

Speciation without Species: A Final Word

W. Ford Doolittle*

Here I present the several models currently popular for understanding speciation in prokaryotes, in particular bacteria. I will argue that “speciation”, as a process or collection of independent but interacting processes sometimes serving to form genotypic and/or phenotypic clusters, can be studied effectively without any definition of “species” or any requirement that all prokaryotic lineages match such a definition. This has always been so, but formal acknowledgement would have a freeing effect.

Keywords

speciation • bacterial species • ecotype • Biological Species Concept • gene transfer

Part of the special issue Species in the Age of Discordance, guest-edited by Matthew H. Haber and Daniel J. Molter.

1 Introduction

This paper, like many before it, aims to solve the “species problem” by declaring it a non-problem. It borrows its title from an earlier article by Jeff Lawrence and its philosophical concepts from Marc Ereshefsky (eliminative pluralism), John Dupré (process ontology), Peter Godfrey-Smith (the use of cubes and the liberating realization that not all cases must be paradigms), Ken Waters (how philosophers should focus on what it is scientists do rather than what should be scientists’ ontology), and Jody Hey (the true nature of the “species problem”). The emphasis is on bacteria, but my pragmatic species anti-realist conclusion may be a general one.

Bacterial systematists long ago agreed that there are immediate and practical (taxonomic, clinical, and epidemiological) needs to designate particular genotypic or phenotypic clusters as “species” (Gevers et al. 2005; Doolittle and Zhaxybayeva 2009). I do not deny these needs for a “species definition”, only that there is some obviously best and universal way to formulate one,

*Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada, ford@dal.ca

or that there is some single evolutionary process or collection of processes (a “species concept”) ensuring that each and every bacterium will belong to a single such (or any) “species”.

Lawrence’s paper “Gene transfer in bacteria: speciation without species” (Lawrence 2002) highlighted one conundrum in bacterial classification. Bacteria (and archaea) engage in both homologous recombination (HR) with close relatives and DNA exchange with much more distant lineages. The former requires substantial DNA sequence similarity between recombining regions of a gene or genome while the later, called lateral gene transfer (LGT), can involve “illegitimate” recombinational processes requiring no sequence similarity, and resulting in the introduction of completely novel material (“new genes”).

What Lawrence noted was that, because of LGT, part of a bacterial genome may be similar enough in sequence to a stretch of DNA in the otherwise phylogenetically distant lineage that was its donor as to engage in frequent HR with it, while the rest may not. The notion that Mayr’s Biological Species Concept (BSC; Mayr 1975, 1996) can be reinterpreted for bacteria by substituting HR for “interbreeding” was already popular then, and is now even more so. So Lawrence reasoned that it would be possible for a bacterium to belong to two quite distantly related BSC-like “species” at the same time, depending on which parts of its chromosome were under examination. In the same vein, if during within-species lineage divergence some chromosomal regions continue to recombine while others (for whatever reason) do not, “speciation” will be completed in part of a bacterial chromosome and scarcely begun in another. Timescales for such temporally fragmented speciation could be immense: the divergence of *Escherichia coli* and *Salmonella enterica* apparently began some 70 million years ago—before the appearance of the first primate!—and is still not complete (Retchless and Lawrence 2007).

Such puzzles are consequences not only of a propensity to exchange genetic material with phylogenetically distant taxa, but of the fact that bacteria, unlike the animals Mayr had in mind when formulating the BSC, do not have to recombine (or exchange DNA at all) in order to reproduce, and when they do, only a small and often functionally biased fraction of a genome is involved. Although many bacteria exchange DNA much more frequently than we might have imagined a few decades ago, they still *reproduce* “asexually”, mostly by fission. Moreover, within-species rates of HR vary over three orders of magnitude among designated “species” (Vos and Didelot 2009; Bobay and Ochman 2017). The BSC cannot reasonably apply to all of them. So, what might?

2 The Ecotype Model

Three speciation mechanisms could be considered, and can work together, but also establish a base for a microbial species pluralism, of an eliminativist sort (Ereshefsky 1994). The first has long been promoted by Fred Cohan (2001) as the “ecotype model”. It involves what experimental bacterial geneticists call “periodic selection” or “clonal sweeps” (Atwood et al. 1951).

Suppose that a fitter-type mutant (one which more efficiently uses the supplied substrate, perhaps) arises in a well-isolated niche containing a pure, asexually-reproducing bacterial culture initially established from a single cell. A useful model system would be a chemostat, in which a constant density of actively growing cells is maintained by continuous introduction of fresh growth medium balanced by release of an equivalent volume of culture. The fitter mutant lineage will take over the entire chemostat and, with no sex and no recombination, fixation of the mutation responsible would entail fixation of the entire chromosome in which it occurred. Any diversity that has accumulated at loci other than that of the selected mutation will be purged. This purging of course includes phylogenetic marker loci, such as the 16S ribosomal RNA (16S rRNA) gene. Subsequently, neutral mutations can produce new sequence diversity at this and

other sites: how extensive such local “microdiversity” is will depend on the time between selective sweeps.

Meanwhile, in neighboring chemostats started with the same culture, different fitter-type mutations and whatever alleles happened to be present on the chromosomes on which they occurred will have been fixed. So sequences in the two chemostats will show *cohesive divergence*—similarity within and dissimilarity between at all loci, not just those under selection. Cohesive divergence is a signal and arguably the definition of “speciation”, taken generally.

There is an obvious null model to which this might be compared (Acinas et al. 2004; Giovannoni 2004; Straub and Zhaxybayeva 2017). Suppose bacterial evolution proceeded as simply as possible, by neutral sequence divergence alone (no sex or recombination), lineages being added by branching (“speciation”) and removed (“extinction”) at an average constant and equal rate over time. If we measure global diversity by sequencing all instances of some universal marker gene (say, again, that for 16S rRNA),¹ the expectation is that at any given time there *will* be clusters, and clusters of clusters, because extinction makes gaps. But a plot of the number of taxa per cluster as a function of the degree of sequence divergence defining the cluster should be smooth (figure 1). There should be no inflections corresponding to taxonomic ranks, including “species”, if these are merely arbitrary, investigator-imposed conventions. Nothing will constrain divergence, birth and death to make some degree of divergence “special”. But if the ecotype model, entailing constraints on population and niche sizes, does hold, and there is some more-or-less restricted average value of these, we expect such an inflection. There will be microdiversity—an excess of very closely related sequences—all but one of which will disappear in the next clonal sweep.

