Opinion

Cell PRESS

A species concept for bacteria based on adaptive divergence

Michiel Vos

Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), 6666 GA Heteren, The Netherlands

Bacterial strains are currently grouped into species based on overall genomic similarity and sharing of phenotypes deemed ecologically important. Many believe this polyphasic taxonomy is in need of revision because it lacks grounding in evolutionary theory, and boundaries between species are arbitrary. Recent taxonomy efforts using multilocus sequence typing (MLST) data are based on the identification of distinct phylogenetic clusters. However, these approaches face the problem of deciding the phylogenetic level at which clusters are representative of evolutionary or taxonomically distinct units. In this review, I propose classifying two phylogenetic clusters as separate species only when they have statistically significantly diverged as a result of adaptive evolution. More than a method for classification, the concept of adaptive divergence can be used in a 'reverse ecology' approach to identify lineages that are in the process of speciation or genes involved in initial adaptive divergence.

The bacterial species problem

The simplest criterion for grouping bacterial strains into species (see Glossary) is that of overall similarity, and this phenetic species concept is the current standard for bacteria [1,2]. In practice, strains are assigned to the same species when DNA–DNA hybridization (DDH) values are greater than 70%, a cut-off value chosen because it roughly agreed with existing notions of what constituted different species [3]. DDH efficiency will decrease in cases of divergence of homologous genes by mutation and of divergence in gene content via lateral gene transfer (LGT). As DDH assays require cultured strains and are time consuming, the percentage similarity of the gen to the 16S ribosomal RNA subunit gene is often used as a proxy. A cut-off of 97% similarity in 16S rDNA is often used to group strains into Operational Taxonomic Units (OTUs). Strains that are less similar have never been found to belong to the same species; however, strains that are more than 97% similar in their 16S rDNA gene sequences can still be different species according to the DDH criterion [2].

In addition to DDH, morphological, physiological and biochemical traits are used to group strains into species. Some traits, often those that are of immediate relevance to humans, are even deemed important enough to override the genomic similarity criterion. For instance, *Mycobacterium tuberculosis* and *Mycobacterium bovis* are considered different species, even though this is not supported by the DDH and 16S rDNA criteria [4]. Furthermore, for species such as *Bacillus thuringiensis, Bacillus cereus* and *Bacillus anthracis*, species designation is based on phenotypes encoded by plasmids, loss or gain of which would change species designation [5]. Related bacterial strains usually differ in numerous phenotypes, and so it is difficult, if not impossible, to decide which differences should be used to classify strains into species and which differences can be ignored. Moreover, defining species based on presence or absence of particular phenotypes or genes is even more arbitrary than using the overall genomic similarity criterion, which at least can be applied equally to all types of bacteria. There is thus a widely recognized need for a novel species concept that: (i) is non-arbitrary, (ii) can be equally applied to all types of bacteria and archaea, and (iii) is based on the evolutionary processes that ultimately underlie bacterial diversity [2,3,6–14].

The concept of adaptive divergence

A recent framework of bacterial evolution and speciation that has gained widespread interest is the stable ecotype model developed by Cohan and coworkers [8,9,14,15]. In this model, a single clone diversifies into a cluster of

Glossary

Allopatric divergence: genetic divergence as a result of reduced gene flow resulting from geographic separation.
Clade: an evolutionary lineage (including all descendants of this lineage).
DNA–DNA hybridization (DDH) assay: measures the degree of genetic similarity between two genomes. Hybridization is less efficient when shared genes have genetically diverged and/or when genomes have differentially acquired or lost genes.
Fixation: the spread of a mutation through the entire population by selection or drift.
Genetic (or neutral) drift: the change in frequency of a mutation in a population as a result of chance, not selection.
Homologous recombination: the uptake and incorporation of a stretch of homologous DNA from a donor cell by a recipient cell.
Lateral gene transfer: the uptake and incorporation of a stretch of non-homologous DNA from a donor cell by a recipient cell.
Multilocus sequence typing (MLST): a typing scheme originally developed for pathogenic bacteria based on sequencing of several (usually seven) housekeeping gene fragments.
Neutrality Index (NI): a measure indicating the relative importance of adaptive evolution or neutral drift in fixing nucleotide differences between two taxa.
Operational Taxonomic Unit (OTU): a pragmatically defined biological type. In 16S rRNA sequence based surveys of bacteria and archaea, sequences that are >97% similar are usually grouped into a single OTU.
Species: a taxonomic rank for which many definitions exist. The species definition advocated here groups individuals into a species when they form a cluster that has diverged from its ancestral cluster as a result of statistically significant adaptive evolution
Sympatric divergence: genetic divergence as a result of reduced gene flow resulting from ecological separation (rather than geographical separation)

Corresponding author: Vos, M. (michiel.vos@nioo.knaw.nl).

