

Genomics and the bacterial species problem

W Ford Doolittle and R Thane Papke

Address: Biochemistry and Molecular Biology, Dalhousie University, 5850 College Street, Halifax, Nova Scotia, Canada B3H 1X5.

Correspondence: W Ford Doolittle. Email: ford@dal.ca

Published: 29 September 2006

Genome **Biology** 2006, **7**:116 (doi:10.1186/gb-2006-7-9-116)The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2005/7/9/116>

© 2006 BioMed Central Ltd

Abstract

Whether or not bacteria have species is a perennially vexatious question. Given what we now know about variation among bacterial genomes, we argue that there is no intrinsic reason why the processes driving diversification and adaptation must produce groups of individuals sufficiently coherent in their genetic and phenotypic properties to merit the designation 'species' - although sometimes they might.

"The species problem is caused by two conflicting motivations; the drive to devise and deploy categories, and the more modern wish to recognize and understand evolutionary groups. As understandable as it might be that we try to equate these two, and as reasonable and correct as it might be to use taxa as starting hypotheses of evolutionary groups, the problem will endure as long as we continue to fail to recognize our taxa as inherently subjective, and as long as we keep searching for a magic bullet, a concept that somehow makes a taxon and an evolutionary group both one and the same."

Jody Hey [1]

Thus Jody Hey [1] dismisses the vast and highly philosophical literature on the meaning of the word 'species'. Of course, this literature overwhelmingly addresses species in the context of eukaryote (especially vertebrate) evolution, and seldom tackles the special problems that microbes pose. We microbiologists, to our credit, have often acknowledged that the exercise of formulating a useful 'species definition' and the quest for an underlying 'species concept' are not the exactly same [2-6]. But we too have a 'species problem'.

Species definition versus species concept

What we want from a species definition is a set of easily applied and stable rules by which to decide when two organisms are

similar enough in their genomic and/or phenotypic properties to be given the same name [5-8]. The needs for such a guide to taxonomic practice in medicine, biotechnology and defense are obvious, and even arbitrary rules to satisfy them would be better than no rules at all [9]. We look to a species concept, on the other hand, for a genetic and/or ecological model of bacterial diversification and adaptation. Ideally, this model would make sense of our definition, justifying the choice of one particular set of rules for defining species as less arbitrary, or more natural, than another [2-4,9-14]. Thus, while acknowledging the dual nature of our quest, we still hope for "a concept that somehow makes a taxon and an evolutionary group both one and the same" [1].

The prevailing bacterial species definition has species as a "category that circumscribes a (preferably) genomically coherent group of individual isolates/strains sharing a high degree of similarity in (many) independent features, comparatively tested under highly standardized conditions" [5]. In practice, degree of similarity is assessed in molecular terms: "a prokaryotic species is considered to be a group of strains (including the type strain) that are characterized by a certain degree of phenotypic consistency, showing 70% of DNA-DNA binding and over 97% of 16S ribosomal RNA (rRNA) gene-sequence identity" [5]. A more precise and appropriate modern measure, but limited in its application to sequenced genomes, is the average nucleotide identity (ANI) calculated from pairwise comparison of all genes shared between any two strains.

An ANI of 94% generally corresponds to other molecular species definitions and to traditional taxonomic practice [7], so a solid consensus definition, genomic in spirit, may be in the offing. The more we learn about genomes, however, the more unlikely it seems that any unifying species concept will be possible. In particular, lateral gene transfer (LGT), within-species genomic variability and homologous recombination all make it harder to imagine how any single model for the maintenance of genomic coherence could be broadly valid or why, when valid, groups that match any single species definition should be the inevitable outcome.

Lateral gene transfer and the origins of evolutionary novelty

In animal species, evolutionary novelties arise as mutant alleles within populations. Because of the presence of sex and recombination, selection can effect their fixation independently of alleles at other loci. Bacteria have been traditionally thought of as asexuals lacking recombination, with their populations being clones [2,15,16]. Favored alleles can still sweep to fixation, but they bring the rest of the genome in which they first occurred along for the ride. Still, even radical (species-founding) evolutionary novelties would originate as mutations occurring within the ancestral bacterial population. And, for both animal and bacterial species, genomic coherence - which we might define as a greater degree of similarity in gene content (the actual number and identity of the genes present) and gene sequence (the sequences of corresponding genes) within species than between species - would be maintained by the selective purging of variability, one gene at a time in sexual species and one genome at a time in asexuals. (In the early days of bacterial genetics, this genomic sweeping process was called 'periodic selection').

