessarily true, and distinct species have been occasionally described with 16S rRNAs that are  $> 98.7\%$  identical. Most uncultured microbes cannot be assigned to a classical species because we do not know their phenotype. In some cases, uncultured microbes can be assigned a provisional ‘*Candidatus*’ designation if their 16S rRNA sequences are sufficiently different from those of recognized species, if experimental *in situ* hybridization can be used to specifically detect them and if a basic description of their morphology and biology has been provided<sup>81</sup>.

**Problems with DNA–DNA hybridization**

Pairwise comparisons of individual strains are not necessarily symmetrical, and different values can be obtained with the same pair of strains depending on which is used as probe and which as target. In addition, the results of this method can be intransitive: if DNA–DNA hybridization levels are  $> 70\%$  between strains A and B and  $> 70\%$  between strains B and C, they need not be  $> 70\%$  for strains A and C, which can lead to ambiguous or inconsistent assignments of strains to a common species. Finally, DNA–DNA hybridization is tedious, difficult to standardize between laboratories and cannot be used to build incremental databases.

**Alternative methods**

The use of newer genomic methods for defining microbial species has been encouraged recently<sup>11</sup>, as long as they are sufficiently consistent with DNA–DNA re-association. One such method, multilocus sequence analysis (MLSA) depends on phylogenetic analyses of multiple (typically 6–8) protein-coding core genes. Several species seem to form discontinuous sequence clusters, which suggests that MLSA clusters could provide a criterion for delimiting the extent of a microbial species<sup>31</sup>. Alternatively, attempts are underway to test whether species can be defined on the basis of the average nucleotide identity (ANI) of all orthologous genes in complete genome sequences of pairs of strains<sup>12</sup>. Data for 28 species indicates that 70% DNA–DNA similarity corresponds to 95% ANI, which could serve as a cut-off value to define a microbial species<sup>15</sup>.

**Concepts**

Various concepts have been suggested for microbial species, but none have been generally accepted<sup>9</sup>. The following quotes represent several published concepts that were chosen to illustrate the lack of consensus:

- “A species could be described as a monophyletic and genomically coherent cluster of individual organisms that show a high degree of overall similarity in many independent characteristics, and is diagnosable by a discriminative phenotypic property.” (REF. 9)
- “Species are considered to be an irreducible cluster of organisms diagnosably different from other such clusters and within which there is a parental pattern of ancestry and descent.” (REF. 82)
- “A species is a group of individuals where the observed lateral gene transfer within the group is much greater than the transfer between groups.” (REF. 83)
- “Microbes ... do not form natural clusters to which the term “species” can be universally and sensibly applied.” (REF. 84)
- “Species are (segments of) metapopulation lineages.” (REF. 7)

Alternatively, it might be possible to replace pairwise DNA–DNA hybridization by the identification of discrete sequence clusters based on multiple core genes<sup>2,10,23</sup>. But these are technical issues that do not address primary conceptual questions, namely, what is a microbial species and do they truly exist<sup>3,13</sup>? In this Review, we propose that microbiologists should adopt an abstract species concept that seems to apply to all biological organisms, we

discuss potential cohesive forces at the species level and we illustrate these concepts through selected examples of the genetic diversity within and between species.

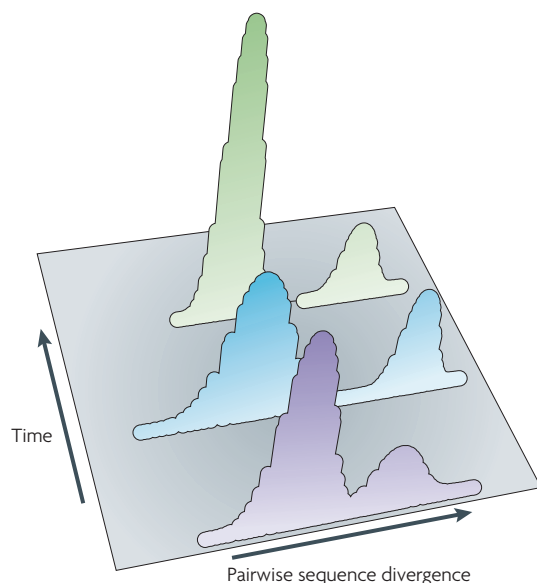
**A species concept for microbes**

Various concepts have been suggested for microbial species (BOX 1), but none have been generally accepted<sup>9</sup>, possibly because all of these concepts include methodological considerations. In contrast to concepts based on particular methods, the method-free unitary species concept of de Queiroz<sup>7</sup> (and similar considerations by Hey<sup>8</sup>) seems to apply well to microbes — that is, “species are metapopulation lineages”. Metapopulations are “sets of connected subpopulations” that are “maximally inclusive” and the limits of which are set by evolutionary cohesive forces. A lineage can be thought of as a metapopulation that extends through time (FIG. 1), “occupies an adaptive zone minimally different from that of any other lineage in its range” and “evolves separately from all lineages outside its range”. Unlike other species concepts, “metapopulation lineages do not have to be phenotypically distinguishable, or diagnosable, or monophyletic, or reproductively isolated, or ecologically divergent, to be species. They only have to be evolving separately from other such lineages.” Microbes that form distinct groups owing to a cohesive force are metapopulation lineages and thus form species, whereas microbes without limits imposed by a cohesive force do not.

Notably, the only criterion for a species according to this concept is their evolutionary fate, and no methodological criterion is required for assigning species designations. If this concept were accepted by microbiologists, it would still be necessary to agree on practical approaches and criteria by which species can be recognized, which might differ from current methodologies. Yet the following description of models of population structures in microbes, and observations on their microevolution, support our interpretation that such a species concept is necessary as a guide for the definition of species. Further experimental data and experience will be needed, however, to determine whether this concept is sufficient for this purpose.

**Ecotypes and periodic selection**

Various ecotype-based population-structure models have been proposed by Cohan<sup>24</sup> that could result in cohesive evolutionary forces in microbes (FIG. 2a–e). For asexual microbes, all related inhabitants of a unique ecological niche are thought to belong to a single, stable ecotype, within which overgrowth by fitter variants (periodic selection) repeatedly eliminates the genetic diversity that accumulates over time (FIG. 2a). Periodic selection places limits on the genetic diversity within an ecotype, but does not prevent the divergence of other, non-competing ecotypes in distinct niches, nor does it limit their diversity. Ecotypes are equated with true biological groupings<sup>24,25</sup>. By contrast, some described species are thought to be too broad, as they encompass multiple ecotypes. Under special conditions, however, single ecotypes can also encompass multiple clusters of genotypes. For example, discrete genotype clusters that



**Figure 1 | A unifying species concept that consists of a metapopulation lineage over time.** Theoretical histograms indicating the frequencies of binned levels of pairwise genetic divergence are shown. A metapopulation consists of one or more related populations at any one time; this schematic shows two populations. Samples taken at different time points, indicated in different colours, yield the same metapopulation, which indicates that it is a lineage, even if quantitative genetic divergence within and between each population varies owing to microevolution. Figure modified, with permission, from REF. 7 © (2005) National Academy of Sciences.

have diverged during geographical separation (allopatry) can compete owing to facilitated global transmission (FIG. 2b) and, when the population size is small, genetic diversity can persist owing to genetic drift, instead of being eliminated by periodic selection (FIG. 2c). Alternatively, homologous recombination between ecotypes can slow the elimination of genetic diversity (FIG. 2d) and LGT can facilitate the continual emergence of new ecotypes, with concomitant, continuous extinction of other competing ecotypes (FIG. 2e).