Box 1. Selection

Two main types of selection are often distinguished: positive selection promoting particular mutations, and negative (or purifying) selection removing particular mutations. Here positive selection is used in the narrow sense of selection promoting a mutation until it reaches fixation (in a 'selective sweep'). This distinguishes it from diversifying selection, in which mutations do not reach fixation as their fitness is negatively correlated with their frequency (for example in surface antigens, recognizable by many nonsynonymous changes).

Because MLST studies routinely test for selection using the d_N/d_S test and usually find only purifying selection, the presence of extensive positive selection in more distantly related housekeeping genes reported here might be unexpected. This seeming discrepancy disappears when one realizes that MLST studies usually focus on a single species cluster. Within such clusters there is indeed selection against most protein changes (d_N/d_S <1). When the occasional beneficial protein change does occur, it will sweep to fixation in relatively short time. These fixed changes can be detected by comparing two clusters using the MK test but are invisible to the d_N/d_S test, which is most useful for detecting diversifying selection within a cluster. One well documented example of an excess of fixed protein changes in a core gene detected by the MK test is the divergence of a homologous regulatory gene controlling lipopolysaccharide modification mediating antibiotic resistance between *Escherichia coli* and *Salmonella enterica* [19].

clones as a result of neutral mutations and random genetic drift. Occasionally, a beneficial mutation will arise in one of these diverged clones and sweep to fixation (positive selection) (Box 1). By pruning the cluster back to a single clone, such successive selective sweeps will restrict genetic divergence within a cluster. Strains that inhabit different niches will be affected by different selective sweeps, and so sequence clusters will diverge from each other because of the fixation of different adaptive mutations (Figure 1a). Because such distinct sequence clusters reflect differential ecological adaptation, they could be taken to represent the fundamental units of bacterial biodiversity [15].

The McDonald–Kreitman (MK) test [16] can be used to detect whether two sequence clusters have diverged as a result of recurrent selective sweeps (periodic selection), as described by the ecotype model, or alternatively as a result of the random accumulation of neutrally behaving mutations (genetic drift). The MK test was originally developed to detect past selection in pairs of named species [16] For instance, it has been used to show that many homologous genes in *Escherichia coli* and *Salmonella enterica* have diverged through differential positive selection [17–19]. In this article, the application of the MK test is reversed; sequence clusters are classified as separate species only when their divergence is found to be the result of significant positive selection. The use of the MK test to identify ecologically distinct bacterial types has been pioneered by Simmons and coworkers [20], and is developed further here.

The MK test recognises two types of nucleotide substitutions: synonymous substitutions that do not result in protein changes and are considered neutral, and nonsynonymous substitutions that do result in protein changes and can be considered either neutral, deleterious or adaptive [21]. When clusters diverge through neutral drift, the differences that are fixed between clusters are expected to