But genomics tells us that bacteria often acquire evolutionary novelties from outside the ancestral population by LGT [16-18]. Best studied, not surprisingly, are bacteria that have become pathogens by the acquisition of novel plasmids, chromosomal genes or mobile pathogenicity islands [19], but non-pathogens also evolve in this saltatory fashion. From a recent comparative genomics/metagenomics study of the cyanobacterium *Prochlorococcus*, the ocean's principal prokaryotic photosynthesizer, Coleman *et al.* [20] conclude that "genetic variability between phenotypically distinct strains that differ by less than 1% in 16S ribosomal RNA sequences occurs mostly in genomic islands. Island genes appear to have been acquired in part by phage-mediated lateral gene transfer, and some are differentially expressed under light and nutrient stress."

In this and many similar cases, many genes conferring a highly complex adaptation can be acquired in one event, instantly dividing a single population into two subpopulations that differ substantially in lifestyle but continue to share in a

common gene pool. LGT radically uncouples the evolution of phenotype from the evolution of the bulk of the genome, as this is reflected in overall genome similarity (coherence). For instance, *Bacillus anthracis* (strain Ames ancestor), *Bacillus cereus* (ATCC1098) and *Bacillus thuringiensis* (serovar konkukian str. 97-27) all show more than 94% ANI (and so are a single species by this criterion and others), and are highly syntenic in chromosome structure. And yet they are famously different in phenotype - a virulent pathogen and potentially lethal bioterror agent, a cause of food poisoning, and a popular eco-friendly organic biopesticide, respectively.

Within-species variability in gene content

For every acquired gene for which a role in a radical species-creating LGT event might be inferred, there will be dozens or hundreds more whose contributions - if any - to evolutionary novelty remain unknown. And even within species as traditionally defined there can be enormous strain-to-strain variation in gene content. In a survey of 33 clusters of strains (with 2-11 genomes per cluster) that would be considered species by the greater than 94% ANI criterion, we find anywhere from 1 to 4,404 genes per cluster that are present in some strains but absent from others (O. Zhaxybayeva, C.L. Nesbø and W.F.D, unpublished work). From a similar study, Konstantinidis and Tiedje [7] observe that strains of the same species by this criterion "can vary up to 30% in gene content", and raise the possibility of resetting the 'species' to something like a 99% ANI cut-off.

Five years ago, when only the tip of the iceberg of variability in gene content was visible, Lan and Reeves [8] suggested that we look at 'species genomes' as comprising a core set (all genes present in at least 95% of strains) and an auxiliary set (present in 1-95% of strains). Something like this notion is embraced in the more recently articulated 'pangenome' concept, this term denoting the total number of genes found in at least one of the strains of a species [21]. In some species (such as *Bacillus anthracis*) the depth of the pangenome may have been plumbed after only a few genomes have been sequenced. For others, such as the ecologically versatile *Streptococcus agalactiae*, Tettelin *et al.* [22] suggest that "unique genes will continue to be identified even after sequencing hundreds of genomes."

This variability, we would argue, makes highly problematic one of the more appealing 'magic bullets' proposed for recognizing species as coherent natural units in the environment, namely as tight clusters of strains with very similar sequences for certain marker genes (sometimes 16S rRNA, sometimes more rapidly evolving genomic regions). Such 'microdiverse' clusters (Figure 1) are often observed in environmental surveys in which marker genes are amplified by PCR from environmental DNA samples, and have been interpreted in terms of Cohan's 'ecotype' model for bacterial species [5,11,23,24]. This model imagines that genomic