The ecotype model is highly attractive as it describes evolutionary and ecological principles that might account for microbial population structures and speciation. For example, different individual genetic clusters of *Bacillus simplex*, *Bacillus subtilis* and *Bacillus licheniformis* that were preferentially associated with shady and sunny slopes of ‘Evolution Canyon’ seem to represent numerous distinct ecotypes<sup>25,26</sup>. The ecotype model has also been invoked to explain genotype clusters in pathogenic microbes; for example, ecotypes have been equated with host specificity within the *Mycobacterium tuberculosis* complex<sup>27</sup>. But we think that some details of the ecotype concept lack support from biological observations or are even contradicted by them. For example, one prominent application of the ecotype concept was to provide ecological meaning to sequence clusters within marine genera (*Pelagibacter*, *Vibrio* and *Prochlorococcus*) that seemed to be associated with

distinct distributions of environmental parameters<sup>28</sup>. Various ecotypes of *Prochlorococcus* are thought to vary in abundance in seawater according to gradients of light, temperature and nutrients, possibly reflecting an association between individual ecotypes with particular genomic islands<sup>29</sup>. However, unlike Cohan’s ecotypes<sup>24</sup>, in which sequence diversity is limited, the nucleotide divergence between marine sequence clusters can be large, and can be greater than the limits that are normally associated with a microbial species by current definitions. Possibly, a designation other than ecotype might be more appropriate for these groupings within marine microbes; alternatively, further modifications of the ecotype concept might be needed to accommodate them. A further problem with the ecotype predictions are recent data, from a range of bacterial pathogens, that represent the most extensive analyses that are currently available on population structure within microbial species. As described below, these data raise questions about the details of Cohan’s ecotype models, and may also require changes in these models.

### Genetically monomorphic pathogens

Bacterial pathogens with low synonymous sequence diversity ( $D_s$  of <0.0002), so-called genetically monomorphic organisms, are promising models to reveal evolutionary mechanisms, and could even be more informative than microbes with greater sequence diversity, in which millions of years of evolutionary history might have blurred genomic signals of phylogenetic history through recombination<sup>30</sup> or eliminated them through genomic reduction<sup>31</sup>.

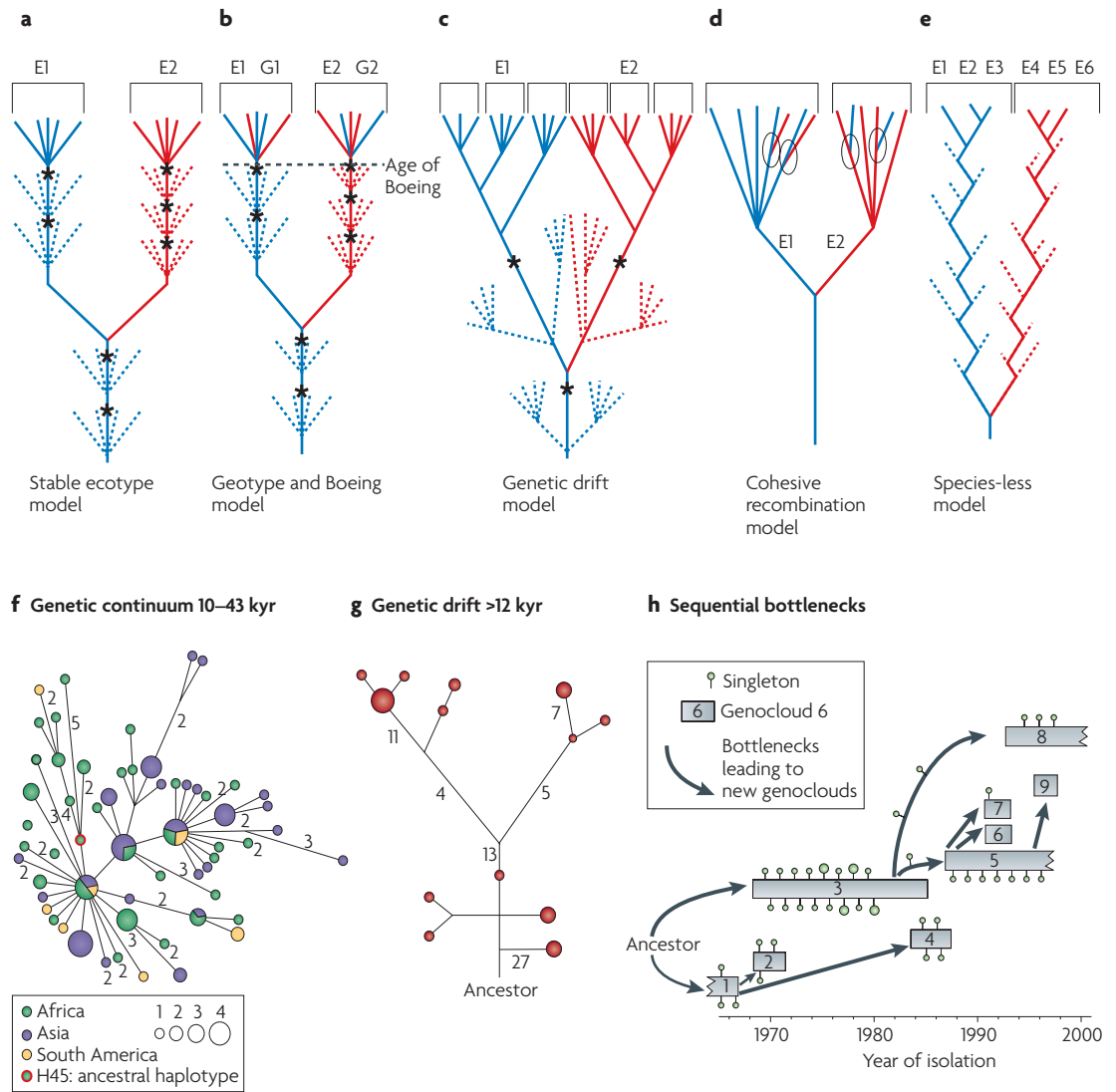
Based on their low levels of neutral sequence diversity, all *Salmonella enterica* subsp. *enterica* serovar Typhi (*S. Typhi*; the cause of typhoid fever) strains are thought to have descended from a common ancestor that infected humans 10,000–50,000 years ago<sup>32</sup>. Various genotypes have evolved over that time period. However, unlike the expectations for periodic selection, which should leave gaps between sequence clusters, many of these old genotypes apparently continue to exist today and form a continuum without gaps (FIG. 2f). In addition, although mutants resistant to nalidixic acid have been selected by extensive disease therapy with fluoroquinolones, none of the *S. Typhi* genotypes have become uniformly resistant to nalidixic acid, and local populations consist of a mixture of sensitive and resistant organisms<sup>32</sup>. This observation also argues against periodic selection, which should lead to complete replacement by fitter variants. Periodic selection may not occur in *S. Typhi*, as a few cases of typhoid fever result in a subsequent healthy carrier state and therefore *S. Typhi* should be protected from competition with bacteria from acute-phase disease. Alternatively, periodic selection might be so rare that all *S. Typhi* strains are derived from a single ecotype that has not undergone periodic selection over many millennia. *S. Typhi* illustrates two main problems with the ecotype concept: our lack of understanding of how rapidly sequence diversity is purged and the paucity of observations from nature that support complete purging.