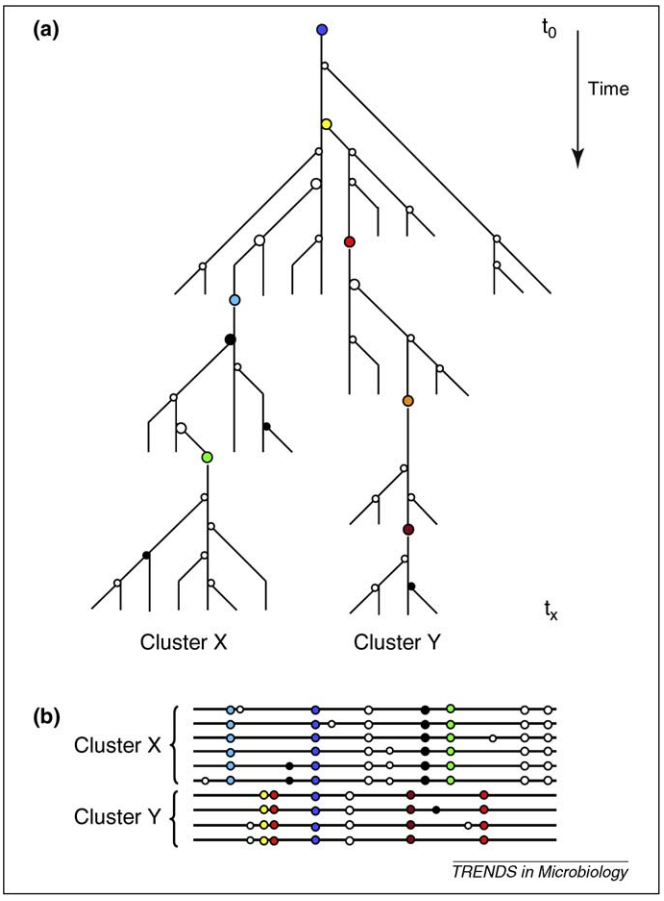

Figure 1. Divergence of two bacterial lineages under the stable ecotype model. (a) An ancestral clone (blue circle) at time t_0 diversifies through neutral mutations (circles at tree nodes). A novel adaptive mutation appears in the population as well (yellow circle). Clones carrying this mutation occupy a niche distinct from the ancestral blue population, and so both types can coexist. The yellow mutant subsequently diversifies through neutral mutations also, until a second adaptive mutation (red circle) appears. The mutant carrying both yellow and red mutations outcompetes the clones carrying only the yellow mutation, and subsequently diversifies again. Likewise, the ancestral blue cluster evolves by the fixation of novel mutations that increase fitness. As both clusters inhabit different niches, selective sweeps in cluster X do not affect clones in cluster Y, and vice versa. Note that mutations that are universally advantageous are expected to homogenize distinct clusters. However, because recombination uncouples the fate of different mutations, homogenization will only affect the genomic region containing this universally beneficial mutation. **(b)** DNA sequences sampled from clusters X and Y at t_x from **(a)**. Polymorphisms present within clusters are represented by small circles, and differences fixed between clusters are represented by large circles. Polymorphisms are either neutral (small white circles) or slightly deleterious (small black circles). Large coloured circles are fixed adaptive differences. Large white circles are neutral polymorphisms that have swept to fixation along with the novel adaptive mutation (a process termed 'genetic hitchhiking'). Note that the ancestral blue mutation is shared by both clusters, whereas the subsequent adaptive mutations are unique to clusters. The NI here indicates an excess of protein-changing fixed differences relative to protein-changing polymorphisms as a result of adaptive evolution.

mirror the frequency of synonymous and nonsynonymous polymorphisms within clusters [22]:

$$NI = (P_N/D_N)/(P_S/D_S), \quad (1)$$

where NI is the Neutrality Index, P_N and P_S are, respectively, nonsynonymous and synonymous polymorphisms within clusters and D_N and D_S are, respectively, nonsynonymous and synonymous fixed differences between clusters. When NI = 1, there is no difference in pattern between nonsynonymous and synonymous substitutions, and clusters diverge neutrally. With NI,<1, fixed differences between species are due to nonsynonymous differences more

Box 2. Challenges to the adaptive divergence species concept

1. ***Species will harbour extensive genetic and ecological diversity***

Species boundaries in the adaptive divergence concept are broader than those currently used in the polyphasic species concept (Figure 2). High intraspecies diversity, especially compared with eukaryotes [59], has been viewed by some as a disadvantage, because it can result in an underestimation of functional diversity [7]. For example, many distinct phenotypes can be present among closely related, clinically relevant strains. This is at least partly the reason why many of the current efforts in bacterial systematics have focused on very closely related sequence clusters [12,14,26,60,61]. However, it could also be argued that broad species boundaries are to be preferred, as environmental pyrosequencing studies are revealing that the number of bacterial taxa could be very large indeed.

2. ***Calculating NI for whole genomes will obscure biologically relevant variation in the degree of adaptive divergence between genes***

The MK test is relatively insensitive to the number of individual sequences used [21]. Indeed, the test can be performed using only one individual sequence for one of two contrasted taxa. However, only minute portions of the core genome have been sampled for the comparisons summarized in Figure 2. Genome-wide analyses have demonstrated that NI values for different genes can be either positive or negative [17,62], which was also evident in analyses of individual loci belonging to the concatemer sequences in Figure 2. Using a small number of genes could therefore generate confounding results (as might be the case for *Staphylococcus* in Figure 2). Moreover, even in the absence of chance effects or varying selection pressure over the genome, the use of a limited number of loci is not ideal, as it can be shown that concatenating two loci that individually have NI = 1 can lead to either NI = 1, N < 1 or NI > 1 [63].

Challenges 1 and 2 could be simultaneously met by also applying the MK test to ecologically distinct lineages found within a species. When different strains are found to exhibit adaptive divergence in certain, niche-specific genes, they could be recognized as distinct ecotypes, perhaps by extending the binomial name to a trinomial name [9]. Likewise, groups of strains harbouring distinct accessory genes conferring ecologically relevant phenotypes could still be recognized.