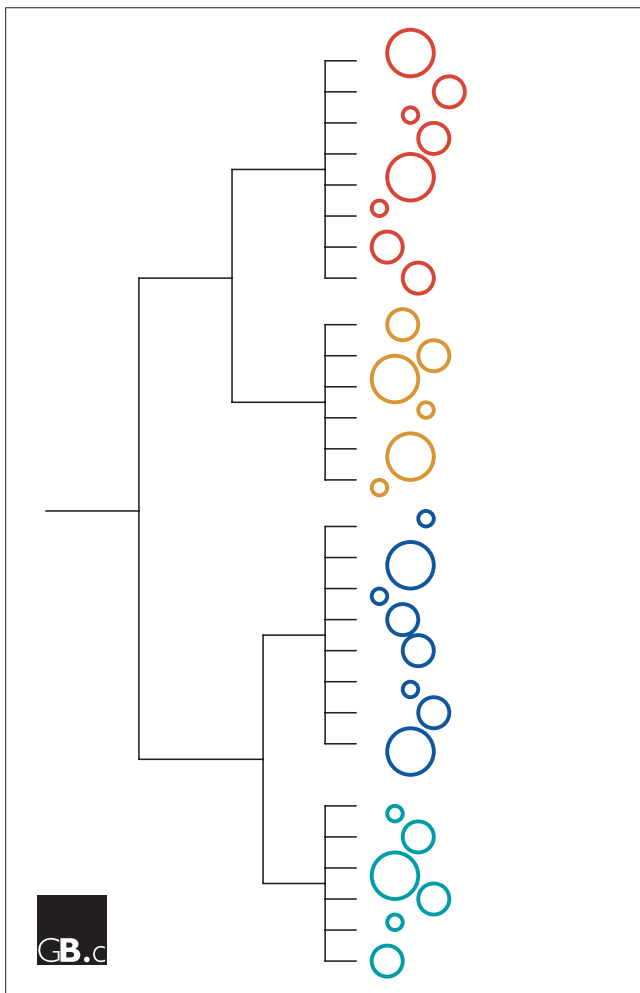


Figure 1
Microdiversity and diversity in gene content. Environmental surveys, using PCR amplification and sequencing of marker genes such as 16S rRNA or more rapidly evolving protein-coding genes and intergenic spacers, often reveal microdiverse clusters of strains with closely related sequences. The diagram shows a hypothetical phylogenetic tree compiled from such sequences, with each cluster indicated by a set of circles of the same color. Such a pattern of clustering by sequence might be expected if there were process other than random divergence and extinction of lineages at play (see Figure 2), and has been attributed [11,23,24] to an ecotype speciation process (see text). In this context, a microdiverse cluster might generally be a species. Comparisons of sequenced genomes for multiple strains of many designated species, and of genome sizes from isolates of others, show, however, that gene content can vary by up to 30% among different lineages of strains, even when the 'species' marker genes are identical in sequence [25]. The different sizes of the circles represent on an exaggerated scale the diversity in genome size in closely related strains found by such studies.

coherence within ecotypes is maintained by periodic selection, as discussed above, while barriers between ecological niches (spatial, temporal or nutritional) prevent genomes that sweep to fixation in one niche from invading another (Figure 2). The minor variations in marker gene sequences within a microdiverse cluster of isolates from a given site

would then just be neutral substitutions accumulated since the last diversity-purging genomic sweep of the ecotype.