#### Evolution Canyon

A mountainous region in Israel with highly contrasting physical and biotic environments on a micro-geographic scale.

#### $D_s$

Mean pairwise sequence diversity at synonymous sites.



**Figure 2 | Theoretical models of population structure versus population structures of genetically monomorphic bacterial pathogens. a–e** | Five types of ecotype models have been previously described in detail<sup>10</sup>. Basic characteristics of these models are indicated by their trivial names (for example, the Stable ecotype model and the Geotype and Boeing model). E1 and E2 represent ecotypes; G1 and G2 represent genotypes. Colours reflect genetic ancestry. Solid lines indicate extant lineages that exist today, whereas dotted lines indicate extinct lineages that have disappeared owing to overgrowth during episodes of periodic selection. **f** | Minimal spanning tree of genotypes within *Salmonella enterica* subsp. *enterica* serovar Typhi based on mutation discovery over 88 kb among 105 strains that were isolated in Africa, Asia and South America. The most striking aspect of these data is the apparent continuity of genetic diversity, which argues against periodic selection. The numbers along the edges that connect haplotypes are numbers of single nucleotide polymorphisms (SNPs); unlabelled edges represent single SNPs. An additional striking aspect is the current existence of ancestral haplotypes, including the ancestral node H45, which is circled in red. Finally, there is a lack of geographic specificity and many old haplotypes are global in their distribution, indicating multiple events of global spread. **g** | Conceptual phylogenetic tree based on genomic synonymous SNPs within 156 strains of *Yersinia pestis* that resembles the expectations of genetic drift (c), rather than periodic selection. The tree has a three-branch structure and contains eight populations, but the number of ecotypes, if any, is unknown. Numbers along branches represent numbers of synonymous SNPs (as in f); the branches are not drawn to scale. **h** | Patterns that resemble the species-less model (e) are caused by repeated extinction of genotype clusters (called ‘genoclouds’) among 502 isolates of *Neisseria meningitidis* serogroup A subgroup III during pandemic spread of cerebrospinal meningitis<sup>36</sup>. However, the rapid, apparently random, successions of genoclouds are attributed to geographic bottlenecks during epidemic spread that amplify the first strain to pass the bottleneck, even if it is a sequence variant. Genoclouds are represented by horizontal rectangles, the widths of which indicate the time periods during which they were isolated. Vertical lines terminated by dots indicate rare sequence variants of each genocloud. Sequence clusters 3, 5 and 8 were isolated globally, whereas others were restricted to China (1 and 4), Russia (2) or Africa (6, 7 and 9). Panels a–e reproduced, with permission, from *Nature Reviews Microbiology* REF. 10 © (2005) Macmillan Publishers Ltd. Panel f modified, with permission, from REF. 32 © (2006) American Association for the Advancement of Science. Panel g modified, with permission, from REF. 34 © (2004) National Academy of Sciences.

*Yersinia pestis*, the cause of plague, illustrates a second distinct population structure. *Y. pestis* is thought to represent a clone of *Yersinia pseudotuberculosis*, a gastrointestinal pathogen that is transmitted by the faecal–oral route. *Y. pestis* became capable of flea-borne transmission between rodents owing to the acquisition of two plasmids 12,000 years ago<sup>33,34</sup> (FIG. 2g). Strains of *Y. pestis* are almost indistinguishable from *Y. pseudotuberculosis*, and would not form a distinct species by the criteria that are normally used to differentiate species (BOX 1). Yet *Y. pestis* does deserve its designation as a distinct species, because it is almost certainly evolving independently of *Y. pseudotuberculosis*, with a totally distinct ecological niche and mechanism of transmission. Based on the few neutral (synonymous) sequence polymorphisms that differentiate genomes of modern strains<sup>34</sup>, the population genetic structure of *Y. pestis* (FIG. 2g) resembles the predictions associated with genetic drift (FIG. 2c). However, similar to *S. Typhi*, the limits of putative ecotypes remain unclear: do all *Y. pestis* strains correspond to a single ecotype or is each population a distinct ecotype? Thus, currently, the predictions of the ecotype models are too imprecise to reliably identify ecotypes among microbial populations.

We now turn to *Neisseria meningitidis* serogroup A subgroup III, the causal agent of epidemic cerebrospinal meningitis, the population structure of which seems to resemble that of the species-less ecotype model (FIG. 2e). Sequence diversity in these organisms accumulates rapidly over a period of several years, largely owing to homologous recombination with other neisserial species or lineages of *N. meningitidis*<sup>35,36</sup>. The most frequent genotype and its sequence variants are referred to as a ‘genocloud’. The diversity within genoclouds is only transient, owing to repeated purging during waves of pandemic spread (FIG. 2h). However, rather than periodic selection, these purification events were attributed to the bottlenecks that are associated with the migration of only a few organisms to new geographical areas<sup>36</sup>; migration is necessary for the continued existence of these bacteria because herd immunity leads to their elimination after several years in any single location. Genoclouds are also transient, being replaced by the progeny of particularly fit single-cell variants (founders) that yield new genoclouds. Bottlenecks and founder events could also explain the lack of genetic diversity within *Mycobacterium bovis* strains from Britain<sup>27</sup> and provide an alternative mechanism to periodic selection for the cohesive forces that generate populations.

### Genotype clusters at the species level

How well does the metapopulation species concept match patterns of genetic diversity in classical species? In the 1980s, microbiologists thought that species consisted of asexual clones, or clusters of related clones, that could be differentiated by neutral markers, such as electrophoretic allozymes<sup>37</sup>. Subsequent extensive comparisons of sequences from multiple housekeeping gene fragments (multilocus sequence typing; MLST) have yielded comparable results<sup>38</sup>. MLST data have been generated for ≥48 species (see [All Species MLST](#)

[databases](#) in Further information), primarily to analyse their epidemiological and environmental diversity. Within each species, individual strains are assigned to sequence types (STs), which are then grouped into genotype clusters that consist of chains of STs that differ from each other at only one or two loci<sup>39</sup>. The levels of genetic diversity in MLST genotype clusters are comparable to, or greater than, those of the individual populations described above for genetically monomorphic pathogens, whereas the genetic diversity of an entire species is typically much greater ( $D_s \leq 0.25$ ) than those of such populations. Currently, the largest MLST database, that of *Neisseria* spp., contains data from >8,000 isolates in 43 genotype clusters and >5,000 STs (FIG. 3f). Eight other MLST databases also contain large numbers of isolates, STs and genotype clusters, and the size of all MLST databases is growing rapidly. The neutral genetic diversity within these datasets indicates that many classical species contain multiple populations and may well correspond to metapopulation lineages.