3. ***Effects of demography can obscure effects of selection***

The assumption of the MK test that deleterious nonsynonymous mutations are efficiently removed by negative selection is actually too simplistic. First, selection is weak when the deleterious effects of mutations are very small. Second, selection is weak when population sizes are small and the effect of chance (genetic drift) is relatively large. Bacterial populations usually harbour many slightly deleterious mutations that are in the process of being removed by selection [64,65]. By elevating P_N/D_N, these slightly deleterious mutations result in an underestimation of adaptive divergence. Moreover, past changes in population size affecting the relative contribution of genetic drift could lead to spurious interpretations of the MK test [21,62,66].

An alternative explanation for NI < 1 is the fixation of slightly deleterious mutations as a result of chance effects caused by population bottlenecks [62]. Although subsequent population growth and a return of efficient selection would restore P_N/P_S, the past increase in D_N/D_S would remain visible. A recent critical examination of the MK test [62] found that low P_N/P_S (efficient purifying selection within clusters) was a better predictor of low NI than was high D_N/D_S (adaptive divergence between clusters), which could fit this scenario. McDonald and Kreitman themselves did not find it parsimonious to assume past bottlenecks (or alternatively, recent population expansions) [16]. Effective population sizes for many of the better known bacterial taxa are assumed to be sufficiently large not to be greatly affected by genetic drift. However, it is possible that the founding of a new cluster as the result of a new adaptation represents a bottleneck in itself [62].

More data on bacterial population biology as well as the development of extensions of the MK test controlling for demography [66] and statistical bias [67] are clearly desirable. It must be noted that even without a complete understanding of the mechanisms that determine its value; the NI could still be useful as a statistic to delineate taxa.

often than expected, and are assumed to be selected for (positive selection). With NI >1, fixed differences between species are due to nonsynonymous divergence less often than expected. This is caused by selection against protein changes (negative selection), with divergence primarily being driven by neutral fixation of synonymous substitutions (drift) (Figure 1b).

Two clusters can each be classified as a distinct species when the MK test yields a significant NI value (<1), whereas they can be considered to belong to the same species when the two clusters have not diverged as a result of adaptive evolution and the NI value is nonsignificant. To demonstrate the utility of this species concept based on adaptive divergence, the MK test was performed on various pairs of sequence clusters downloaded from public databases. Ideally, the MK test is performed on all genes shared by related strains. However, as a proxy of the core genome, a concatemer of multiple protein encoding 'housekeeping' loci is often used for population genetic analyses, a method termed 'multilocus sequence typing' (MLST) [23]. Only a small number of studies focusing on population level variation in two related taxa exist, and few used the MLST method. Therefore, lower resolution, single locus datasets were also analyzed. Note that the limited amount of sequence data used per comparison means that this analysis is used only as an illustration of how the adaptive divergence species concept works (Box 2).

Sequence clusters that are not very distinct were found to have diverged predominantly through drift (Figure 2). According to the stable ecotype model, this is because both clusters are subject to the same selection pressures and thus are affected by the same periodic selection events, regularly capping inter-cluster divergence [15]. Genetic divergence between more distantly related clusters is increasingly found to be the result of positive selection (Figure 2). Different adaptive mutations have swept to fixation in each of the paired sequence clusters, reflecting adaptation to different niches. The continuous nature of speciation [24] inevitably results in a grey area in which the MK test is indicative of adaptive divergence but is not (yet) significant (Figure 2, open circles below N I= 1, but see Box 2 for the *Staphylococcus* case).

Based on the genes used here, *B. cereus* and *B. thuringiensis* do not represent two separate species according to the adaptive divergence species definition (Figure 2), which is in line with current opinion [5]. *Burkholderia thailandensis* has been proposed to form a species separate from *Burkholderia pseudomallei* because it forms a distinct sequence cluster [6,25], but because the core genomes of both lineages seem to have diverged through drift rather than adaptation, they do not classify as separate species either (Figure 2). *Neisseria meningitidis* and *Neisseria lactamica* strains are separated by very few fixed nucleotide differences in the genes used here (Figure 2). Strains

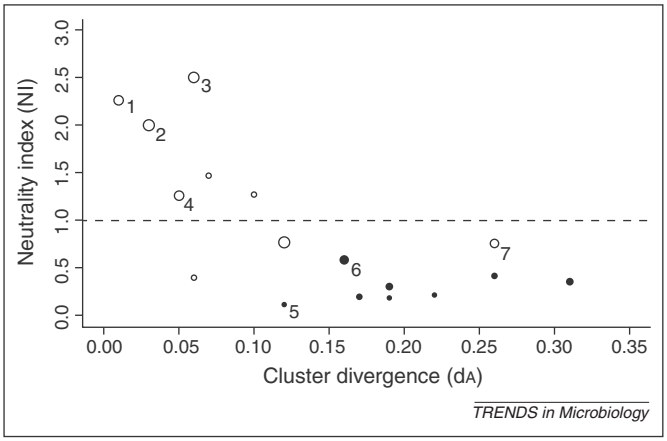

Figure 2. Adaptive divergence (NI) between sequence clusters plotted as a function of genetic divergence between clusters (d_A). Each data point represents a comparison of two sequence clusters. The dotted line (NI = 1) represents completely neutral divergence. Two clusters that have diverged significantly through differential adaptation (NI < 1) belong to separate species according to the definition of adaptive divergence, whereas clusters that have diverged through the fixation of synonymous mutations, with nonsynonymous mutations being selected against (NI > 1), are considered to belong to the same species. Open symbols indicate nonsignificance; closed symbols indicate significance (G-test, $P < 0.05$). Symbol size reflects the number of loci used in the pairwise comparison (one to seven loci, data summarized in Supplementary Table S1,online). For the labelled symbols, the following pairs of sequence clusters were used: **(1)** *Sulfolobus islandicus* Kamchatka population and *Sulfolobus islandicus* California population; **(2)** *Neisseria meningitidis* and *Neisseria lactamica*; **(3)** *Burkholderia thailandensis* and *Burkholderia pseudomallei*; **(4)** *B. cereus* and *B. thuringiensis*; **(5)** *Pseudomonas stutzeri* and *Pseudomonas syringae*; **(6)** *E. coli* and *S. enterica*; and **(7)** *Staphylococcus epidermis* and *Staphylococcus aureus*.

assigned to different *Neisseria* species frequently engage in homologous recombination with each other [26], and mutations are thus transferred between different lineages, preventing divergence.

It must be emphasized that the adaptive divergence concept only applies to those genes that are shared by the complex of strains under study, a set of genes known as the core genome [27,28]. The adaptive divergence species concept does not consider accessory genes that intersperse this frame of core genes. Although accessory genes are likely to play central roles in defining the ecological niche of a strain, they do not offer an opportunity to consistently delineate species, as clusters of related core genomes usually harbour many different combinations of such genes as a result of LGT or differential gene loss. This is exemplified by the weak correlation between the divergence of homologous genes and the percentage of gene content shared between genomes [29].

Insights into bacterial speciation

The previous section focused on a non-arbitrary way of classifying individuals into species. Although species delineation is of great practical use, species concepts ideally also shed light on how species evolve. The next section discusses the role of two processes that are key to speciation: biogeography and recombination.

Biogeography

It is immediately noticeable in Figure 2 that data points are absent from the top right corner. This indicates that genetic drift is limited in its power to generate distinct sequence clusters. Different nucleotide changes will accumulate, and occasionally fix, in geographically isolated populations between which migration is limited. This process, termed 'isolation by distance', has been shown to occur in several types of bacteria and archaea [e.g. 30–34]. However, in none of these cases was isolation by distance sufficiently strong to generate sequence clusters separated by many fixed differences. The relatively limited extent of isolation by distance is demonstrated by the comparison of two isolated *Sulfolobus* hot spring populations from Kamchatka and California (Figure 2). The hypothesis that allopatric speciation could occur in free-living bacteria [35,36] is therefore not supported by this analysis. Instead, it is more likely that adaptive mutations arise locally, sweep to fixation globally [37], and that genetic divergence is primarily a result of niche adaptation.

Homologous recombination

Homologous recombination occurs frequently in many bacterial taxa [38], and could impede adaptive divergence by homogenizing emerging lineages adapting to different niches [15]. Some barrier to homologous recombination must thus exist to allow sequence clusters to diverge. The efficiency of homologous recombination decreases exponentially with sequence divergence [39,40] (without a sharp drop that could correspond to a boundary between clusters [41]). Geographic isolation is not expected to result in genetic divergence sufficient to greatly reduce homologous recombination between clusters [12,20,42]. Purely mechanistic barriers to homologous recombination, such as changes in restriction–modification systems or differential regulation of competence [43], can evolve readily. However, such barriers would have to coincide with a shift in niche in order for speciation to take off (or the actual mechanism preventing recombination must 'magically' also have ecological significance), which might not be likely [44].

At the level of (currently recognized) species, bacteria seem to be cosmopolitan, and so sister taxa overlap in their geographical distribution. However, although on a global scale "everything is everywhere", "the environment selects" on smaller scales [45]. It could be envisaged that niche shifts result in a different spatial or temporal occupation of the same location relative to the ancestral cluster, preventing the close contact between different cells that is needed for homologous recombination. This tight linkage between ecological specialization and the likelihood of homologous recombination could enable sympatric speciation, as has been observed in some phytophagous insects after a switch in host plant [44]. The observation that recombination is more likely to occur within rather than between distinct phylogenetic clades even when they are geographically intermingled [46] is consistent with this scenario. Sympatric speciation is generally considered uncommon in animals and plants, but could well be the rule, rather than the exception, for bacteria and archaea [9,15,47–50].