That such a conclusion (non-random “speciation”, possibly as ecotypes) might be appropriate for *some* designated bacterial species but not for others was recently shown by Straub and Zhaxybayeva (2017), who recommend their birth–death model as the null against which claims for the operation of constraining genetic or ecological speciation processes should be tested. And even when significant microdiversity *is* observed, the constraint could be artifactual, resulting from failure to sample intermediate lineages present in low abundance at the targeted site, or lacking some single gene whose phenotypic expression (pathogenesis or antibiotic resistance for instance) actually directed sampling.²

¹Note that 97% 16S rRNA sequence similarity is widely accepted as defining “operational taxonomic units” and to be necessary if not sufficient to identify species. But the taxon-relative and *ad hoc* nature of this practice is widely acknowledged. As more complete genome sequences come online, average nucleotide identity (ANI) of shared orthologs will likely replace it, with a “species” cut off of around 95–96% ANI (Richter and Rosselló-Móra 2009). This may still be arbitrary, but does entail sampling of much more of each genome.

²From a very recent ANI-based analysis of all possible pairs of 90,000 prokaryotic genomes Jain et al. (2017) conclude that 99.8% of pairs show > 95% or < 83% inter-species ANI, with a clear gap between, suggesting that the former “represents an accurate threshold for demarcating almost all currently named prokaryotic species”. They suggest that the “biological mechanisms underlying this genetic discontinuity are not clear but should be a subject for future research.” My point in this essay is not that *no* tightly-clustered bacterial “species” exist but that not all bacterial lineages need belong to one. Still, 99.8% is an overwhelming preponderance, and does not square with the claim of Shapiro et al. (2017, see below) that “More impressive than the number of species is the number of intermediate forms ...”, nor with the general argument advanced here. Included in the massive data set of Jain et al. may be many genome assemblies represented precisely *because* they are closely related, a source of considerable bias, and a possible explanation for our disagreement. That an overwhelming proportion of prokaryotic species should be so narrowly constrained in its divergence and that there should be so few “intermediate” forms seems inconsistent with the operation of the processes described here. One possibility is that such clusters comprise clonal ecotypes emerging through selection within much more extensive (ANIs much lower than 95%) BSC-type species, most such lower-similarity “con-specifics” being present at much lower abundance, or in different locations, and thus not sampled. Boundaries to such BSC-like species could be established by barriers to gene flow such as restriction-modifications systems or narrow bacteriophage host ranges, but would be variable in effectiveness and

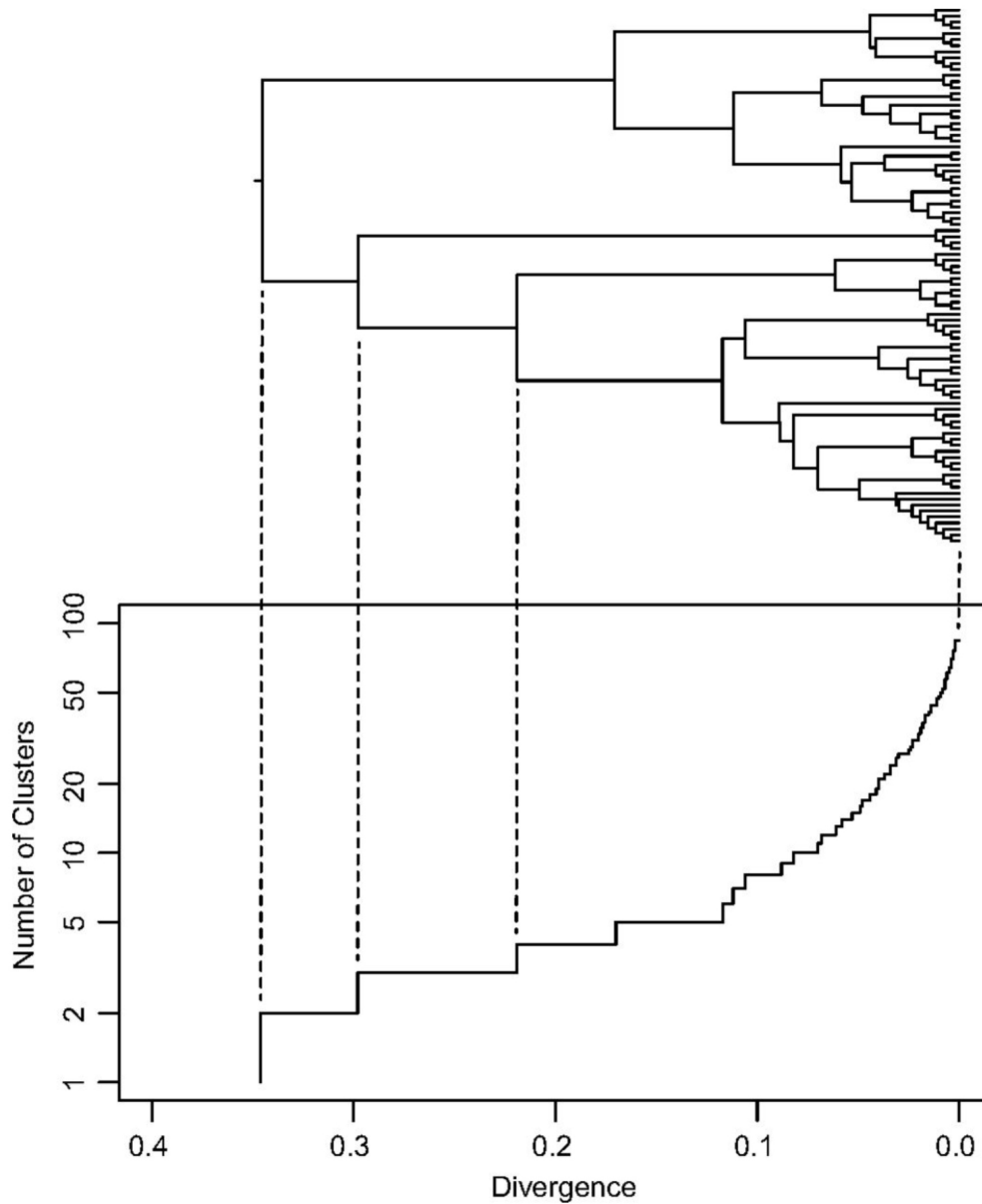


Figure 1: The relationship between clustering and divergence in the random birth-death model for lineage diversification of Straub and Zhaxybayeva (2017). The upper part of the figure is a “phylogeny” of surviving lineages generated by the model, with lineages that have gone extinct (“died”) not being shown. The lower part plots the number of clusters observed at any level of sequence divergence (this assumed to be constant with time). Unless some biological process makes species “special”, such plots (averaged over sufficient data) should be a smooth curve. Reprinted with permission from Straub and Zhaxybayeva (2017).