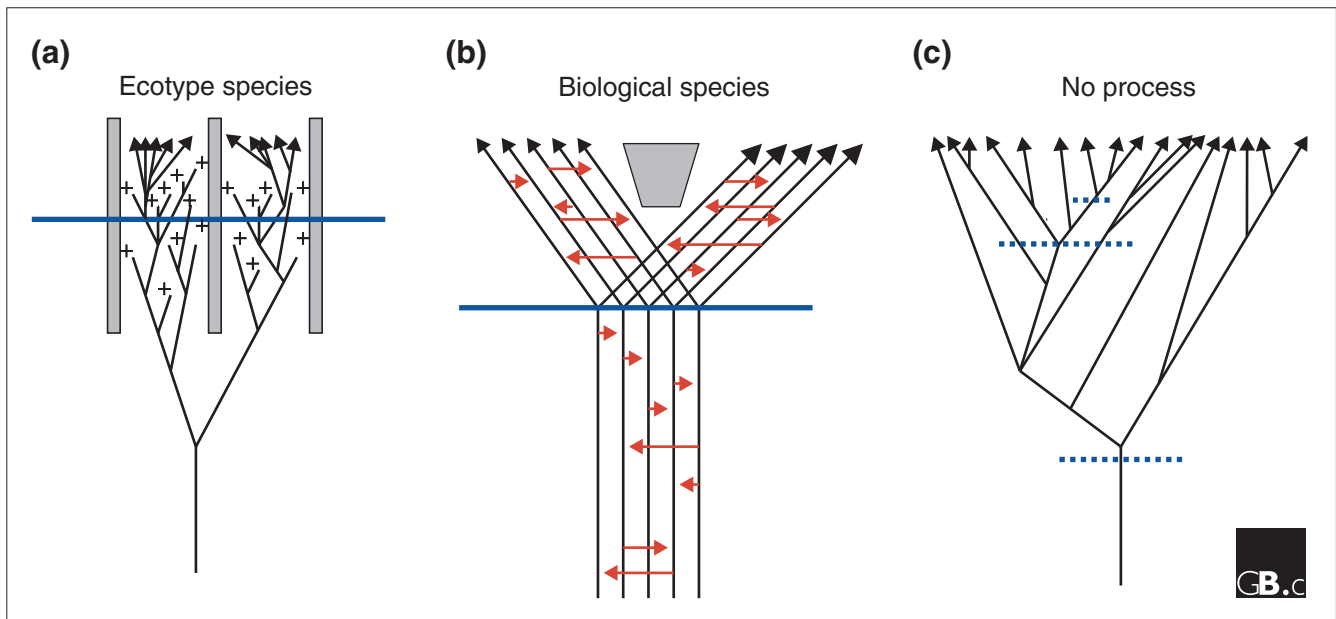
The problem here (as we might have predicted from the comparisons of sequenced 'conspecific' genomes discussed above) is that these same strains may be enormously more diverse in gene content than they are in gene sequence (see Figure 1). In a survey of genome sizes of *Vibrio splendidus* isolates by pulsed-field gel electrophoresis, in which all the isolates were greater than 99% identical at the 16S level and all taken from a single site (albeit at multiple times) on the coast of Massachusetts, Thompson *et al.* [25] concluded that "this group consists of at least a thousand distinct genotypes, each occurring at extremely low environmental concentrations (on average less than one cell per milliliter)." Genome sizes varied by as much as 1 Mb among them. The authors' suggestion that much of the observed genome size (and hence gene content) variation may be selectively neutral is attractive. What clearly cannot be supported, however, is the notion that species *qua* ecotypes are genomically coherent.

Homologous recombination in bacteria

Another surprise of the past decade is that bacteria are not all asexuals lacking recombination, but that in some homologous recombination is so frequent that it easily outperforms mutation as a source of strain-to-strain sequence differences [26]. The evidence for this comes from multi-locus sequence analysis (MLSA) based on sequences from five to seven unlinked core housekeeping genes amplified from scores or hundreds of strains of a species and, more recently, from the use of recombination detection algorithms [27] with aligned long segments or entire genomes (from fewer strains). As Dykhuizen and Green presciently observed some 15 years ago [12], we might apply to such recombining groups something like Ernst Mayr's 'biological species concept' (BSC). In this context the BSC would require that a bacterial species maintains genomic coherence because its members share an exclusive common gene pool (see Figure 2). Different species would have separate gene pools, and diverge and adapt through the separate fixation within them of favorable mutations or laterally acquired genetic novelties.

If we are to base a robust bacterial species concept on such a traditional model we must know first, whether biological barriers to exchange between gene pools of related species can be expected to define species boundaries with anything like the sharpness that various prezygotic (for example, mating behavior) and postzygotic (for example, hybrid sterility) factors define animal species [2], and second, whether such sharpness is indeed observed. Both are in question.