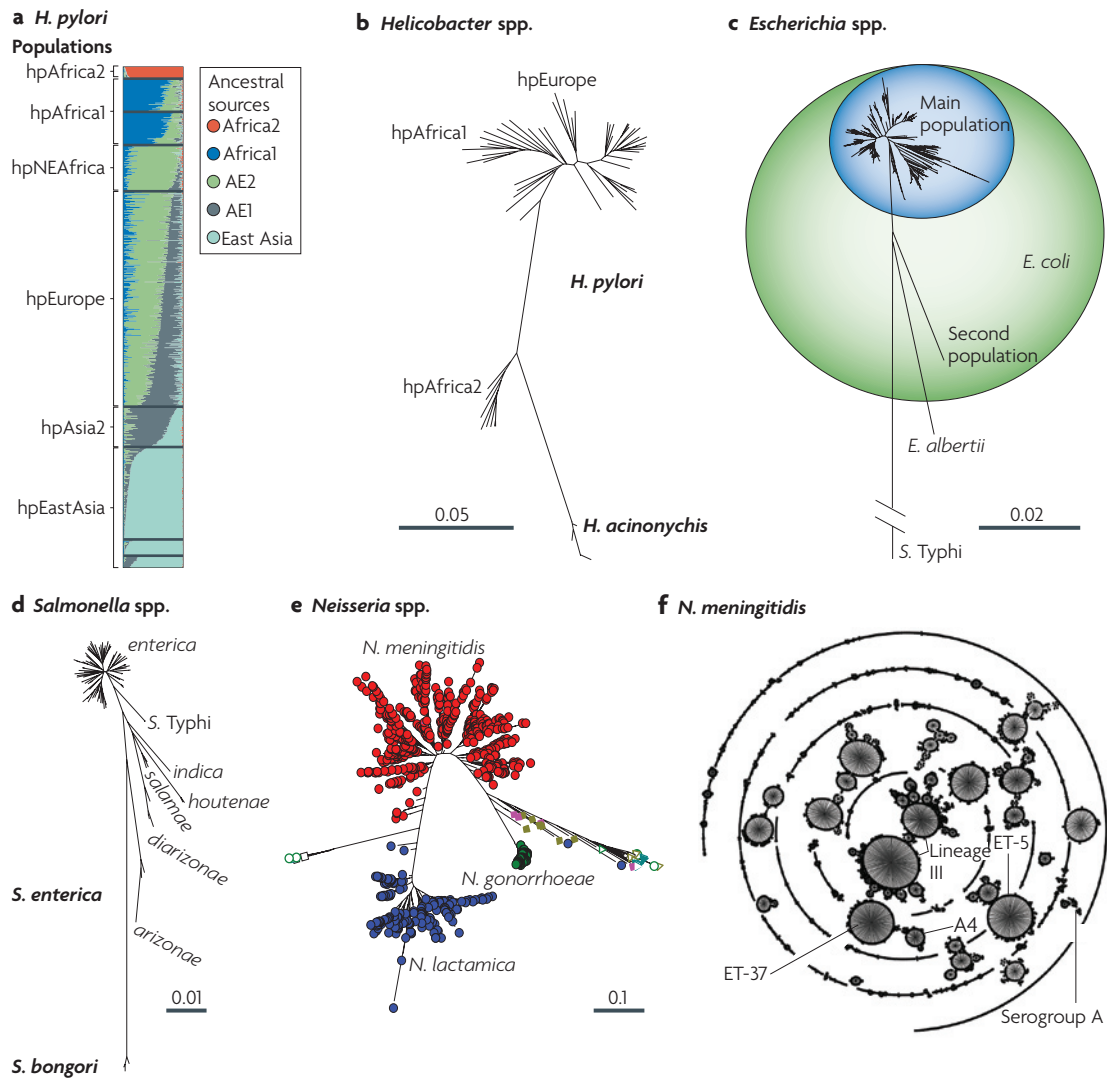
The predominance of asexual vertical inheritance within microbial species has been debated since 1993, when Maynard Smith *et al.*<sup>40</sup> posed the deceptively simple question “How clonal are bacteria?”, and then concluded that homologous recombination has influenced bacterial population genetic structures<sup>39</sup>. Traces of homologous recombination within a species have now been observed in a wide range of microbes, including marine bacterioplankton<sup>41</sup>, cyanobacterial mats<sup>42</sup>, intracellular symbionts<sup>31</sup> and Archaea<sup>16,23,43,44</sup>, but the frequency of recombinants varies between species. In one extreme example, recombination has been so frequent in *Helicobacter pylori* that almost each isolate defines a new ST<sup>45</sup> and 50% of the genome can be replaced by homologous recombination within 40–2,000 years<sup>46</sup>. *H. pylori* consists of a number of populations that accompanied ancient human migrations and then differentiated owing to isolation by distance<sup>45</sup>. But subsequent recombination has been so frequent between *H. pylori* from neighbouring geographical areas that these old genetic boundaries are now starting to blur (FIG. 3a). Signs of frequent recombination have even been detected within species that were previously thought to be largely clonal, such as *Escherichia coli* and *S. enterica*, resulting in star-like trees that contain little phylogenetic information<sup>30,47</sup> (FIG. 3c,d).

Distinct species are separated by large genetic distances that can act as barriers to homologous recombination, resulting in discontinuities in the genetic–distance distributions of multilocus sequences that delineate the boundaries of individual species<sup>2,47,48</sup>. Such major gaps can also result from other neutral processes, even in the absence of recombination, including the random extinction of individual clades<sup>2,49</sup>. Thus, the identification of genetic gaps within multilocus sequence data from related strains has been proposed to represent a promising approach for future biological species definitions<sup>10,28,49,50</sup>.

Delimited clusters of concatenated multilocus sequences do correlate well with classical species definitions for species of *Burkholderia* and *Streptococcus*<sup>50,51</sup>, and have been used to define species