Evidence in nature for clean and clear ecotype evolution in real time has been hard to find (Cordero and Polz 2014). A recent exhaustive study by Bendall et al. (2016) is thought to provide the most persuasive example of such a clonal sweep. These investigators, after a nine-years of sequencing metagenomic DNA samples from a lake in Wisconsin, concluded that in several microbial populations, and in particular that of a species of *Chlorobium* (a photolithotrophic green sulfur bacterium), there was genome-wide purging of diversity due to selection at some (unidentified) chromosomal locus.

3 Biological Species Concept for Bacteria

The second process that can produce cohesive divergence is HR. This, as already mentioned, is the basis of claims that some bacteria conform to something like Mayr's BSC, claims that date back to the 1991 discovery by Dykhuizen and Green (1991) of frequent recombination in the most popular model bacterium, *Escherichia coli*. This theory too entails that all alleles at a given locus will, with their accumulated mutations, ultimately be traceable to a single (randomly chosen) ancestral allele in an earlier population—but likely a much earlier ancestor if there is no selection. This is a simple application of coalescent theory (Kingman 1982).

Given frequent within-population recombination, however, different loci will trace to different ancestral alleles present in different genomes at different times, so within-species phylogenetic trees derived from the sequences of different genes will be different: not what is observed in the ecotype model. But in multiple diverging (speciating) subpopulations between which gene exchange has been restricted longer than these coalescent times, trees for different genes will all have the same topology (figure 2). So such contrasting patterns (rather than microdiversity) is the hallmark of cohesive divergence in HR-dominated bacterial evolution and, as Dykhuizen and Green asserted, “an operational criterion for the defining of bacterial species.”

The required restriction of exchange could be by a physical barrier or incompatibilities of the agents of DNA exchange such as bacteriophages. Many bacteria and archaea do seem to show cohesive divergence implemented in this way: from a recent survey of 105 designated prokaryotic species based at least indirectly on Dykhuizen and Green's principle, Bobay and Ochman (2017) find that “only a minor fraction (< 15%) undergoes too little [within-species] gene flow to be assigned to species based in the BSC.”

It is possible to imagine that, if rates at which recombination falls off with sequence divergence and rates at which divergence actually happens are just right, clustering into BSC-type species will occur spontaneously, without selection on any locus. Any microdiversity would be a consequence of this selection-independent process. But the current consensus is that unreasonable rates of such “falling off” are required for this, and so in general selection for occupancy of a new niche and/or some reduction in inter-population HR by geographic separation or barriers to transfer seems to be required (Fraser et al. 2007; Cordero and Polz 2014; Shapiro et al. 2017).

Very often it is likely a mixture of ecotype and BSC-like processes that separate populations. Shapiro and Polz (2015) present an appealing five-step speciation model in which “speciation” is initiated by an advantageous selectable mutation or (as below) LGT of a gene allowing occupancy of a new niche (use of a previously ignored substrate for instance) and proceeds differently depending on the lineage's propensity to engage in HR. If this is high, then the acquired gene will spread though the population while leaving diversity at other loci intact. If low, then periodic selection (“clonal expansion” as in the ecotype model) will initially dominate, purging such diversity. But, as they note, “given enough time before any further selective events, and assuming

unstable over time (Oliviera et al. 2016). Because bacteria do not have to exchange genetic information in order to reproduce, reinforcement mechanism would likely not be selected for (for that effect) (Servedio and Noor 2003).