Speciation could be expected to proceed more easily in taxa that primarily evolve clonally. No studies exist examining the relationship between homologous recombination rate and species diversity. However, recombination rate is known to be high in some *Vibrio* species [38],

and this genus seems at least as speciose as some of the genera of which representatives are known to have a much lower homologous recombination rate. The lack of a negative correlation between species richness and homologous recombination (if confirmed) could indicate that ecological barriers are sufficient to prevent frequent homologous recombination. It could also indicate that homologous recombination can promote speciation in ways that outweigh its homogenizing effect. It is likely that at any given time, different adaptive mutations will be present in different individuals within a population. The ensuing competition between these individuals will slow down the fixation of these beneficial mutations, a process termed 'clonal interference'. Homologous recombination could facilitate the response of populations to natural selection by combining these adaptive mutations into single individuals, as is commonly assumed for eukaryotic sex [51]. Adaptive speciation thus could proceed faster in bacteria with high homologous recombination rates because they are expected to respond faster to divergent selection.

Any homologous recombination that does occur between two adaptively diverging lineages is expected to result in offspring of intermediate phenotype that have lowered fitness in each of the two parent niches. Note, however, that homologous recombination in genes that are not (at least initially) involved in adaptive divergence has no fitness consequence. Strong evidence for ongoing recombination between such genes long after the onset of divergence of niche specific genes has been presented for *E. coli* and *S. enterica* [52]. Only those niche shifts that result in the spatial or temporal segregation of populations will result in decreased levels of homologous recombination between two diverging clusters. Although named species often harbour a wide array of niche specialists, few of those will end up evolving into a separate lineage or species, as homologous recombination at loci not involved in niche differentiation will continue to tie the evolutionary fates of the different strains together. The enormous genetic and phenotypic diversity present in many bacterial species, especially compared with animals and plants, is consistent with this scenario.

The inability of two clusters to engage in homologous recombination classifies them as species according to the biological species concept [24]. This is probably the most well known species concept (although it is frequently overlooked that many eukaryote classifications are not actually based upon it). It has been proposed to apply this species concept to bacteria by testing whether signatures of homologous recombination are evident within, but not between, clusters of related strains [53]. Because the MK test takes into account both mutations shared through ancestry and mutations shared through homologous recombination, the adaptive divergence species concept fully encapsulates the biological species concept.

Conclusions

Darwin was much struck by the 'entirely vague and arbitrary distinctions between species and varieties' [54]. The question of what exactly constitutes a species continues to captivate biologists, including microbiologists. The advent of genomics has brought proposals for high resolution methods to delineate species based on levels of overall nucleotide similarity or shared gene content [29,55–57]. Although these approaches are a great improvement over DDH, the problem remains of deciding on a cut-off point at which genomic similarity between strains ends and genomic dissimilarity between species begins. Statistically significant adaptive divergence of the core genome offers a non-arbitrary measure to assign species status to sequence clusters retrieved from isolates or the environment.

The adaptive divergence species concept is explicitly based on evolutionary theory, specifically the stable ecotype model, and incorporates the processes of evolutionary descent, ecological adaptation and homologous recombination. It can be applied to MLST data, population genomic data and, preferentially, whole genome data. It could even be used to classify protein coding marker sequences from environmental samples into species.

More than a method for classification, the concept of adaptive divergence can shed light on how bacterial populations diversify. The approach outlined in Figure 2 could be used to identify lineages that are in the process of speciation, to study this process subsequently in more detail using genomic, phenotypic and environmental data. Intragenomic differences in adaptive divergence will offer valuable insights into the nature of bacterial speciation. For instance, plotting NI values of every individual gene shared by two closely related taxa could reveal if there is a tendency for certain types of genes to be involved in initial adaptive divergence.

The adaptive divergence concept is not likely to be applicable to eukaryotes, as the percentage of protein-coding sequence is generally small in eukaryotic genomes, and sequence evolution is less extensive in larger organisms, because of their smaller effective population sizes [58]. However, research on bacterial speciation will ultimately benefit from integration into the extensive body of speciation theory developed for plants and animals. The increased affordability of next generation sequencing technologies, and increased collaboration between microbial ecologists and population geneticists will undoubtedly help to answer many important questions on bacterial speciation (some of which are listed in Box 3) in the near future.

Box 3. Outstanding questions

- Can the enormity of bacterial diversity be explained solely by adaptation? Or can genetic drift, via isolation by distance or via bottlenecks, also play a significant role in the evolution of certain species clusters?
- At what spatial scales do distinct bacterial niches occur?
- How common is despeciation, the process by which two previously distinct clades hybridize after renewed ecological contact [68]?
- The higher the number of sampled strains, the smaller will be the total core genome, or 'pan genome', (and conversely the larger will be the accessory genome) [69,70]. It can be argued that those genes that are present in a relatively wide range of genomic backgrounds are likely to confer a more general selective advantage [28,71]. Do more widely shared genes exhibit lower degrees of adaptive divergence?

Acknowledgments
I thank Mark Achtman, Frederick Cohan, Adam Eyre-Walker, Bill Hanage, Austin Hughes and Noel Smith for helpful discussion and/or comments on the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tim.2010.10.003.

References

1 Stackebrandt, E. *et al.* (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52, 1043–1047
2 Rossello-Mora, R. and Amann, R. (2001) The species concept for prokaryotes. *FEMS Microbiol. Rev.* 25, 39–67
3 Achtman, M. and Wagner, M. (2008) Microbial diversity and the genetic nature of microbial species. *Nat. Rev. Microbiol.* 6, 431–440
4 Smith, N.H. *et al.* (2006) Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. *Nat. Rev. Microbiol.* 4, 670–681
5 Helgason, E. *et al.* (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*–one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* 66, 2627–2630
6 Gevers, D. *et al.* (2005) Opinion: Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.* 3, 733–739
7 Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1899–1909
8 Cohan, F.M. (2006) Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1985–1996
9 Cohan, F.M. and Perry, E.B. (2007) A systematics for discovering the fundamental units of bacterial diversity. *Curr. Biol.* 17, R373–386
10 Konstantinidis, K.T. *et al.* (2006) The bacterial species definition in the genomic era. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1929–1940
11 Fraser, C. *et al.* (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323, 741–746
12 Fraser, C. *et al.* (2007) Recombination and the nature of bacterial speciation. *Science* 315, 476–480
13 Ward, D.M. (1998) A natural species concept for prokaryotes. *Curr. Opin. Microbiol.* 1, 271–277
14 Koeppel, A. *et al.* (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *Proc. Natl. Acad. Sci. U. S. A.* 105, 2504–2509
15 Cohan, F.M. (2002) What are bacterial species? *Annu. Rev. Microbiol.* 56, 457–487
16 McDonald, J.H. and Kreitman, M. (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351, 652–654
17 Charlesworth, J. and Eyre-Walker, A. (2006) The rate of adaptive evolution in enteric bacteria. *Mol. Biol. Evol.* 23, 1348–1356
18 Shapiro, B.J. and Alm, E.J. (2008) Comparing patterns of natural selection across species using selective signatures. *PLoS Genet.* 4, e23
19 Winfield, M.D. and Groisman, E.A. (2004) Phenotypic differences between *Salmonella* and *Escherichia coli* resulting from the disparate regulation of homologous genes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17162–17167
20 Simmons, S.L. *et al.* (2008) Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *PLoS Biol.* 6, e177
21 Eyre-Walker, A. (2006) The genomic rate of adaptive evolution. *Trends Ecol. Evol.* 21, 569–575
22 Rand, D.M. and Kann, L.M. (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* 13, 735–748
23 Maiden, M.C.J. *et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145
24 de Queiroz, K. (2005) Ernst Mayr and the modern concept of species. *Proc. Natl. Acad. Sci. U. S. A.* 102, 6600–6607
25 Hanage, W.P. *et al.* (2006) Sequences, sequence clusters and bacterial species. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1917–1927
26 Hanage, W.P. *et al.* (2005) Fuzzy species among recombinogenic bacteria. *BMC Biol.* 3, 6
27 Lan, R. and Reeves, P.R. (2000) Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends Microbiol.* 8, 396–401
28 Wertz, J.E. *et al.* (2003) A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J. Evol. Biol.* 16, 1236–1248
29 Konstantinidis, K.T. and Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2567–2572
30 Whitaker, R.J. *et al.* (2003) Geographic barriers isolate endemic populations of hyperthermophilic archaea. *Science* 301, 976–978
31 Vos, M. and Velicer, G.J. (2008) Isolation by distance in the spore-forming soil bacterium *Myxococcus xanthus*. *Curr. Biol.* 18, 386–391
32 Silva, C. *et al.* (2005) Evolutionary genetics and biogeographic structure of *Rhizobium gallicum* sensu lato, a widely distributed bacterial symbiont of diverse legumes. *Mol. Ecol.* 14, 4033–4050
33 Papke, R.T. *et al.* (2003) Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* 5, 650–659
34 Martiny, A.C. *et al.* (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ. Microbiol.* 11, 823–832
35 Whitaker, R.J. (2006) Allopatric origins of microbial species. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 1975–1984
36 Papke, R.T. and Ward, D.M. (2004) The importance of physical isolation to microbial diversification. *FEMS Microbiol. Ecol.* 48, 293–303
37 Majewski, J. and Cohan, F.M. (1999) Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* 152, 1459–1474
38 Vos, M. and Didelot, X. (2009) A comparison of homologous recombination rates in bacteria and archaea. *Isme J.* 3, 199–208
39 Roberts, M.S. and Cohan, F.M. (1993) The effect of DNA sequence divergence on sexual isolation in *Bacillus*. *Genetics* 134, 401–408
40 Zawadzki, P. *et al.* (1995) The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics* 140, 917–932
41 Costechareyre, D. *et al.* (2009) Homologous recombination in *Agrobacterium*: potential implications for the genomic species concept in bacteria. *Mol. Biol. Evol.* 26, 167–176
42 Hanage, W.P. *et al.* (2006) Modelling bacterial speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 361, 2039–2044
43 Maughan, H. and Redfield, R.J. (2009) Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution* 63, 1852–1866
44 Bolnick, D.I. and Fitzpatrick, B.M. (2007) Sympatric speciation: Models and empirical evidence. *Annu. Rev. Ecol. Evol. Syst.* 38, 459–487
45 Baas Becking, L.G.M. (1934) *Geobiologie of inleiding tot de milieukunde.* W.P. Van Stockum & Zoon N.V.
46 Didelot, X. and Maiden, M.C. (2010) Impact of recombination on bacterial evolution. *Trends Microbiol.* 18, 315–322
47 Sikorski, J. and Nevo, E. (2005) Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15924–15929
48 Bailly, X. *et al.* (2007) Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of *Medicago* species. *J. Bacteriol.* 189, 5223–5236
49 Miller, S.R. *et al.* (2009) Ecological specialization in a spatially structured population of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 75, 729–734
50 Oakley, B.B. *et al.* (2010) Evolutionary divergence and biogeography of sympatric niche-differentiated bacterial populations. *Isme J.* 4, 448–497
51 Vos, M. (2009) Why do bacteria engage in homologous recombination? *Trends Microbiol.* 17, 226–232
52 Retchless, A.C. and Lawrence, J.G. (2007) Temporal fragmentation of speciation in bacteria. *Science* 317, 1093–1096
53 Dykhuizen, D.E. and Green, L. (1991) Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* 173, 7257–7268
54 Darwin, C. (1859) On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. John Murray. http://darwin-online.org.uk/

55 Konstantinidis, K.T. and Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* 187, 6258–6264
56 Goris, J. *et al.* (2007) DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91
57 Deloger, M. *et al.* (2009) A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J. Bacteriol.* 191, 91–99
58 Lynch, M. (2006) The origins of eukaryotic gene structure. *Mol. Biol. Evol.* 23, 450–468
59 Staley, J.T. (1997) Biodiversity: Are microbial species threatened? Commentary. *Curr. Opin. Biotechnol.* 8, 340–345
60 Papke, R.T. *et al.* (2007) Searching for species in haloarchaea. *Proc. Natl. Acad. Sci. U. S. A.* 104, 14092–14097
61 Hunt, D.E. *et al.* (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* 320, 1081–1085
62 Hughes, A.L. *et al.* (2008) Synonymous and nonsynonymous polymorphisms versus divergences in bacterial genomes. *Mol. Biol. Evol.* 25, 2199–2209
63 Shapiro, J.A. *et al.* (2007) Adaptive genic evolution in the *Drosophila* genomes. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2271–2276
64 Hughes, A.L. (2005) Evidence for abundant slightly deleterious polymorphisms in bacterial populations. *Genetics* 169, 533–538
65 Rocha, E.P. *et al.* (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. *J. Theor. Biol.* 2, 226–235
66 Eyre-Walker, A. and Keightley, P.D. (2009) Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* 26, 2097–2108
67 Stoletzki, N., and Eyre-Walker, A. (2011) Estimation of the neutrality index. *Mol. Biol. Evol*. 28, 63–70
68 Sheppard, S.K. *et al.* (2008) Convergence of *Campylobacter* species: implications for bacterial evolution. *Science* 320, 237–239
69 Tettelin, H. *et al.* (2008) Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* 11, 472–477
70 Lefebure, T. and Stanhope, M.J. (2007) Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome Biol.* 8, R71
71 Lan, R. and Reeves, P.R. (2001) When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends Microbiol.* 9, 419–424