**Bottleneck**  
Extreme reduction in population size.

**Sequence type**  
A unique allelic combination for all of the loci that have been sequenced.



**Figure 3 | Microbial species and sequence clusters.** **a** | Nearly continuous genetic clines have arisen via recombination between five ancestral populations in *Helicobacter pylori*. The multi-coloured rectangle contains 769 horizontal lines, one per isolate, each of which is composed of segments that are colour coded by ancestral source. The length of each segment is proportional to the proportion of ancestry from that source. Designations of modern populations are indicated at the left, whereas the colour codes for ancestral populations are at the right. **b** | NJ (neighbour joining) tree of concatenated sequences of *H. pylori*, which shows that the hpAfrica2 population is almost as distinct from the other populations of *H. pylori* as is the distinct species *Helicobacter acinonychis*. **c** | NJ tree of concatenated sequences of *Escherichia coli*, *Escherichia albertii* and *Salmonella enterica* subsp. *enterica* serovar Typhi. The green circle encloses the main population within *E. coli*, which contains 460 isolates (enclosed by a blue circle), plus a minor second population, which contains only 2 isolates. The second population recombines with the main population even though it is almost as distinct as *E. albertii*, a completely distinct species. **d** | NJ tree of concatenated sequences from *S. enterica* and *Salmonella bongori*. The italicized designations of clusters within *S. enterica* indicate distinct subspecies. The *S. Typhi* population appears to be almost as distinct as a subspecies owing to its having imported one allele by recombination from one of the subspecies. **e** | Bayesian tree of third-position sites within concatenated sequences of *Neisseria* spp. Three species, *Neisseria meningitidis* (red), *Neisseria lactamica* (blue) and *Neisseria gonorrhoeae* (green), are well resolved, but various others (other colours) are not. This poor resolution may reflect problems with taxonomic assignments or an inability to separate distinct species purely by genotype clusters. **f** | Illustration of the complexity of genetic diversity within a microbial species by eBurst analysis of *N. meningitidis* sequence types (STs). Each ST is represented by a single dot. Dots are arbitrarily arranged concentrically in seemingly continuous circles unless they are linked to other STs in a genotype cluster (fractal-like elements). Genotype clusters that were already known in 1998 (REF. 85) are labelled with their prior designations. Panel **a** modified, with permission, from Nature REF. 45 © (2007) Macmillan Publishers Ltd. Panel **b** modified from REF. 52. Panel **c** modified, with permission, from REF. 30 © (2006) Blackwell Publishing. Panel **d** modified, with permission, from REF. 47 © (2006) Royal Society Publishing. Panel **e** modified, with permission, from REF. 50 © (2005) BioMed Central. Data in panel **f** courtesy of B.G. Spratt and E.J. Feil, based on 5,811 STs containing 7,973 isolates from the PubMLST database (see Further information).

of *Ferroplasma*<sup>48</sup>. However, for other ‘fuzzy’ species, not all isolates fall into a sharply circumscribed genotype cluster because of occasional recombination between the species clusters<sup>23,48,50</sup>. For example, *N. meningitidis*, *Neisseria gonorrhoeae* and *Neisseria lactamica* form distinct genotype clusters, but strains assigned to various other *Neisseria* spp. simply form an ill-defined group of twigs at the tips of the tree (FIG. 3e). Few species have been tested rigorously by these criteria and additional analyses and extensive data from global sources are needed. However, it is already clear that current species definitions do not always correspond to single, circumscribed genotype clusters. Most *H. pylori* strains fall into one genotype cluster, but a second distinct genotype cluster (hpAfrica2) has been found in South Africa<sup>45</sup>. This second cluster is genetically almost as distinct from the primary cluster as is *Helicobacter acinonychis*, a separate species that arose by a host jump to large felines 200,000 years ago<sup>52</sup> (FIG. 3b). Similarly, two distinct genetic populations exist within *E. coli*<sup>30</sup>, one of which is almost as distant from the main population as is *Escherichia albertii* (FIG. 3c). Another problem with this approach to species definition is that multilocus sequence data based on seven loci might not have sufficient resolution for reliably recognizing genotype clusters. For example, *S. Typhi* seems to be distinct from other serovars of *S. enterica* (FIG. 3d), but this reflects the import of one allele from a distinct *S. enterica* subspecies<sup>47</sup> rather than speciation. The opposite situation is observed with sequence clusters that are based on rRNA sequences from environmental sources of *Francisella* spp. — one sequence cluster extends over various classical species rather than being limited to a single species<sup>53</sup>. Finally, genetic differentiation between species is a gradual process by which genes that are involved in speciation, and proximate genes that are linked to them, differentiate more quickly than others that are less important for adaptation to different niches<sup>54</sup>.

If genotype clusters were to become a primary criterion for recognizing species boundaries, species containing several, distinct genotype clusters would need to be subdivided into multiple species that each correspond to a single genotype cluster. However, we note that metapopulation lineages within the theoretical species concept do not preclude multiple genotype clusters, but simply dictate that multiple discrete populations within a metapopulation lineage do not evolve separately. Consistent with the metapopulation concept, homologous recombination continues to occur between genotype clusters in both *H. pylori*<sup>55</sup> and *E. coli*<sup>30</sup>. Discrete populations (genotype clusters) can also represent stages in speciation, such that the criteria needed for their designation as a separate species has not yet been achieved, but might be in the future<sup>54</sup>.

### LGT, gene loss and adaptation

Possibly the greatest challenge to the species concept<sup>3</sup> has been driven by the recent recognition that LGT has occurred on innumerable occasions and potentially represents a major disruptive force that could invalidate the

existence of microbial species. LGT has been recognized as a major cause of reduced susceptibility to antibiotics since the 1960s. For example, 28–67 kb long SSCmec genetic elements that encode resistance to methicillin and other antibiotics have been imported into *Staphylococcus aureus* on at least 20 independent occasions<sup>56</sup>. But many genes other than those involved in antibiotic resistance have also been imported by LGT, including selfish DNA on plasmids and prophages, resulting in unexpectedly high levels of variation in gene content within individual species. Genomic comparisons between strains of *E. coli* revealed that 8–21% of the genes within each genome were specific to single strains and many strain-specific genes were associated in genomic islands, each containing several genes<sup>54,57,58</sup>. Genomic islands are often associated with LGT and can introduce complete metabolic pathways and environmental adaptations through a single quantum event<sup>59,60</sup>. Genomic islands are probably widespread in microbes, and have also been identified by metagenomic studies of sea water<sup>29</sup>, marine sediments<sup>61</sup> and acid-mine-drainage biofilms<sup>44</sup>.

The diversity in content between related microbial genomes can be amazingly high. Extrapolations from 8 genomes of *Streptococcus agalactiae*<sup>62</sup> indicated that each additional genome would add at least 30 novel genes to the pan-genome. These extrapolations also indicated that the gene content of the pan-genome would still not be completely known after hundreds of genomes had been sequenced.

The differential existence of variable genes or genomic islands in some strains is compatible with a concept of rapid environmental adaptation owing to LGT. However, with the notable exception of antibiotic resistance, little evidence exists for recent, real-time evolution that is due to frequent and ongoing LGT. Strains of *E. coli* in which the genomes have been compared<sup>57,58</sup> differentiated millions of years ago<sup>30</sup>, and attempts to enumerate how often and when genes were introduced into an individual species indicate that most imported genes were probably introduced only rarely and before written history<sup>63,64</sup>. For example, one ‘recently’ introduced pathogenicity island (PAI), the *cagPAI* of *H. pylori*, seems to have been acquired only once, early in the history of this species, and most of the variably present genes in *H. pylori* seem to reflect differential gene loss<sup>65</sup>. Similarly, the well-known *Salmonella* pathogenicity islands 1 and 2 were apparently introduced into *Salmonella* spp. millions of years ago, long before sub-differentiation into modern subspecies and serovars<sup>54,66,67</sup>.