One barrier to exchange could be a precipitous decline in the frequency of homologous recombination as sequences diverge. The strength of this barrier will vary between

**Figure 2**

Models of processes that promote genomic coherence. **(a)** The ecotype species concept and **(b)** the biological species concept both entail processes that lead to genomic coherence within populations and divergence (horizontal dimension) between populations. Black arrowheads indicate organisms or isolates. The crosses in **(a)** indicate the clones eliminated in the process, while the red arrows in **(b)** indicate recombination between genomes. Blue lines indicate speciation. **(c)** If only random lineage splitting and lineage extinction occurred, coherence would not be expected, and the designation of speciation events (dashed blue lines) would be arbitrary. In the ecotype (periodic selection) model in **(a)**, which is applicable to organisms without significant genetic recombination, favorable mutations sweep to fixation, carrying the genome in which they first occurred along, so that diversity is reduced to zero at all loci. Accumulation of neutral mutations, prior to the next sweep, generates the sort of microdiversity illustrated in Figure 1. Gray bars are niche boundaries. In the biological species model, it is individual favorable mutations that are fixed, because recombination (indicated by red arrows) separates them from alleles at other loci in the genome in which they first occurred. Still, recombination at all loci will in time promote genomic coherence within populations and divergence between populations, because with time all alleles at all loci will be traceable to mutations that occurred within the population. The gray block indicates a barrier to recombination.

species because of idiosyncrasies of the recombinational machinery. More interestingly, it should also vary between genes because of their different rates of sequence divergence. And it does vary within species, thanks to mutations in the mismatch repair system, which can increase homologous recombination between moderately diverged (1-2%) genomes 1,000-fold, and permit homologous recombination between highly divergent (20%) sequences. Townsend *et al.* [28] calculate that such mutations elevate rates of adaptive evolution several thousand-fold, and the facts that mismatch repair mutants are common in nature (as if hitchhiking on the favorable recombination events they encourage) and that mismatch repair genes are often themselves mosaics (as if frequently themselves restored by homologous recombination) are good evidence that much adaptive evolution occurs through this transiently open window.

Other barriers to exchange would be peculiarities of the molecular machineries of transduction (transfer of bacterial DNA as part of a phage genome), conjugation and (to a lesser extent) transformation. The host specificity of phages, for instance, might be the principal factor defining the scope of the gene pools for those bacteria for which transduction is

the principal mode of genetic exchange. But some agents of bacterial gene transfer (plasmids and conjugation machinery) are highly promiscuous, mobilizing DNA transfer between phyla or even across domain boundaries: *Escherichia coli* can in fact conjugate with yeast [29]! Unlike the reproductive machineries of eukaryotes, these agents are clearly selfish genetic elements, whose own evolutionary interests are best served by violating, not maintaining, species boundaries. Furthermore, the introduction of substantial segments of novel DNA by LGT - which such agents also promote - can have interesting positive and negative effects on barriers to homologous recombination. Lawrence [13] argues that advantageous LGT acquisitions, by suppressing recombination in regions flanking their insertion sites, will permit sequence substitutions to accumulate, further strengthening regional barriers to homologous recombination. Contrariwise, we [14] have suggested that long segments introduced by LGT should be receptive to subsequent homologous recombination events involving the donor species, which might indeed share the same physical environment. Thus one organism could be a member of two or more otherwise quite distinct 'species' simultaneously, if species are defined by shared gene pools (Figure 3).

Species boundaries: sharp, fuzzy, or nonexistent?

Although the periodic selection process at the heart of Cohan's ecotype model [11] will produce both genomic coherence and ecologically driven divergence if operating alone, homologous recombination between ecotypes can disrupt both these properties at all but the loci under selection. Although homologous recombination operating within, but not between, populations will promote both coherence and divergence, the barriers to between-population homologous recombination are contingent on many factors and unlikely to produce species of similar genomic coherence across the board. And crucially, LGT has the potential to radically disrupt any genomic coherence achieved by either model. Contingent ecological and biological factors (like the host specificities of phages, the prevalence of mismatch-repair mutants or the

selective advantages of acquiring specific long DNA segments) will all affect coherence one way or another. We know too little about the frequencies of any of the underlying processes to predict their net effect - but enough to guess that it will not always be the same. We do know that coherence at the level of gene sequence (as measured by any single marker gene or by ANI) is very poorly coupled to coherence at the level of gene content (see Figure 1), however that might be maintained. And yet gene content is quite possibly the better predictor of coherence at the level of phenotype.