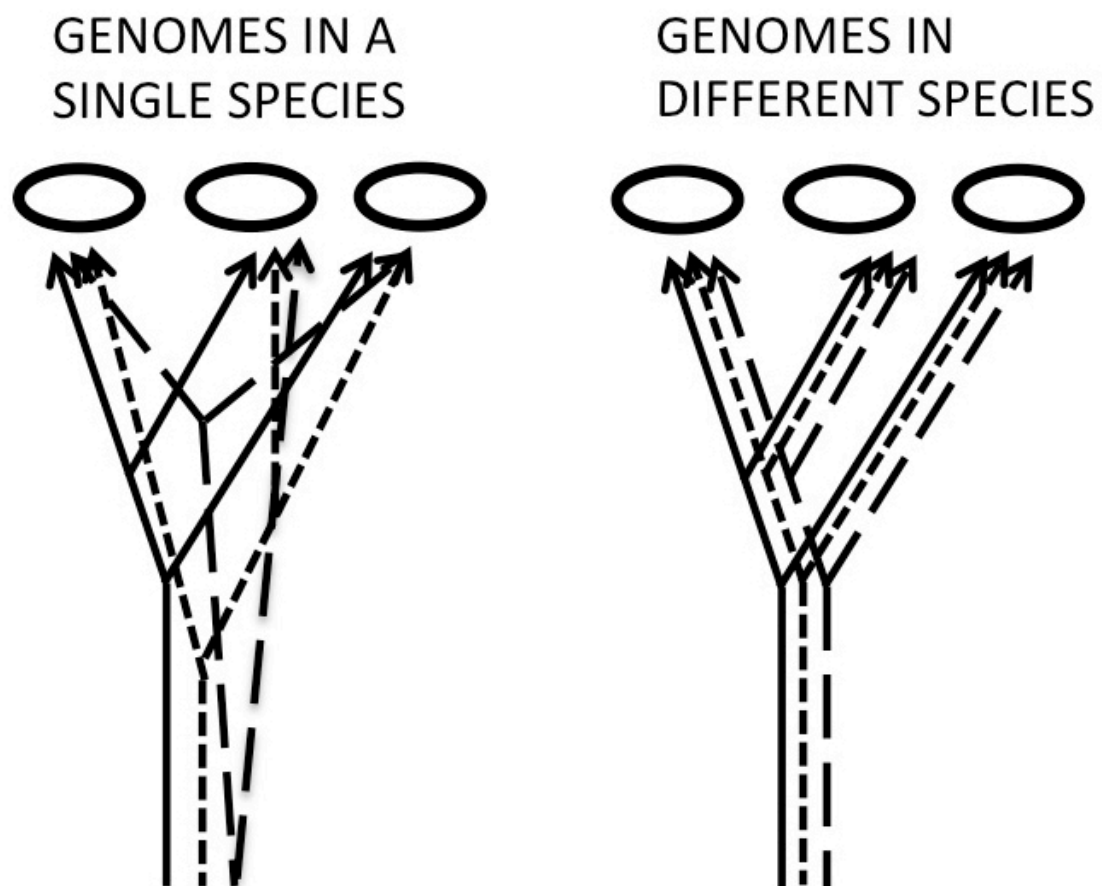


Figure 2: The Biological Species Concept for bacteria. Given sufficient within-species recombination, trees based on the sequences of different genes in the same genome should have different topologies (left), But for different species (not exchanging genetic information) they should be the same. Solid, dashed and dotted lines represent different genes.

that the two [substrate-defined] niches remain sympatric, neutral alleles will eventually become randomly distributed across genotypes, with only adaptive alleles being selectively maintained.”

4 Lateral Gene Transfer and Phenotypic Speciation

There is as well a third process, contributing to but arguably also disruptive of both the above, and manifested in the phenomenon of species’ “pangenomes”. As of two years ago, some 2,000 strains of *E. coli* had been sequenced. With an average of 5,000 genes each, these showed a “core” of around 3,000 genes shared by all (or almost all) and a “pangenome” of about 90,000 present only in some—usually a minority and often only one strain (Land et al. 2015). Currently the *E. coli* pangenome stands at about 120,000 (D. W. Ussery, personal communication) so this species harbors (so far) about four times as many genes as we humans can boast, just not all in the same cell at the same time. This makes it problematic to say that the genome of *H. sapiens* is more complex than that of *E. coli*, or even to be clear on what the latter *is*. It depends on whether we want to compare individuals or “species”. And many designated bacterial species will have pangenomes as large as *E. coli*’s.

Pangenomic strain-to-strain variation reflects the (more or less balanced) gain of genes by LGT and loss by deletion (Puigbo et al. 2014). Some of the gains are selfish elements (phages or plasmids) and some are probably deleterious and/or irrelevant to fitness, but some enable different strains or sub-specific lineages of a species such as *E. coli* to occupy different niches and use different resources (see for instance Luo et al. 2011). Ecotype and HR-based species definitions tend to focus on the shared core of genes and measures of their similarity and relationships (for instance in developing a phylogeny of lineages), while relegating LGT-derived genes to an important but quantitatively minor role, principally in initiating the ecological diversification that can lead to speciation by either model. So it is easy to overlook the great variation in gene content manifested by a single designated bacterial species. The average strain-to-strain identity in gene content for designated bacterial species is around 80%, with some intra-specific comparisons as low as 60%, in the range generally observed for genera (Varghese et al. 2015; Tu and Lin 2016). Bacterial groups designated (for whatever reason) as “orders” show 50–60% similarities in gene content.

Efforts to use gene-content variation rather than sequence similarity to circumscribe microbial taxa are still in their infancy (Nowell et al. 2014; Tu and Lin 2016). But one can imagine an ecological species definition based on such data, and suggest that gene content might be more reliably reflected in the sorts of phenotypic tests traditionally used to differentiate strains of a species than is 16S rRNA similarity, or even ANI1.

5 Eliminative Pluralism for Bacteria

Ereshefsky’s widely-cited 1992 paper “Eliminative Pluralism” is eliminative in that it argues that the category “species” has no single defining property and cannot play the unambiguous role assigned to it in traditional taxonomy. It is pluralist in that it recognizes the value and validity of phylogenetic, interbreeding, and ecological species definitions, roughly translatable into microbial terms according to the three processes described above (Doolittle and Zhaxybayeva 2009). The associated methods of circumscribing “species” carve the world up somewhat differently, so that the selfsame organisms might find themselves in a single species under one concept and two or more different species under another—and sometimes perhaps in no species by any. This problem is thus even more complex than that identified by Lawrence.

Indeed Ereshefsky's pluralism is also more radical than one that holds that different criteria are appropriate for different groups but retains the requirement that any one individual belongs to only one and the same "species", however defined. Importantly, it empties out the bathwater ("species" as a category, class, or natural kind) while saving the babies (individual cohesive clusters of organisms designated "species" *for one reason or another*). As Ereshefsky (1998) notes:

Nothing I have said casts doubt on the existence of those taxa we call 'species'. We remain confident that there are such taxa as *Homo sapiens* and *Canis familiaris*. Of course, it might be odd to call them 'species' in light of the heterogeneity argument The important point here is that the nonexistence of the species category does not imply that the taxa we call 'species' are mere artifacts.

As argued earlier (Doolittle and Zhaxybayeva 2009), such a view holds with equal force for microbes (bacteria and archaea, also viruses). Defining bacterial species by microdiversity (as in the ecotype model), by gene flow (as in the BSC) or gene content (taking pangenomes into account) can distribute the same organisms differently between taxa, and there is no single set of criteria by which the bacterial species category can be best defined—no justification for a bacterial species monism. Microbiologists might by negotiation agree on a degree of genotypic or phenotypic cohesion within (and divergence between) clusters necessary for the identification of species, but no single mechanism for achieving this, no single "species concept" (Roselló-Mora and Amann 2001) would be implied.

That said, we might allow the existence of a heterogeneous collection of *processes* that we could call *biological speciation*. This collection would encompass the above mechanisms as separately or together a complex way of effecting the cohesive divergence that is what most investigators mean by "speciation". This would sometimes result in the formation of taxa every bit as real as *Homo sapiens* or *Canis familiaris*, but sometimes it would not: not all microbes would have to belong to such species. We could have *speciation without species*. We might consider cohesive divergence as the essential property of speciation as a process without implying that there is any single underlying biological mechanism, or (I hope) being obliged to consider whether this move makes of speciation a processural "natural kind" (Dupré and Nicholson 2018)

6 Species as Processes

Dupré (2017) argues persuasively that for many purposes a *process ontology* is preferable to a *substance (or "thing") ontology*. For most of us, this may be easiest to understand in terms of our continued existences as individual organisms. We are clearly more process-like than thing-like: few of our cells are "the same" as those with which we were born. Instead, our identity is sustained over a life-time, in spite of turnover at the cellular and molecular level, by the continuity and integrated character of developmental and regenerative processes. Similarly, Dupré claims, sexual biological species are processes—comprising the recurring acts of reproduction that define the maximally inclusive inter-breeding populations foundational to Mayr's BSC. It is these that stabilize a species over time, not the fleeting lives of individual organisms. As processes, sexual species are also stabilized by abiotic and biotic parameters of their ecological niches and niche construction—the establishment by a species of the conditions under which it continues to evolve.

Dupré integrates this *processual* view with the widespread acceptance of the realization of Michael Ghiselin (1974) and David Hull (1978), that individual species (like *H. sapiens* or *C. familiaris*, whose reality Ereshefsky does not question) are also not categories or classes, but individuals. Moreover, Dupré writes:

I suggest that the invention or emergence of sex is also the emergence of species as individuals. Without sex there are no horizontal relations between the members of a species and they are connected only by their ancestry. But unless every individual, or at least every individual with a minimal novelty (e.g. a point mutation), is the ancestor of a new species there must be some horizontal connections that establish a group of individuals as an appropriate set of ancestors to found a species, and we appear to be launched on an infinite regress. If there existed species-like processes prior to sexual reproduction, these lacked any coherence or integration that could qualify them as processual individuals with persistence as such through time. This proposal also puts Mayr's familiar biological species concept in a slightly different light. Reproductive connections are indeed fundamental to the existence of species *as individuals*.

I'd be a bit more generous here. Some bacteria do recombine and/or exchange genes by LGT a lot. For these, Dykhuizen and Green's criteria may be met and a marginal processual individuality seems not unreasonable. Recall, however, that bacteria do not require mating or recombination and this last, when it occurs, affects only a fraction of a genome. So claims that "bacteria have sex" are merely metaphorical (Franklin 2007), and some bacteria exchange genes very rarely. Indeed, though Bobay and Ochman (2017) title a recent paper "Biological species are universal across life's domains", they admit that "20–30% of plants self-fertilize or reproduce vegetatively and are not suited for classification by the BSC" and that as many as 15% of designated bacterial species may be similarly unsuitable.³

Thus Dupré's criteria of "coherence and integration" are not so clearly definable and clearly do come in degrees. Something like Cohan's ecotype model also must to some variable extent apply to all bacteria, just as ecological boundaries do constrain (and thus "stabilize") the evolution of recognized asexual animals and plants (Van Valen 1976). Even organisms that never exchange genetic information are integrated by competition for resources within their shared environment. Godfrey-Smith's (2009) promotion of the spectrum between marginal and paradigm case can help us here (see next section). Indeed, it is the spectrum-like character of individual designated species' speciation processes and their degree of completion that must be taken into account. Shapiro et al. (2017) write:

More impressive than the number of species is the number of intermediate forms—Darwin's "doubtful cases"—suggesting that speciation is a continuous process that happens all the time. This apparent fluidity has led us and others to propose that most organisms can probably be placed somewhere along a "spectrum" of speciation. Of course, speciation may not happen at all, or at least not go to completion.

7 Paradigms and Marginal Cases

Godfrey-Smith's (2009) use of three-dimensional representations of entities undergoing evolution by natural selection (ENS) introduced some welcome slack in our understanding of the process. For instance, in analyses of reproduction-related dimensions he plots degrees of reproductive specialization (germ/soma distinction), level of functional integration, and the extent to which progeny are produced through a bottleneck (single cells, for instance) as the three axes of

³Perhaps for these authors "universal" means "found within", not "applies to all", an important distinction for my final conclusion.

a cube. Paradigm individuals (we humans for instance) would have high values of all three variables and be located at one corner of the cube, while “marginal” cases of individuals (a buffalo herd) would have minimal values and map to the opposite corner.

Such representations oversimplify biology, of course: an *E. coli* cell seems even more clearly to be a Darwinian individual than I am, but has no germ/soma distinction. Figure 3 is an attempt to place a few designated species in a cube whose three dimensions correspond to the three processes discussed above. Our own species, with obligatory HR, minimal if any LGT (Salzburg 2017) and (because of the HR and population structure, clonal sweeps of limited scope) is at the lower back corner, a paradigm of Mayr’s BSC. *Mycobacterium tuberculosis* on the other hand has low rates of HR and its adaptation to a succession of mammalian hosts appears to have been by a succession of selective sweeps, a paradigm for an ecotype species (Smith et al. 2006; Ward et al. 2008). *Helicobacter pylori*, common in the human gut, seems to evolve largely by HR (Budendorfer et al. 2016). Many designated bacterial species will occupy intermediate positions and very loosely circumscribed positions (fuzzy circles).

Representational cubes aside, the notion that some lineages or entities may conform only marginally to ENS algorithms and yet still be subject to it has especially strong appeal when re-purposed to deal with the bacterial species problem. As Dennett (2011) writes, Godfrey-Smith “shows us how to think about all the many marginal cases—the semi-Darwinian processes, proto-Darwinian phenomena, quasi-Darwinian relationships—without marginalizing them”.

If we take genotypic or phenotypic clustering to some agreed-upon extent to be necessary and sufficient for “species” status—regardless of the underlying genetic or ecological process—we are one third of the way to resolving the bacterial “species” problem. We can go the next third by admitting that *there is no fact of the matter* as to how tight such clustering must be, but whatever criterion we set will surely leave some cases as marginal (by Bobay and Ochman’s criteria as many as 15%)—or out altogether. Such non-recombining lineages are of course subject to some ecotype evolutionary processes because they do occupy some *particular* niche. But there is no *a priori* reason why this needs to cluster them so tightly as to conform to any agreed-upon definition of “species”. So one advantage of thinking of species as a collection of overlapping mechanisms is that it does not always have to work.

Speciation itself, taken as a collection of processes, occupies a place similar to that of the “species” category in Ereshefsky’s eliminative pluralism. It is useful term to describe clustering mechanisms and their potential outcomes in many particular lineages, and in some lineages it may have sufficient coherence and long term stability to produce (or be) a processual individual ‘a la Dupré. But there is no single biological mechanism that is speciation, which we here consider to be any and all genetic and ecological constraints conferring cohesive divergence. It is like “erosion” or “inflation” as processes. Dupré writes:

When I refer to a process I shall, henceforth unless otherwise stated or obvious, mean an individual process, a process with the sort of coherence and persistence that might suggest treating it as a thing. Organisms, on my view, are paradigms of such coherent individual processes, though less controversially processual entities such as storms or rivers also have good claims to be individuals. Some processes—erosion, inflation, evolution—lack any such coherence.

8 Speciation without Species

The title of Kenneth Waters’s recent “Ask not ‘What is an Individual?’” problematizes the “is”. I think microbiologists are now implicitly making the same ontology-avoiding move and that

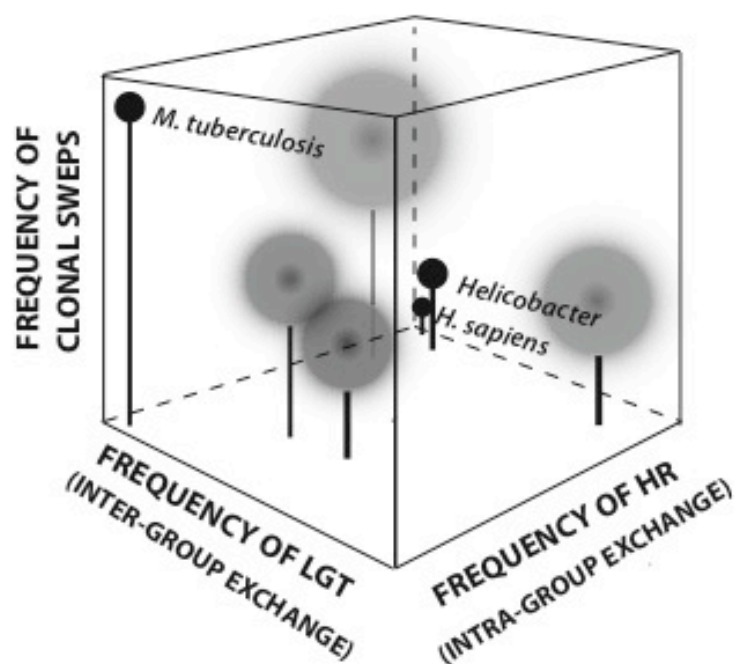


Figure 3: A three-dimensional representation of speciation. Evolution of human genomes is dominated by homologous recombination (HR), with little if any lateral gene transfer (LGT) and few if any population-wide clonal sweeps. *H. sapiens* is a paradigm “species” according to the Biological Species Concept. See text for discussion of *M. tuberculosis* and *Helicobacter (pylori)* information. Fuzzy circles are meant to represent typical bacterial species and the ranges of their behaviors.

philosophers need to pick up on this: there has been too much focus on biologists' apparent conceptual commitments and not enough on what they actually do. Thus there should be a more deliberate distancing from concerns about what "species" actually are and a refocusing on how the various genetic and ecological processes that might create cohesively divergent clusters are investigated. Sometimes such clusters will be sufficiently tight that microbiologists might for some purely practical reason decide to call them "species", but this is almost beside the point, a motivation and a tool but not the actual endpoint. The passage from Shapiro et al. (2017) quoted above concludes "Here, we are less concerned with the number and exact definition of species and more with why speciation happens (or not) and the nature of the speciation process". Earlier, Shapiro and Polz (2015) demurred: "We wish to make a strong distinction between this process of speciation—which we define as any stage of the dynamic process of ecological and genetic differentiation—and the concept of species, which we are not attempting to address." Similarly, Lasalle et al. (2015) write "the real challenge is to understand the underlying mechanisms that led to the emergence of this structure of bacterial diversity. This is intimately linked to the current inability of evolutionary microbiologists to find a unifying model of prokaryotic cladogenesis."

Waters's philosophical pragmatism leads him to conclude, in the case of biological "individuals" (with gene and holobionts as examples) that we can draw a three-part lesson. First, "The processes of evolution are extremely complex, and the results of these complex processes have not yielded a neat parsing of life into individuals": we could say the same for "species" (after all, a kind of individual). Second, "Concepts should be conceived as tools, used to achieve particular purposes": the same can be said for the role of any species concept in studies that seek to understand how ecological and genetic differentiation occur, and might be coupled. Third, Waters suggests that "instead of asking 'what is a biological individual?' philosophers should ask 'what ways of conceiving biological individuality could be useful?'" The same can be said for species, when these are the individuals in question. The task for philosophers of (micro) biology should similarly be to ask what microbiologists who recognize that species designations are arbitrary though invaluable in research and clinical practice—or who study clustering processes under the banner of "speciation" without feeling obliged to define "species"—are actually doing.

The key "is" question about bacterial "species" is, I think, "Are there forces compelling bacterial lineages (a parent bacterium and its progeny and theirs, and so on) to conform to whatever degree of genotypic and phenotypic clustering are required by our agreed-upon species definition?". Again, such a definition need not claim to be diagnostic of any natural kind in order to be of use, and I'm guessing that the answer to that question is, in any case, "No". This would take us the final third of the way toward resolving the bacterial "species problem".

There are of course relevant forces. As Dupré puts it:

To the extent that there are strong divisions between kinds, this is likely to be because natural selection favours and disfavors particular areas of morphospace. Put differently, the combinations of traits that satisfy the conditions of existence [Darwin's term] occupy discontinuous regions of trait space. (This does rather oversimplify the matter, as the conditions of existence depend on what other organisms concurrently exist.)

What is not obvious is that such forces, or any of the processes discussed here, will compel *all* lineages to cluster as tightly as we might collectively require for any useful species definition. There might be bacteria that cannot be placed in species, by any reasonable criteria, these being the result of negotiation, not discovery. Nor is it obvious that such forces impinge differently (more strongly) on what we chose to call species compared to what we chose to call strains, or

genera, or families. If there are lineages that can't be put in species and if there are not forces that impinge especially strongly only at that rank, then arbitrarily negotiated species definitions cannot underwrite or be underwritten by any single or single combination of clustering processes in any principled or uniform way.

Jody Hey rightly notes that (mostly in the case of eukaryotes) it is the attempt to see species as both evolutionary and taxonomic units that *is* “the species problem”:

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.

For bacteria, not only do the motivations conflict, but also neither lends itself to any sort of monistic realism. We microbiologists can fruitfully study cohesive divergence without committing ourselves to the belief, a vestige of pre-Darwinian “typological thinking”, that there is single underlying process or that there are uniformly definable entities called species whose existence we must thereby explain (Doolittle 2009).

Acknowledgments

I thank Olga Zhaxybayeva and Jess Shapiro for valuable comments, and the Natural Science and Engineering Research Council (grant GLDSU447989) for support.

Literature cited

- Acinas, Silvia G., Vanja Klepac-Ceraj, Dana E. Hunt, Chanathip Pharino, Ivica Ceraj, Daniel L. Distel, and Martin F. Polz. 2004. “Fine-scale Phylogenetic Architecture of a Complex Bacterial Community.” *Nature* 430 (6999): 551–554.
- Atwood, Kimball C., Lillian K. Schneider, and Francis J. Ryan. 1951. “Periodic Selection in *Escherichia coli*.” *Proceedings of the National Academy of Sciences of the United States of America* 37 (3): 146–155.
- Bendall, Matthew L., Sarah L.R. Stevens, Leong-Keat Chan, Stephanie Malfatti, Patrick Schwien-tek, Julien Tremblay, Wendy Schackwitz, Joel Martin, Amrita Pati, Brian Bushnell, Jeff Froula, Dongwan Kang, Susannah G. Tringe, Stefan Bertilsson, Mary A. Moran, Ashley Shade, Ryan J. Newton, Katherine D. McMahon, and Rex R. Malmstrom. 2016. “Genome-wide Selective Sweeps and Gene Specific Sweeps in Natural Bacterial Populations.” *The ISME Journal* 10 (7): 1589–1601.
- Bobay, Louis-Marie, and Howard Ochman. 2017. “Biological Species are Universal Across Life’s Domains.” *Genome Biology and Evolution* 9 (3): 491–501.
- Bubendorfer, Sebastian, Juliane Krebes, Ines Yang, Elias Hage, Thomas F. Schulz, Christelle Bahlawane, Xavier Didelot, and Sebastian Suerbaum. 2016. “Genome-wide Analysis of Chromosomal Import Patterns After Natural Transformation of *Helicobacter pylori*.” *Nature Communications* 7: 11995.
- Cohan, Frederick M. 2001. “Bacterial Species and Speciation.” *Systematic Biology* 50 (4): 513–524.

- Cordero, Otto X. and Martin F. Polz. 2014. "Explaining Microbial Genomic Diversity in Light of Evolutionary Ecology." *Nature Reviews Microbiology* 12 (4): 263–273.
- Dennett, Daniel C. 2011. "Homunculi Rule: Reflections on Darwinian Populations and Natural Selection by Peter Godfrey-Smith." *Biology and Philosophy* 26: 475–488.
- Doolittle, W. Ford. 2009. "Eradicating Typological Thinking in Prokaryotic Systematics and Evolution." *Cold Spring Harbor Symposia on Quantitative Biology* 74: 197–204.
- Doolittle, W. Ford and Olga Zhaxybayeva. 2009. "On the Origin of Prokaryotic Species." *Genome Research* 19: 744–756.
- Dupré, John. 2017. "The Metaphysics of Evolution." *Interface Focus* 7 (5): 20160148.
- Dupré, John, and Daniel J. Nicholson. 2018. "A Manifesto for a Processural Philosophy of Biology." In *Everything Flows: Towards a Processural Philosophy of Biology*, edited by D. J. Nicholson and J. Dupré. Oxford: Oxford University Press.
- Dykhuisen, D. E, and L. Green. 1991. "Recombination in *Escherichia coli* and the Definition of Biological Species." *Journal of Bacteriology* 173 (22): 7257–7268.
- Ereshefsky, Marc. 1992. "Eliminative Pluralism." *Philosophy of Science* 59 (4): 671–690.
- Ereshefsky, Marc. 1998. "Species Pluralism and Anti-realism." *Philosophy of Science* 65 (1): 103–120.
- Franklin, L. R. 2007. "Bacteria, Sex, and Systematics." *Philosophy of Science* 74(1): 69–95.
- Fraser, Christophe, William P. Hanage, and Brian G. Spratt. 2007. "Recombination and the Nature of Bacterial Speciation." *Science* 315 (5811): 476–480.
- Gevers, Dirk, Frederick M. Cohan, Jeffrey G. Lawrence, Brian G. Spratt, Tom Coenye, Edward J. Feil, Erko Stackebrandt, Yves Van de Peer, Peter Vandamme, Fabiano L. Thompson, and Jean Swings. 2005. "Re-evaluating Prokaryotic Species." *Nature Reviews Microbiology* 3: 733–739.
- Ghiselin, Michael T. 1974. "A Radical Solution to the Species Problem." *Systematic Zoology* 23: 536–544.
- Giovannoni, Stephen. 2004. "Oceans of Bacteria." *Nature* 430: 515–516.
- Godfrey-Smith, Peter. 2009. *Darwinian Populations and Natural Selection*. Oxford: Oxford University Press.
- Hey, Jody. 2001. "The Mind of the Species Problem." *Trends in Ecology and Evolution* 16 (7): 326–329.
- Hull, David L. 1978. "A Matter of Individuality." *Philosophy of Science* 45 (3): 335–360.
- Jain, Chirag, Luis M. Rodriguez-R., Adam M. Philippy, Konstantinos T. Konstantinidis, and Srinivas Aluru. 2017. "High-throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries." *bioRxiv*. doi:10.1101/225342
- Kingman, J. F. C. 1982. "On the Genealogy of Large Populations." *Journal of Applied Probability* 19: 27–43.
- Land, Miriam, Loren Hauser, Se-Ran Jun, Intawat Nookaew, Michael R. Leuze, Tae-Hyuk Ahn, Tatiana Karpinets, Ole Lund, Guruprasad Kora, Trudy Wassenaar, Suresh Poudel, and David W. Ussery. 2015. "Insights from 20 Years of Bacterial Genome Sequencing." *Functional and Integrative Genomics* 15 (2): 141–161.
- Lasalle, Florent, Daniel Muller, and Xavier Nesme. 2015. "Ecological Speciation in Bacteria: Reverse Ecology Approaches Reveal the Adaptive Part of Bacterial Cladogenesis." *Research in Microbiology* 166 (10): 729–741.

- Lawrence, Jeffrey G. 2002. "Gene Transfer in Bacteria: Speciation without Species?" *Theoretical Population Biology* 61: 449–460.
- Luo, Chengwei, Seth T. Walk, David M. Gordon, Michael Feldgarden, James M. Tiedje, and Konstantinos T. Konstantinidis. 2011. "Genome Sequencing of Environmental *Escherichia Coli* Expands Understanding of the Ecology and Speciation of the Model Bacterial Species." *Proceedings of the National Academy of Sciences of the United States of America* 108: 7200–7205.
- Mayr, Ernst. 1975. *Evolution and the Diversity of Life*. Cambridge: Harvard University Press.
- Mayr, Ernst. 1996. "What is a Species and What is Not?" *Philosophy of Science* 63 (2): 262–277.
- Nowell, Reuben W., Sarah Green, Bridget E. Laue, and Paul M. Sharp. 2014. "The Extent of Genome Flux and its Role in the Differentiation of Bacterial Lineages." *Genome Biology and Evolution* 6 (6): 1514–1529.
- Oliveira, Pedro H., Marie Touchon, and Eduardo P. C. Rocha. 2016. "Regulation of Genetic Flux between Bacteria by Restriction-modification Systems." *Proceedings of the National Academy of Sciences of the United States of America* 113: 5658–5663.
- Puigbo, Pere, Alexander E. Lobkovsky, David M. Kristensen, Yuri I. Wolf, and Eugene V. Koonin. 2014. "Genomes in Turmoil: Quantification of Genome Dynamics in Prokaryotic Supergenomes." *BMC Biology* 12: 66.
- Retchless, Adam C., and Jeffrey G. Lawrence. 2007. "Temporal Fragmentation of Speciation in Bacteria." *Science* 317 (5841): 1093–1096.
- Richter, Michael, and Ramon Roselló-Móra. 2009. "Shifting the Genomic Gold Standard for the Prokaryotic Species Definition." *Proceedings of the National Academy of Sciences of the United States of America* 106 (45): 19126–19131.
- Roselló-Mora, Ramon, and Rudolf Amann. 2001. "The Species Concept for Prokaryotes." *FEMS Microbiology Reviews* 25 (1): 39–67.
- Salzburg, Steven L. 2017. "Horizontal Gene Transfer is not a Hallmark of the Human Genome." *Genome Biology* 18: 85.
- Servedio, Maria, and Mohamed Noor. 2003. "The Role of Reinforcement in Speciation: Theory and Data." *Annual Review of Ecology, Evolution, and Systematics* 34 (1): 339–364.
- Shapiro, B. Jesse, and Martin F. Polz. 2015. "Microbial Speciation." *Cold Spring Harbor Perspectives in Biology* 7 (10): a018143.
- Shapiro B. Jesse, Jean-Baptiste Leducq, and James Mallet. 2016. "What is Speciation?" *PLoS Genetics* 12 (3): e1005860.
- Smith, Noel H., Kristin Kremer, Jacqueline Inwald, James Dale, Jeffrey R. Driscoll, Stephen V. Gordon, Dick van Soolingen, R. Glyn Hewinson, and John Maynard Smith. 2009. "Ecotypes of the Mycobacterium tuberculosis Complex." *Journal of Theoretical Biology* 239 (2): 220–225.
- Straub, Timothy J., and Olga Zhaxybayeva. 2017. "A Null Model for Microbial Diversification." *Proceedings of the National Academy of Sciences of the United States of America* 114 (27): E5414–E5423.
- Tu, Qichao, and Lu Lin. 2016. "Gene Content Dissimilarity for Subclassification of Highly Similar Microbial Strains." *BMC Genomics* 17: 647.
- Van Valen, Leigh. 1976. "Ecological Species, Multispecies, and Oaks." *Taxon* 25 (2/3): 233–239.
- Varghese, Neha J., Supratim Mukherjee, Natalia Ivanova, Konstantinos T. Konstantinidis, Kostas Mavrommatis, Nikos C. Kyrpides, and Amrita Pati. 2015. "Microbial Species Delineation Using Whole Genome Sequences." *Nucleic Acids Research* 43 (14): 6761–6771.

- Vos, Michiel and Xavier Didelot. 2009. "A Comparison of Homologous Recombination Rates in Bacteria and Archaea." *The ISME Journal* 3 (2): 199–208.
- Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kuhl, M. and A. Grossman. 2008. "Genomics, Environmental Genomics and the Issue of Microbial Species." *Heredity* 100: 207–219.
- Waters, C. Kenneth. 2018. "Ask Not 'What is an Individual?'" In *Individuation, Process, and Scientific Practices*, edited by Otávio Bueno, Ruey-Lin Chen, and Melinda Bonnie Fagan. Oxford: Oxford University Press.
- Wernegreen, Jennifer J. 2017. "In it for the Long Haul: Evolutionary Consequences of Persistent Endosymbiosis." *Current Opinion in Genetics and Development* 47: 83–90.

© 2019 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits anyone to download, copy, distribute, display, or adapt the text without asking for permission, provided that the creator(s) are given full credit.

ISSN 2475-3025