Although selective pressures can amplify rare genetic variants over extremely short time periods, and much more rapidly than neutral processes<sup>32,68</sup>, most sequence changes are eventually lost, especially if they reduce fitness. Examples of rapid disappearance owing to a lessening of fitness include: isolates of *Mycobacterium tuberculosis* that are resistant to rifampin<sup>69</sup>; variants of *E. coli* with increased adhesion to the bladder epithelium owing to changes in FimH adhesion<sup>70</sup>; non-synonymous polymorphisms that arose in *Pseudomonas aeruginosa* during chronic infection of a human lung (cystic fibrosis)<sup>71</sup>; and antigenic variants in *N. meningitidis*

#### Pan-genome

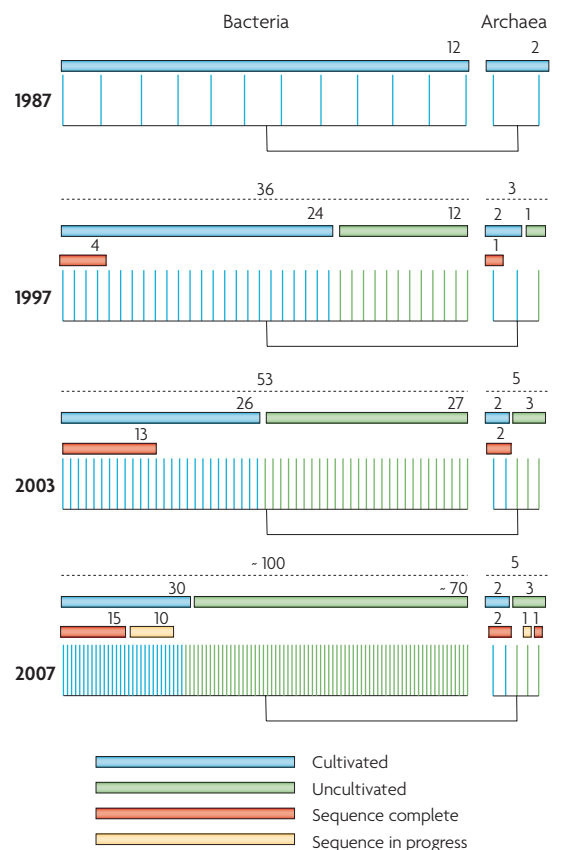
The combination of all coding sequences within the various strains of a species.

that were selected by herd immunity<sup>36</sup>. Loss of function during adaptation to a new host has also been observed after a host jump of *Helicobacter* spp. that occurred 200,000 years ago<sup>52</sup>, as well as in other pathogens, mutualists and symbionts<sup>72</sup>. Thus, rather than reflecting adaptive increases in fitness, a considerable proportion of the genetic diversity between individual strains within a species, including genes introduced by LGT, is transient and lost over time<sup>63</sup>. Given that circumscribed sequence clusters have already been observed in numerous species, we conclude that the existence of occasional LGT is not sufficient grounds to cast doubt on the concept of microbial species.

**Environmental microbes**

The examples of microbial population structures described above are compatible with the metapopulation species concept, but almost all are from bacteria that infect animal hosts. Within vertebrate-associated microbes, including both pathogens and commensals, branches of genetic diversity increase exponentially in frequency near the tips of dendrograms, at the level of species or strains<sup>73</sup>. However, the frequency of appearance of new branches may be much more continuous within environmental isolates<sup>73</sup>, whether from terrestrial or aquatic environments, which might explain why doubt has been cast on the existence of microbial species among environmental isolates<sup>3</sup>. This problem is particularly acute because millions of microbial species are thought to exist that have not yet been cultivated<sup>74,75</sup>, and we are only beginning to appreciate just how much microbial genetic diversity exists in nature. More than 85 novel bacterial phyla have been discovered since 1987 through molecular tools of microbial ecology (FIG. 4). For most of these phyla, we have not yet cultured even one representative isolate nor recovered a single genome sequence (FIG. 4). Furthermore, nothing is yet known about the population structure, frequency of recombination and cohesive forces for the vast majority of microbes on the Earth.

Many descriptions of genetic diversity in environmental samples have been published, but most of these have attempted to estimate the numbers of distinct OTUs rather than examining the diversity within individual clusters of related sequences. However, recent metagenomic approaches are beginning to address this issue. For example, there has been extensive sequencing of microbial DNA from the ocean surface at various geographical locations<sup>19</sup>. Similarly, genomic sequence diversity has been compared between *Prochlorococcus* spp. from various depths of the ocean and different geographical sources<sup>29,76</sup>, and sequence diversity within one gene has been determined over a 1-year period among *Vibrio splendidus* strains from coastal bacterioplankton<sup>77</sup>. A crucial test of the metapopulation species concept would be facilitated by understanding the global population structure of multiple environmental microbial taxa that might equate to species. Unfortunately, such data are still rare<sup>18,43,78</sup>.



**Figure 4 | Numbers of phyla and genomic sequences among Bacteria and Archaea since 1987.** Each schematic tree shows the numbers of known phyla — defined as groups of organisms that share <85% sequence identity in their 16S ribosomal RNA with other groups<sup>86</sup> — that had been recognized by each indicated year. Each phylum is represented by a vertical line, the colour of which indicates whether any of its members have been cultivated (blue) or not (green). Coloured horizontal bars span phyla for which at least one genomic sequence is complete (red) or in progress (yellow). Numbers above each line indicate the numbers of phyla that they span. Genome data for 2007 were extracted from the [GOLD Genomes OnLine Database](#)<sup>87</sup> (see Further information) in March 2007. Figure modified, with permission, from REF. 88 © (2005) American Society for Microbiology.

**Implications of a method-free species concept**

What would be the consequences of accepting the abstract concept of microbial species that we describe? First, the concept does not specify a ‘magic’ level of diversity. *N. meningitidis* is a metapopulation that consists of numerous, semi-discrete lineages that are linked by recombination<sup>50</sup>. *E. coli* is a metapopulation that contains two populations, one of which results from the admixture of four ancestral populations<sup>30</sup>. Second, this concept provides a rationale for maintaining the species designations for distinct ecotype-like groupings, such as *Y. pestis*, which is largely indistinguishable from *Y. pseudotuberculosis* at the sequence level<sup>34</sup>, but that represents a distinct lineage that is unlikely to merge with its parent owing to their



distinct ecological niches. If *Prochlorococcus* ecotypes were linked by cohesive forces, such as recombination, this concept would also support grouping them into one species, despite their extremely pronounced sequence differences<sup>29</sup>. However, the metapopulation species concept does have a potentially negative aspect, namely, it would place the onus on traditional systematics to justify why 70% DNA–DNA hybridization or 95% average nucleotide identity should represent a magic and invariable definition of microbial species. Optimistically, analyses addressed to answer this onus might reveal biological mechanisms that can explain these magic levels. Pessimistically, some species designations might need to be reconsidered if they did not reflect evolutionary distinct metapopulation lineages, resulting in splitting of some currently accepted taxa and merging of others. The taxonomic issues of designating microbes that do not form a species would also need to be addressed.

The acceptance of a method-free species concept would free microbiologists to choose from a range of techniques those that are most suitable for the target organism, as the concept specifies neither a specific methodology nor a specific cut-off, but rather invokes a general evolutionary criterion for speciation. Such freedom would potentially allow the definition of species among some uncultivated microbes from the

environment, while possibly demonstrating that others do not form species. A method-free concept would certainly stimulate methodological discussions that would need to be linked to evolutionary theory, a development that we fully endorse. However, changing to a theory-based approach to species definitions will not happen overnight. It will take considerable time for microbiologists to reach a consensus on the population genetic patterns that are reliable markers of the boundaries of a coherent microbial species. Time will also be needed for the examination of potential taxa, particularly among environmental samples. During this interim period, current methodologies will remain the only available tools for the definition of species and should only be replaced gradually once it is clear which methodologies correlate most strongly with the metapopulation concept.