Indeed, genomics has given us too many processes with too many possible synergistic and antagonistic effects on genomic coherence - and in most cases we know too little about their relative magnitudes - to predict outcomes. If coherence were the usual observation, that is, if bacteria almost always fell into discrete clusters defined genomically (even if not phenotypically), then we would have an ample repertoire of known processes to explain this behavior - although still no reason to presume that the explanation would always be the same. But if such coherence were not the usual observation, then we could use what we know about process to explain that too.

So what is the usual observation? Opinions on this seem unstable. In 2002, Cohan [11] wrote that "bacterial species exist - on this much bacteriologists can agree", while Stackebrandt *et al.* [6] asserted that "experimental and theoretic evidence is compelling that the 'lumpy diversity' present in prokaryotes is recognizable as discrete centers of variation when appropriate methods are applied." In 2005, however, both Cohan and Stackebrandt were authors on a publication that suggested that "it might not be possible to delineate groups within a continuous spectrum of genotypic variation: that is, clustering might not occur ..." [5].

A path more squarely down the middle was taken by Hanage *et al.* [10] in summarizing an MLSA study of *Neisseria*.

"The bacterial domain of life is not uniform. Instead we see clumps of similar strains that share many characteristics, and with an innate human urge to classify, we have defined these as species. This work shows that by applying a simple approach using sequence data from multiple core housekeeping loci, we can resolve those clusters, provided such clusters exist. However, these species clusters are not ideal entities with sharp and unambiguous boundaries; instead they come in multiple forms and their fringes, especially in recombinogenic bacteria, may be fuzzy and indistinct." [10].

The solution to the bacterial species problem

To return to our original quotation, Hey [1] is right in the case of bacteria too: the species problem is very much in our

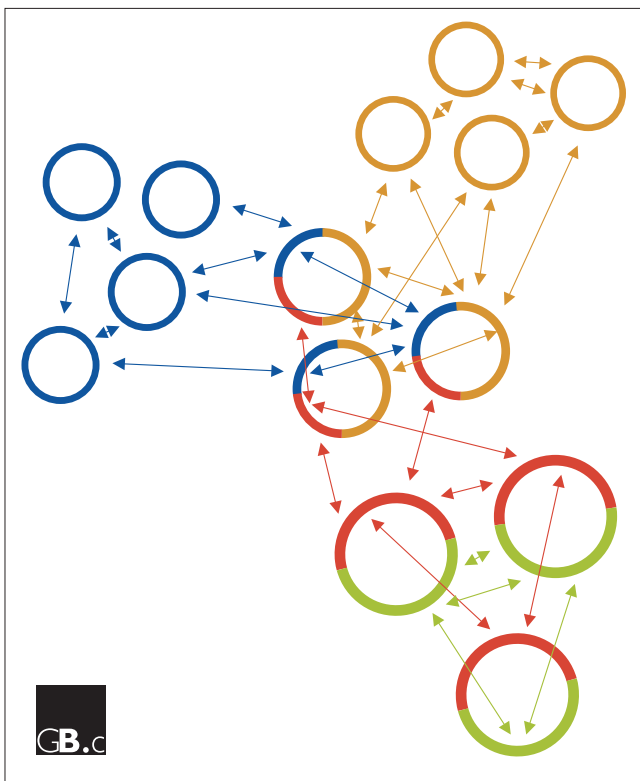


Figure 3

Lateral gene transfer and homologous recombination together can produce organisms effectively belonging to several species at once. The all-blue, all-gold and red/green circles represent genomes from three different bacterial groups that might be designated species. Each circle represents an individual genome. There is effectively no homologous recombination (arrows) between genomes or areas of different colors. LGT has, however, recently created a mosaic genome (center), with segments derived from blue, gold and red/green species (itself a mosaic). Homologous recombination can occur between a segment introduced by LGT and the corresponding region of the original donor strain. Coherence is maintained between the segments and the donor DNA, as in the biological species model. This cartoon is of course unrealistic in several respects: regions shared between species are more likely to be scattered as islands in the genome, and the number of species to which some part of any genome belongs could be much greater.