Our credo is that many microbes fall into natural species with cohesive properties, but that one size cannot fit all. Species are biological units, some simple and others more complicated. The extent of these biological units is determined by the environment and their genomic potential. Concerted interdisciplinary in-depth comparisons of the nature of various microbial species are needed to provide the data that is essential for microbiologists to focus on unifying principles and understand exceptions.

- Pace, N. R. Time for a change. *Nature* **441**, 289 (2006).
- Discourages the use of the term prokaryotes to cover Bacteria and Archaea.
- Fraser, C., Hanage, W. P. & Spratt, B. G. Recombination and the nature of bacterial speciation. *Science* **315**, 476–480 (2007).
- Doolittle, W. F. & Papke, R. T. Genomics and the bacterial species problem. *Genome Biol.* **7**, 116 (2006).
- Questions whether microbial species represent a true biological grouping.
- Mayr, E. *Systematics and the Origin of Species — From the Viewpoint of a Zoologist* (Harvard Univ. Press, Cambridge, 1942).
- Coyne, J. A. & Orr, H. A. *Speciation* (Sinauer Associates, Sunderland, 2004).
- Vandamme, P. *et al.* Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**, 407–438 (1996).
- de Queiroz, K. Ernst Mayr and the modern concept of species. *Proc. Natl Acad. Sci. USA* **102** (Suppl. 1), 6600–6607 (2005).
- The source of the metapopulation lineage concept used in this Review.
- Hey, J. On the failure of modern species concepts. *Trends Ecol. Evol.* **21**, 447–450 (2006).
- Rossello-Mora, R. & Amann, R. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25**, 39–67 (2001).
- Gevers, D. *et al.* Opinion: re-evaluating prokaryotic species. *Nature Rev. Microbiol.* **3**, 735–739 (2005).
- Stackebrandt, E. *et al.* Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **52**, 1043–1047 (2002).
- Konstantinidis, K. T., Ramette, A. & Tiedje, J. M. The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B* **361**, 1929–1940 (2006).
- Buckley, M. & Roberts, R. J. *Critical Issues Colloquia Reports. Reconciling Microbial Systematics and Genomics* (American Academy of Microbiology, Washington DC, 2007).
- Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1025 (2006).
- Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
- Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
- Strous, M. *et al.* Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature* **440**, 790–794 (2006).
- Hallam, S. J. *et al.* Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl Acad. Sci. USA* **103**, 18296–18301 (2006).
- Halpern, A. L. *et al.* The Sorcerer II global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
- Marcy, Y. *et al.* Dissecting biological “dark matter” with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc. Natl Acad. Sci. USA* **104**, 11889–11894 (2007).
- Mussmann, M. *et al.* Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol.* **5**, e230 (2007).
- Bentley, D. R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545–552 (2006).
- Papke, R. T. *et al.* Searching for species in haloarchaea. *Proc. Natl Acad. Sci. USA* **104**, 14092–14097 (2007).
- Cohan, F. M. & Perry, E. B. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* **17**, R373–R386 (2007).
- Koeppel, A. *et al.* Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl Acad. Sci. USA* **105**, 2504–2509 (2008).
- Sikorski, J. & Nevo, E. Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at “Evolution Canyons” I and II, Israel. *Proc. Natl Acad. Sci. USA* **102**, 15924–15929 (2005).
- Smith, N. H., Gordon, S. V., Rua-Domenech, R., Clifton-Hadley, R. S. & Hewinson, R. G. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nature Rev. Microbiol.* **4**, 670–681 (2006).
- Poiz, M. F., Hunt, D. E., Preheim, S. P. & Weinreich, D. M. Patterns and mechanisms of genetic and phenotypic differentiation in marine microbes. *Philos. Trans. R. Soc. Lond. B* **361**, 2009–2021 (2006).
- Coleman, M. L. *et al.* Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**, 1768–1770 (2006).
- Wirth, T. *et al.* Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol. Microbiol.* **60**, 1136–1151 (2006).
- Dale, C. & Moran, N. A. Molecular interactions between bacterial symbionts and their hosts. *Cell* **126**, 453–465 (2006).
- Roumagnac, P. *et al.* Evolutionary history of *Salmonella* Typhi. *Science* **314**, 1301–1304 (2006).
- Achtman, M. *et al.* *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc. Natl Acad. Sci. USA* **96**, 14043–14048 (1999).
- Achtman, M. *et al.* Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc. Natl Acad. Sci. USA* **101**, 17837–17842 (2004).
- Linz, B., Schenker, M., Zhu, P. & Achtman, M. Frequent interspecific genetic exchange between commensal *Neisseria* and *Neisseria meningitidis*. *Mol. Microbiol.* **36**, 1049–1058 (2000).
- Zhu, P. *et al.* Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. *Proc. Natl Acad. Sci. USA* **98**, 5234–5239 (2001).
- Selander, R. K. *et al.* Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl. Environ. Microbiol.* **51**, 873–884 (1986).
- Maiden, M. C. Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.* **60**, 561–588 (2006).
- Feil, E. J. Small change: keeping pace with microevolution. *Nature Rev. Microbiol.* **2**, 483–495 (2004).
- Maynard Smith, J., Smith, N. H., O’Rourke, M. & Spratt, B. G. How clonal are bacteria? *Proc. Natl Acad. Sci. USA* **90**, 4384–4388 (1993).
- A revolutionary publication that altered our thought processes on the importance of homologous recombination.
- Vergin, K. L. *et al.* High intraspecific recombination rate in a native population of *Candidatus Pelagibacter ubique* (SAR11). *Environ. Microbiol.* **9**, 2430–2440 (2007).