heads. Sometimes the many contingent genetic and ecological forces driving bacterial genome evolution will have produced clusters of genomes so much like each other and so much unlike any others in the world that even the tightest species definition will be satisfied. Sometimes this will merely appear to be so, because we have selected as medically interesting, or have been able to culture, certain organisms only by virtue of their possession of a single gene, while a spectrum of otherwise genomically similar relatives lacking it have gone unnoticed. Sometimes it will not be so, the contingent genetic and ecological forces working against each other and producing 'clusters' so fuzzy and with gene content versus genome sequence incongruities so striking that even the loosest criteria for genomic coherence cannot be met. We might, in an effort to match definition and concept, choose to think of genuine 'species' as those evolutionary groups that both satisfy an accepted species definition based on genomic coherence and whose coherence can be understood as the product of a biological process, as in the ecotype or BSC model. But many bacteria will not belong to such groups - and it is not a given that any such 'genuine' species exist.

There will, of course, always be a need to have some agreed-upon way of naming organisms, some species definition. Konstantinidis and Tiedje [7] suggest, primarily because of variability in gene content among closely related strains, that "standards could be as stringent as including only strains that show a greater than 99% ANI, or are less identical at the nucleotide level but share an overlapping ecological niche." But they do not endorse such a tightening up, because this "would instantaneously increase the number of existing species probably by a factor of 10, and cause considerable confusion in the diagnostic and regulatory (legal) fields". Without a magic bullet that makes our species definition and our species concept (or concepts) "one and the same", such expediency considerations will always - and legitimately - play a role in defining species.

It will often also be expedient to think in terms of lineages of strains within species and of phylogenetic relationships between species. There seems to be no other sensible way of doing this than to use concatenated shared (core) genes, and to represent the results as trees [18,30,31]. Useful as such trees may be, we must realize that they will not represent the true intergenomic relationships in recombinogenic groups, which will be reticulate, not tree-like - nor will they describe the evolutionary behavior of the non-core part of the pangenome of any species, which may be much larger than the core [32].

In understanding genome evolution, the 'species concept' does limited work. The ecotype and BSC models (see Figure 2) are useful heuristics, but calling them models for speciation does not make them more useful. In biogeography and biodiversity studies, the word 'species' may actually work

some mischief. Questions such as 'How many species of bacteria are there?' or 'Are bacterial species cosmopolitan?' are invaluable in stimulating research into the diversity and distribution of microbial genotypes and phenotypes. But without a species definition coupled to a magic bullet concept that guarantees that defined species are natural biological entities, these questions would be better reformulated in terms of genotypes and phenotypes. There will never be such a magic bullet. In using species concepts, we microbiologists would do well to follow the advice of a philosopher, William James, who wrote: "Since it is only the conceptual form which forces the dialectic contradictions upon the innocent sensible reality, the remedy would seem to be simple. Use concepts when they help, and drop them when they hinder, understanding."

Acknowledgements

We thank Joe Bielawski, Eric Bapteste, Paco Rodriguez-Valera and Olga Zhaxybayeva for invaluable comments, and CIHR and Genome Atlantic for support.

References

- Hey J: **The mind of the species problem.** *Trends Ecol Evol* 2001, **16**:326-330.
- Franklin, L: **Bacteria, sex and systematics.** *Philosophy of Science* 2006, in press.
- Ward DM: **A natural species concept for prokaryotes.** *Curr Opin Microbiol* 1998, **1**:271-277.
- Rosselló-Mora E, Amann R: **The species concept for prokaryotes.** *FEMS Microbiol Rev* 2001, **25**:39-37.
- Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coeyne T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J: **Opinion: Re-evaluating prokaryotic species.** *Nat Rev Microbiol* 2005, **3**:733-739.
- Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, Nesme X, Roselló-Mora R, Swings J, Trüper HG, et al.: **Report of the ad hoc committee for the reevaluation of the species definition in bacteriology.** *Int J Syst Evol Microbiol* 2002, **52**:1043-1047.
- Konstantinidis KT, Tiedje JM: **Genomic insights that advance the species definition for prokaryotes.** *Proc Natl Acad Sci USA* 2005, **102**:2567-2572.
- Lan R, Reeves PR: **Intraspecies variation in bacterial genomes: the need for a species genome concept.** *Trends Microbiol* 2000, **8**:396-401.
- Godreuil S, Cohan F, Shah H, Tibayrenc M: **Which species concept for pathogenic bacteria? An E-Debate.** *Infect Genet Evol* 2005, **5**:375-87.
- Hanage WP, Fraser C, Spratt BG: **Fuzzy species among recombinogenic bacteria.** *BMC Biol* 2005, **3**:6.
- Cohan FM: **What are bacterial species?** *Annu Rev Microbiol* 2002, **56**:457-487.
- Dykhuizen DE, Green L: **Recombination in *Escherichia coli* and the definition of biological species.** *J Bacteriol* 1991, **173**:7257-7268.
- Lawrence JG: **Gene transfer in bacteria: speciation without species?** *Theor Popul Biol* 2002, **61**:449-460.
- Nesbø CL, Dlutek M, Doolittle WF: **Recombination in *Thermotoga*: implications for species concepts and biogeography.** *Genetics* 2006, **172**:759-769.
- Levin BR, Bergstrom CT: **Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes.** *Proc Natl Acad Sci USA* 2000, **97**:6981-6985.
- Gogarten JP, Doolittle WF, Lawrence JG: **Prokaryotic evolution in light of gene transfer.** *Mol Biol Evol* 2002, **19**:2226-2238.
- Doolittle WF: **Phylogenetic classification and the universal tree.** *Science* 1999, **284**:2124-2129.

18. Lerat E, Daubin V, Ochman H, Moran NA: **Evolutionary origins of genomic repertoires in bacteria.** *PLoS Biol* 2005, **3**:e130.
19. Dobrindt U, Hochhut B, Hentschel U, Hacker J: **Genomic islands in pathogenic and environmental microorganisms.** *Nat Rev Microbiol* 2004, **2**:414-424.
20. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, Chisholm SW: **Genomic islands and the ecology and evolution of *Prochlorococcus*.** *Science* 2006, **311**:1768-1770.
21. Fraser-Liggett CM: **Insights on biology and evolution from microbial genome sequencing.** *Genome Res* 2005, **15**:1603-1610.
22. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin A, et al.: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome".** *Proc Natl Acad Sci USA* 2005, **102**:13950-13955.
23. Acinas SG, Klepac-Ceraj V, Hunt DE, Pharino C, Ceraj I, Distel DL, Polz MF: **Fine-scale phylogenetic architecture of a complex bacterial community.** *Nature* 2004, **430**:551-554.
24. Giovannoni S: **Evolutionary biology: oceans of bacteria.** *Nature* 2004, **430**: 515-516.
25. Thompson JR, Pacocha S, Pharino C, Klepac-Ceraj V, Hunt DE, Benoit J, Sarma-Rupavtarm R, Distel DL, Polz MF: **Genotypic diversity within a natural coastal bacterioplankton population.** *Science* 2005, **307**:1311-1313.
26. Feil EJ, Spratt BG: **Recombination and the population structures of bacterial pathogens.** *Annu Rev Microbiol* 2001, **55**:561-90.
27. Mau B, Glasner JD, Darling AE, Perna NT: **Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*.** *Genome Biol* 2006, **7**:R44.
28. Townsend JP, Nielsen KM, Fisher DS, Hartl DL: **Horizontal acquisition of divergent chromosomal DNA in bacteria: effects of mutator phenotypes.** *Genetics* 2003, **164**:13-21.
29. Heinemann JA, Sprague GF Jr: **Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast.** *Nature* 1989, **340**:205-209.
30. Lan R, Reeves PR: **When does a clone deserve a name? A perspective in bacterial species based on population genetics.** *Trends Microbiol* 2001, **9**:419-424.
31. Wertz JE, Goldstone C, Gordon DM, Riley MA: **A molecular phylogeny of enteric bacteria and implications for a bacterial species concept.** *J Evol Biol* 2003, **16**:1236-1248.
32. Legault BA, Lopez-Lopez A, Alba-Casado JC, Doolittle WF, Bolhuis H, Rodriguez-Valera F, Papke TR: **Environmental genomics of "*Haloquadratum walsbyi*" in a saltern crystallizer indicates a large pool of accessory genes in an otherwise coherent species.** *BMC Genomics* 2006, **7**:171.