42. Lodders, N., Stackebrandt, E. & Nübel, U. Frequent genetic recombination in natural populations of the marine cyanobacterium *Microcoleus chthonoplastes*. *Environ. Microbiol.* **7**, 434–442 (2005).
43. Whitaker, R. J. Allopatric origins of microbial species. *Philos. Trans. R. Soc. Lond. B* **361**, 1975–1984 (2006).
44. Allen, E. E. *et al.* Genome dynamics in a natural archaeal population. *Proc. Natl Acad. Sci. USA* **104**, 1883–1888 (2007).
45. Linz, B. *et al.* An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**, 915–918 (2007).
- Bacterial biogeography as a mirror of ancient human migrations.**
46. Falush, D. *et al.* Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size and minimal age. *Proc. Natl Acad. Sci. USA* **98**, 15056–15061 (2001).
47. Falush, D. *et al.* Mismatch induced speciation in *Salmonella*: model and data. *Philos. Trans. R. Soc. Lond. B* **361**, 2045–2053 (2006).
48. Eppley, J. M., Tyson, G. W., Getz, W. M. & Banfield, J. F. Genetic exchange across a species boundary in the archaeal genus *Ferroplasma*. *Genetics* **177**, 407–416 (2007).
49. Cohan, F. M. What are bacterial species? *Annu. Rev. Microbiol.* **56**, 457–487 (2002).
50. Hanage, W. P., Fraser, C. & Spratt, B. G. Fuzzy species among recombinogenic bacteria. *BMC Biol.* **3**, 6 (2005).
51. Hanage, W. P., Fraser, C. & Spratt, B. G. Sequences, sequence clusters and bacterial species. *Philos. Trans. R. Soc. Lond. B* **361**, 1917–1927 (2006).
- One of a number of articles from a special symposium that are cited here and that illustrate various aspects of diversity within microbial species. The whole symposium is highly recommended as background material.**
52. Eppinger, M. *et al.* Who ate whom? Adaptive *Helicobacter* genomic changes that accompanied a host jump from early humans to large felines. *PLoS Genet.* **2**, e120 (2006).
53. Keim, P., Johansson, A. & Wagner, D. M. Molecular epidemiology, evolution, and ecology of *Francisella*. *Ann. NY Acad. Sci.* **1105**, 30–66 (2007).
54. Retchless, A. C. & Lawrence, J. G. Temporal fragmentation of speciation in bacteria. *Science* **317**, 1093–1096 (2007).
55. Falush, D. *et al.* Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582–1585 (2003).
56. Grundmann, H., Aires-de-Sousa, M., Boyce, J. & Tiemersma, E. Emergence and resurgence of methicillin-resistant *Staphylococcus aureus* as a public-health threat. *Lancet* **368**, 874–885 (2006).
57. Welch, R. A. *et al.* Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **99**, 17020–17024 (2002).
58. Lloyd, A. L., Rasko, D. A. & Mobley, H. L. Defining genomic islands and uropathogen-specific genes in uropathogenic *Escherichia coli*. *J. Bacteriol.* **189**, 3532–3546 (2007).
59. Hacker, J. & Carniel, E. Ecological fitness, genomic islands and bacterial pathogenicity. A Darwinian view of the evolution of microbes. *EMBO Rep.* **2**, 376–381 (2001).
60. Gal-Mor, O. & Finlay, B. B. Pathogenicity islands: a molecular toolbox for bacterial virulence. *Cell. Microbiol.* **8**, 1707–1719 (2006).
61. Mussmann, M. *et al.* Clustered genes related to sulfate respiration in uncultured prokaryotes support the theory of their concomitant horizontal transfer. *J. Bacteriol.* **187**, 7126–7137 (2005).
62. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl Acad. Sci. USA* **102**, 13950–13955 (2005).
63. Lerat, E., Daubin, V., Ochman, H. & Moran, N. A. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* **3**, e130 (2005).
64. Dagan, T. & Martin, W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Natl Acad. Sci. USA* **104**, 870–875 (2007).
65. Gressmann, H. *et al.* Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet.* **1**, e43 (2005).
66. Ochman, H. & Groisman, E. A. Distribution of pathogenicity islands in *Salmonella* spp. *Infect. Immun.* **64**, 5410–5412 (1996).
67. Vernikos, G. S. & Parkhill, J. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* **22**, 2196–2203 (2006).
68. Shih, A. C., Hsiao, T. C., Ho, M. S. & Li, W. H. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc. Natl Acad. Sci. USA* **104**, 6283–6288 (2007).
69. Gagneux, S. *et al.* The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. *Science* **312**, 1944–1946 (2006).
- The role of fitness in natural microbial populations.**
70. Weissman, S. J. *et al.* Differential stability and trade-off effects of pathoadaptive mutations in the *Escherichia coli* FimH adhesin. *Infect. Immun.* **75**, 3548–3555 (2007).
71. Smith, E. E. *et al.* Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc. Natl Acad. Sci. USA* **103**, 8487–8492 (2006).
72. Moran, N. A. & Plague, G. R. Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.* **14**, 627–633 (2004).
73. Dethlefsen, L., McFall-Ngai, M. & Relman, D. A. An ecological and evolutionary perspective on human–microbe mutualism and disease. *Nature* **449**, 811–818 (2007).
74. Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99**, 10494–10499 (2002).
75. Huber, J. A. *et al.* Microbial population structure in the deep marine biosphere. *Science* **318**, 97–100 (2007).
76. Johnson, S. I. *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740 (2006).
77. Thompson, J. R. *et al.* Genotypic diversity within a natural coastal bacterioplankton population. *Science* **307**, 1311–1313 (2005).
78. Wilhelm, L. J., Tripp, H. J., Givan, S. A., Smith, D. P. & Giovannoni, S. J. Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biol. Direct* **2**, 27 (2007).
79. Wayne, L. G. *et al.* Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37**, 463–464 (1987).
80. Stackebrandt, E. & Ebers, J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol. Today* **33**, 152–155 (2006).
81. Murray, R. G. & Stackebrandt, E. Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int. J. Syst. Bacteriol.* **45**, 186–187 (1995).
82. Staley, J. T. The bacterial species dilemma and the genomic–phylogenetic species concept. *Philos. Trans. R. Soc. Lond. B* **361**, 1899–1909 (2006).
83. Dykhuizen, D. Species numbers in bacteria. *Proceedings Calif. Acad. Sci.* **56**, 62–71 (2005).
84. Nesbø, C. L., Dlutsek, M. & Doolittle, W. F. Recombination in *Thermotoga*: implications for species concepts and biogeography. *Genetics* **172**, 759–769 (2006).
85. Maiden, M. C. J. *et al.* Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl Acad. Sci. USA* **95**, 3140–3145 (1998).
- The first publication to describe the MLST concept.**
86. Hugenholtz, P., Pitulle, C., Hershberger, K. L. & Pace, N. R. Novel division level bacterial diversity in a Yellowstone hot spring. *J. Bacteriol.* **180**, 366–376 (1998).
87. Liolios, K., Tavernarakis, N., Hugenholtz, P. & Kyrpides, N. C. The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res.* **34**, D332–D334 (2006).
88. Ribosomal gene milestone met, already left in dust. *ASM News* **71**, 6–7 (2005).

## Acknowledgements

The authors thank F. Balloux, H. Daims, S.J. Giovannoni, K. Harris, M. Horn, B. Linz, A. Loy, W. Martin, Y. Moodley, N.R. Pace, M.F. Polz, P. Roumagnac, F.M. Cohan and B.G. Spratt for helpful comments on earlier versions of this manuscript. M.A. is supported by grant 05/FE1/B882 from the Science Foundation of Ireland and M.W. is supported by the University of Vienna in the framework of the “Symbioses and molecular principles of recognition” research focus.

## DATABASES

Entrez Genome Project: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj>  
[Bacillus licheniformis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Bacillus subtilis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Escherichia coli](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Helicobacter pylori](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Mycobacterium bovis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Mycobacterium tuberculosis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Neisseria gonorrhoeae](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Neisseria lactamica](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Neisseria meningitidis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Pseudomonas aeruginosa](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [S. Typhi](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Staphylococcus aureus](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Streptococcus agalactiae](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Vibrio splendidus](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Yersinia pestis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj) | [Yersinia pseudotuberculosis](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeproj)

## FURTHER INFORMATION

Mark Achtman's homepage:

<http://web.mpiib-berlin.mpg.de/mlst/AGroup>

Michael Wagner's homepage:

<http://www.microbial-ecology.net>

All Species MLST databases:

<http://pubmlst.org/databases.shtml>

GOLD Genomes OnLine Database:

<http://www.genomesonline.org>

PubMLST database: <http://pubmlst.org>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF