Institut Pasteur

Research in Microbiology 166 (2015) 729–741

www.elsevier.com/locate/resmic

Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis

Florent Lassalle [a,*], Daniel Muller [b,c,d], Xavier Nesme [b,c,d,e]

[a] University College London, Gower Street, London, WC1E 6BT, United Kingdom
[b] Université de Lyon, F-69622, Lyon, France
[c] Université Lyon 1, Villeurbanne, France
[d] CNRS, UMR5557, Ecologie Microbienne, Villeurbanne, France
[e] INRA, USC1364, Ecologie Microbienne, Villeurbanne, France

Received 23 February 2015; accepted 30 June 2015
Available online 17 July 2015

Abstract

In this review, we synthesise current models and recent comparative genomic studies describing how bacterial species may emerge through adaptation to a new ecological niche and maintain themselves in the same niche over long time periods. We notably consider the impact of genetic exchange with phylogenetically close relatives living in sympatry and how this leads to the heterogeneous evolution of different genes within the bacterial genome. This heterogeneity provides landmarks to recognise genes that determine adaptation to the ecological niche, and we present reverse ecology strategies to unravel ecological properties of bacterial populations.

Keywords: Bacterial speciation; Ecological speciation; Ecological niche; Ecotype; Core genome; Core niche

1. Introduction

The notion of species in prokaryotes is intensely debated [1]. Currently, bacterial species are simply defined by their genomic homogeneity, leading to the conventional division of the bacterial world into genomically homogeneous units, genomic species, that attempt to reflect the natural occurrence of clusters of diversity — though this apparent structure might be due to under-sampling of the extant diversity [2,3]. While this definition of bacterial species is operational, the real challenge is to understand the underlying mechanisms that led to the emergence of this structure of bacterial diversity [4]. This is intimately linked to the current inability of evolutionary microbiologists to find a unifying model of prokaryotic cladogenesis. The classical "Biological Species Concept" (BSC) [5], which was originally defined for animals, places the sexual isolation of clades as the central condition for their divergence. In this model, sexual isolation can mostly arise through the appearance of prezygotic barriers or geographical separation of lineages. Prokaryotes usually co-occur in the environment, often with no geographical structure of populations, and clonally reproduce, but are subject to intra- and trans-specific sex through homologous recombination and horizontal gene transfer (HGT). In this context, it is impossible to understand the emergence of prokaryotic species in the framework of the BSC [6].

Ecological speciation appears as an alternative, and was proposed to be the major way through which prokaryotes diversify [7]. According to Cohan, within a bacterial population, some mutant genotypes may arise with a new function that changes the definition of their ecological niche by providing a newly accessible resource, or changing the way a previously accessible one is used. When the environment provides the resource in question, these mutants may

[*] Corresponding author.
E-mail addresses: florent.lassalle@ucl.ac.uk (F. Lassalle), daniel.muller@univ-lyon1.fr (D. Muller), xavier.nesme@univ-lyon1.fr (X. Nesme).

http://dx.doi.org/10.1016/j.resmic.2015.06.008
0923-2508/

successfully colonise the new niche, leading to the emergence of an ecotype, an ecologically differentiated lineage. The ecological distinctness of the new ecotype relative to the ancestral population or to sister ecotype lineages allows related ecotypes to co-exist indefinitely without competing, and by evolving along independent paths to form genetically distinct populations, i.e. to speciate [8]. More than any other kind of genomic change such as base substitution, gene duplication or loss, the acquisition of a new gene is likely to bring functional novelty to a genome and thus to open a new niche. HGT therefore appears as a major drive for ecological speciation. Paradoxically, HGT events cause prokaryotic genomes and genes to have (partially) decoupled histories, generating methodological and conceptual problems in the definitions of units of prokaryotic diversity [9,10]. The essence of an ecological definition of species is thus tightly associated with the source of conflict in the appreciation of bacterial diversity in terms of species. Nonetheless, various models of ecotype formation have been proposed to detail the different steps and conditions necessary for ecological speciation to occur in prokaryotes [11,12]. Only a small set of recent studies of nascent ecotype populations can, however, support or refute these theories [13]. Conversely, studies comparing genomes of older taxa provided insights into how variation, notably in gene content, could be linked to the ecology of clades of various depth [14−19], but little theoretical ground has been laid for understanding the evolutionary processes linking ecological speciation events to the advent of these clades.

Here, we will explore the virtues and limits of ecological speciation models in understanding evolution of prokaryotes over large time scales. In the light of comparative genomics and population genomics, we will attempt to reconcile organism/genome-centred and gene-centred notions of adaptation to ecological niches. Finally, we will see how the various predictions of these models might allow us to back-trace ancient events of adaptation of bacterial lineages and how this reverse ecology approach could help microbiologists to bridge the increasing gap between the assessment of genomic diversity and the knowledge of bacterial ecology.

2. Emergence of ecological species

According to Cohan's “stable ecotype” model [20], the acquisition of an adaptive gene (or mutation in general) determining colonisation of a new niche immediately imposes strong selection for the single adapted mutant. Clonal multiplication of the founder mutant yields a nascent ecotype population, and the potential providence in the environment of the newly exploitable resource may lead to its rapid expansion. Even though this scenario can take place in sympatric conditions with the ecotype living amongst non-adapted closely related populations, the latter are not expected to contribute to the genetic pool of the nascent ecotype. Indeed, recombination between closely related bacteria, as estimated in experimental microbial populations [21,22], is thought to be rare relative to the intensity of selection. This causes the founder genotype to spread in the new niche before linkage between the niche-specifying gene and the rest of the genome has had the time to be disrupted [23]. In other words, the ecotype progenitor clonally populates the new niche, founding a population genomically homogeneous where the selected niche-specifying gene is ubiquitous. (Fig. 1, steps 1 and 2). During the clonal reproduction that follows, clonal variants arise, among which those losing the niche-specifying gene are systematically removed from the population occupying this niche. Other variants may prove fitter than the progenitor genotype and may clonally replace it. This selective process causes a purge of genome-wide diversity within the ecotype population, and its recurrent occurrence is called periodic selection [24]. The frequent genotype replacement by any fitter variant is associated with the fixation of all neutral mutations that were linked to the selected one, thus accelerating the divergence of the ecotype from its parent population (Fig. 1, step 9).

However, recent studies showed that divergent ecotypes living in sympatry (i.e. occupying distinct niches in the same micro-habitat) exchange significant amounts of DNA, which makes them maintain a coherent population, with no structure of the genomic variation at loci other than those under ecological selection [25,26]. This conflicts with the classic ecotype model predictions (see Box 1: “Ecotypes: Definitions and concepts”), which can be explained by the departure from the assumption that the ratio of the recombination rate over the intensity of selection for the niche-specifying trait would be low; in fact, recombination rates are likely no higher than previously estimated, but the intensity of selection is probably lower than thought, which could be caused by selective trade-offs experienced by wild prokaryotic populations. Indeed, genomic monomorphy induced by strong selection for the niche-specifying gene could be counterselected to promote the diversity of genomic backgrounds [13]. For instance, there are several reported cases of selection to maintain many rare alleles of cell surface receptors in order to reduce the pressure from predators (phage, amoeba, immune system) on the whole population [27] or to maximise the number of different bacteriocins produced in the population [28]. Under these conditions, periodic selection will instead lead to gene-specific sweeps, meaning that the niche-specifying gene will be conserved within an ecotype population, but not the genomic background, which may recombine intensely with neighbouring populations (Fig. 1, steps 3−5). A group of ecotypes that are still able to recombine frequently with each other form a meta-population, a “maximally inclusive cluster of independently evolving lineages” that is nonetheless structured into several groups of preferential genetic exchange and higher phenotypic similarities [1,12]. Actually, frequent sex within meta-populations can promote the emergence of new ecotypes by transferring generally adaptive mutations between subpopulations, thus preventing their competition on shared niche dimensions [24]. Also, recombination between subpopulations could help ecotypes to more rapidly invent the combination of mutations necessary to reach a more stable adapted state in their respective niches [20].

In the “stable ecotype” model, early sexual isolation is not required for the initial separation of the ecotype lineage, but

Box 1. Ecotypes: definitions and concepts.

The seminal models of ecotypes defined in Cohan's framework [7,11,20,60] refer to ecologically homogeneous populations adapted to one single well-defined niche, which we can call ecotypes *sensu stricto* (s.s.). Because of their definition as an ecologically homogeneous population, any mutation resulting in phenotypic change can be seen as the emergence of a new ecotype s.s., which leads these entities to be extremely short-lived and to represent the smaller differential of evolution of a bacterial population. In these ecotypes s.s., it is fair to assume the population will diversify clonally only.

We here introduce the concept of ecoytpes *sensu lato* (s.l.), which are longer-lived populations resulting from a series of episodes in which an adaptive mutation has been followed by the successful speciation of an ecotype s.s. This suite of adaptations determined the nominal ecology of the lineage. However, ecotypes s.s. are constantly emerging within an ecoytpe s.l. population and then, for most, go extinct without founding any ecologically differentiated population, a process that maintains genotypic and phenotypic variation within the ecotype s.l. population. Importantly, standing variation within an ecotype s.l. population can possibly be mixed by homologous recombination and thus it is not assumed that it is composed of clonal lineages. Ecotypes s.l. are thus equivalent to meta-populations of ecotypes s.s., each with their own precise ecological niche definition that is derived from the core niche definition of the global ecotype s.l.

The presence of various ecotypes s.s. with different ecological niche definitions rarely leads to the founding of a new independent ecotype s.l. because the speciation event relies on complete realisation of this new niche, i.e. colonisation of an environment matching conditions specified by the ideal niche definition [99]. Also, the new environmental setting would need to sufficiently contrast with the original environment so as to provide the ecotype s.s. with a significant fitness advantage over all relatives, which could rely on rare events of long-distance transportation. Hence, an ecotype s.l. population may harbour variation with potential adaptive consequences that has remained neutral or almost neutral in the population's environmental context. In addition, variants inducing moderate changes in adaptation to the realised ecological niche may also be retained in the population, since they provide a buffer of adaptability when the environmental conditions fluctuate around the nominal definition of the ecotype's niche. This is particularly possible when recombination unlinks the niche-specifying loci and allows multiple combinations of variants to occupy the space of a broadly defined niche. In this context, periodic selection is expected to maintain (combinations of) variants at intermediate frequencies rather than purging them, reflecting the frequency spectrum of environmental conditions where they are the fittest.

Speciating ecotypes s.s. are probably hard to spot in the field given their ephemeral nature and the very few – if not unique – genomic and ecological traits that differentiate them from sister populations; they can, however, be identified and their evolutionary course well characterized using experimental evolution, the principles of which are extensively described in [100]. On the other hand, observing speciation of ecotypes s.l. in the environment is more probable, with S and L ecotypes of *Vibrio vulnificus* [26] as a good example.

later happens as a consequence of high sequence divergence [23]. In a more recent interpretation of the ecotype model [13], continued homologous recombination within the meta-population at niche-neutral loci maintains lower levels of sequence divergence and free mixing of polymorphisms between ecotypes. While genome-wide recombination delays global genomic divergence and hence the occurrence of recombination inhibition, it does not prevent niche differentiation as niche-specifying genes diversify rapidly under positive selection. The appearance of some form of assortative mating or sexual barrier, which would reduce the genomic cohesion of the global meta-population compared to that of ecotype subpopulations, thus seems required to trigger the process of divergence of ecotype lineages. This can occur through ecological isolation of the new ecotype from its parent population [26], with physical separation of habitats resulting in a decreased frequency of genetic exchange (Fig. 1, step 10). Similarly, the appearance of physiological barriers, such as plasmid or phage host-range differences or changes in the modification-restriction systems [20], can lead to sexual isolation (Fig. 1, step 11).

Whatever the timing of the appearance of sexual isolation between ecotypes, it is predicted that the genomic cohesion of the meta-population will eventually disappear. As the ecotype diverges from its ancestral population, its potential to recombine gradually decreases. This is notably due to the mechanical loss of efficiency of the recombination machinery linked to mismatch-repair (MMR), with a significant effect arising typically after a few percent genomic DNA divergence [21,29]. Over time, when a threshold of recombination over divergence is crossed, the ecotype lineage is no longer able to re-hybridize with its relatives, causing the irremediable separation of descendant clades, i.e. speciation [12,30]. One should note that the ultimate formation of genetically and ecologically distinct clusters of ecotype populations might be quite rare, given that most nascent ecotypes will be extinct before this step due, to contingent events of mass extinction and also possibly because only a few ecotypes will actually develop an evolutionarily stable strategy.

The "stable ecotype" and derived models seem appropriate to explain the bacterial speciation process from its initiation until the total separation of lineages. However, many limits can be found to this view. Notably, genetic exchange between related ecotypes might continue at loci that are not linked to niche-specifying genes long after their ecological separation. Recombination frequencies and levels of divergence between loci often appear to evolve heterogeneously, showing that the process of ecological speciation is asynchronous across the genome [31,32].

3. Genome-wide heterogeneity of the ecological speciation process

In situations where ecological isolation does not cause the complete rupture of physical contact between emerging cell lineages (i.e. sympatry), homologous recombination will still occur efficiently at shared loci like housekeeping genes that generally evolve slower than niche-specifying genes and hence can durably provide material fit for MMR-controlled recombination [33]. Conversely, the insertion of transferred genes bearing ecological adaptation will prevent recombination around the inserted locus. Indeed, following the "fragmented speciation" model proposed by Retchless and Lawrence [32], recombination with members of another population that do not possess the niche-specifying gene at the homologous locus would cause its loss, which would be counterselected [32,34]. Population genetics modelling showed that this causes a mechanical arrest of recombination at loci surrounding clade-specific insertions [35]. The multiplication of such sexually isolated clade-specific loci [31,34] or the propagation of recombination arrest around chromosomes [35] would eventually lead to complete loss of sexuality between sister lineages. Phylogenetic analyses of the genomes of Enterobacteriaceae revealed that, for several pairs of genera,

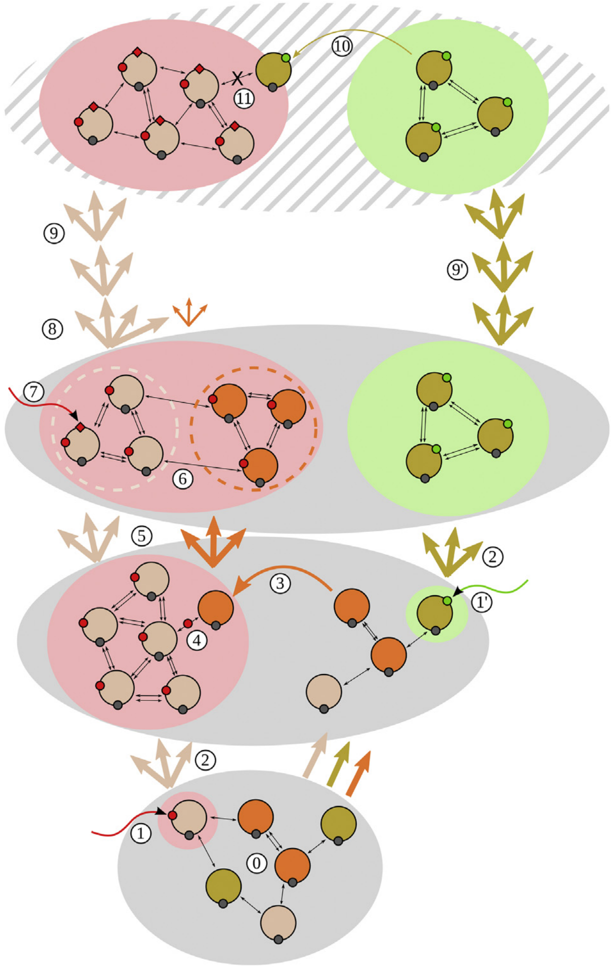

Fig. 1. Emergence, maintenance and diversification of ecological species. Lined discs of different colours represent bacterial genotypes or genomic backgrounds and peripheral pearls on these discs symbolise niche-specifying genes borne by these genotypes. Coloured backgrounds represent preferred microhabitats determined by genes of the same colour, i.e. to ecological niches. (0) A heterogeneous population occupying one ecological niche (grey background), with several genotypes (discs of different colours) sharing a common gene (grey pearl on disc perimeters) coding for adaptation to the common niche, exchange genetic material at rates dependent on genomic relatednesses (thin two-sided arrows, double for more intense exchange). (1 and 1′) Individuals receive a new gene (red or green pearl) by HGT from unrelated taxa, providing new functions that modify the definition of the ecological niche. The new ecological niches (red or green background, respectively) are derived from the ancestral one (grey background), with an additional dimension and/or changes in pre-existing dimensions of the niche definition; however, most other dimensions remain unchanged. The overlay of the grey and red or green backgrounds thus represents sharing of a core niche (grey) by populations with derived adaptations to red or green niches. We note that coexistence of several derived ecotypes is conditional on the appearance of a trade-off in exploiting the ancestral niche while adapting to the derived niche, which prevents colonisation of the ancestral niche by the new ecotype [60]. (2) The successful exploitation of the new niche triggers the clonal multiplication of the founder individual to form a nascent ecotype population. (3) A shared core ecology allows members of the parent population to meet and exchange genetic material, (4) with the possibility of transfer of the novel niche-specifying gene, allowing the mutant from the divergent genotype (orange disc) to compete equally with the ecotype clone (beige discs) in the new niche. (5) Owing to drift, the minor migrant genotype may rise in frequency and (6) quantitative differences in metabolism may let both genotypes exploit subtly different nano-niches (discontinuous circles), allowing them to cohabit temporarily with little competition but with possibility of gene flow. (7) A secondary HGT event brings another new gene (red diamond) in one of the genomic backgrounds, conferring a higher global or niche-specific fitness, (8) leading to the quashing of the diversity of nano-niche-specialized ecotypes. Later invasion of the ecotype becomes more difficult with the increasing number of niche-specifying genes to be transferred so as to sustain periodic selection. (9 and 9′) Episodes of periodic selection maintain genomic homogeneity within ecotype populations, while the accumulation of neutral substitutions drives whole genome divergence of different ecotypes. Gene flow between diverged ecotypes becomes limited (10) by the increased importance of the derived dimensions of ecological niches relative to the core niche, restricting the occurrence of shared habitats and thus the encounter rate, and (11) by the progressive appearance of barriers to DNA exchange, notably due to the sharp decrease in homologous recombination efficiency with genome divergence.

the inferred age of their respective last common ancestors varied across loci, supporting the hypothesis that this gradual speciation process was at work [31].

Thus, sexual isolation and ecologically-driven divergence might occur at a few loci, while most of the genome keeps recombining and evolving cohesively between the emergent ecotype lineage and the parent meta-population. This phenomenon was observed by Shapiro et al. [26] in *Vibrio cyclitrophicus* ecotypes that co-occur in coastal sea waters but are adapted to the attachment to particles of different sizes. In those genomes, the divergence between ecotypes is strongly marked at only a few divergent “islands” the size of one or a few genes. In fact, islands of divergence pointed to ecologically relevant gene functions, such as attachment and biofilm production [26], showing that heterogeneous levels of genomic divergence between ecotypes can be used to identify loci that might be adaptive in each niche, and hence infer the ecological properties of an ecotype. However, the ecotype speciation process is at too early a stage in that case to allow recognition of ecotypes at other (neutral) loci, such as housekeeping genes targeted in multi-locus sequence analyses (MLSA) and derived model-based approaches [36—39]. A similar process is thought to operate at a later stage, in *Sulfolobus islandicus* sympatric populations [25], where nucleotide sequence divergence has already spread from the adaptive “islands” into megabase-sized “continents”. In this case, the signal of lower recombination that marks the loci of the putative niche-specifying genes predicted by the hypothesis of Lawrence and Rechtless [32] is diluted by neutral divergence occurring across the genome, and the hypothesised ecological nature of the observed speciation could not be validated as well.

4. Maintenance of ecologically differentiated clades

One common limitation of the models described above is to neglect the possibility of hybridisation of separated lineages. In fact, periodic selection leads to the arrest of recombination in only one direction, by counter-selecting (within the niche) recombinants that have lost the niche-specifying locus [32]. It does not preclude the segment containing the niche-specifying gene being transferred to a related lineage. Only a weak argument has been provided to refute this possibility based on the potential loss of fitness induced by the gene gain [34], which could be linked to the cost of maintaining genes that are not useful, or to their toxicity out of their original context — for instance, due to the synthesis of a toxic chemical intermediate missing the subsequent enzyme to destroy it. The latter trait is unlikely to hold, as the archetypal niche-specifying mutation would more generally consist in the gain of functionally self-sufficient genes or operons — selfish operons [40] — that by definition do not need a genomic context to express properly, at least no more than in the donor ecotype's genome. Transfer of the niche-specifying gene between lineages is therefore possible provided it is neutral or beneficial in the recipient's niche. The latter case means that HGT will lead to periodic selection events leading to the purge of diversity in the recipient population with fixation of the transferred gene. Finally, the transfer event could also impact the recipient's niche delineation (e.g. by making accessible a new resource) and potentially trigger new ecological speciation (Fig. 2A, step 1—2).

This possibility may explain the accounts of sister species of *Neisseria*, with supposedly distinct ecologies, to be made of overlapping clusters of diversity as determined by multi-locus sequence typing (MLST). These 'fuzzy species' [41] indeed present multiple instances of hybrid lineages between the *bona fide* species lineages [42]. The group integrating both *N. lactamica* and *N. meningitidis* thus forms a meta-population whose diversity is unusually large. This may simply reflect on-going gene flow between recent ecotypes at loci that are neutral relative to the distinctive parameters of their ecological niches. In such a case, the possibility of hybridisation between ecologically differentiated groups of the meta-population may not persist after the later stages of their differentiation [43] and they may ultimately arise as distinct species (Fig. 2A, step 3). Alternatively, a hybrid lineage may persist and result in a new species successfully adapted to yet another similar, but distinct, ecological niche (Fig. 2A, step 4 and 2B).

This hybridisation process was observed with substantial introgression in *Campylobacter coli* with genes from *Campylobacter jejuni* after a relatively long time of isolation and divergence [44]. Interestingly, the authors advocate that the admixture of lineages could have a role in the ecological adaptation of agricultural lineages [44] (Fig. 2C). These genetic exchanges between distant relatives can be the source of many adaptations, as they can provide genes that sustained long-term selection in a genomic context which is essentially the same, and for this reason may integrate more easily into the recipient genome (see below on “Gene domestication”). In general, related species live in similar environments [45,46] and often live in sympatry [47—49]. Thus, certain adaptations on which one of these species relies to survive, such as the degradation of a particular sugar or attachment to a particular host cell, would probably provide a substantial selective advantage to a cognate species. The transfer of genes serving an adaptation to the shared aspects of their niche (or “core niche”, see below) may thus be promoted between sister ecotypes. Similarly, Cohan and Koeppel proposed the model of “nano-niches” [50], where cognate ecotype lineages share a pool of resources but use them in quantitatively different ways, thus partitioning the whole ecological niche. Niche partition allows close relatives to escape exclusive competition and thus to be able to live in sympatry [51]. The nano-niche model seems to apply to very recently diverged strains of *Bacillus subtilis* subsp. *spizizenii* (with ≥99.4% ANI and ≥95% shared genes), which may represent the earliest characterized stage of ecological speciation [8]. Such ecotypes may arise very frequently as quantitative variation in the niche dimensions are more likely to arise by mutation, leading to expression changes (e.g. mutations in promoters or gene copy number variation) and could even happen in a closed gene pool. However, the frequent emergence of these young nano-niche-specialized ecotypes may be balanced by their high extinction rate, linked to the high risk of appearance of a successful

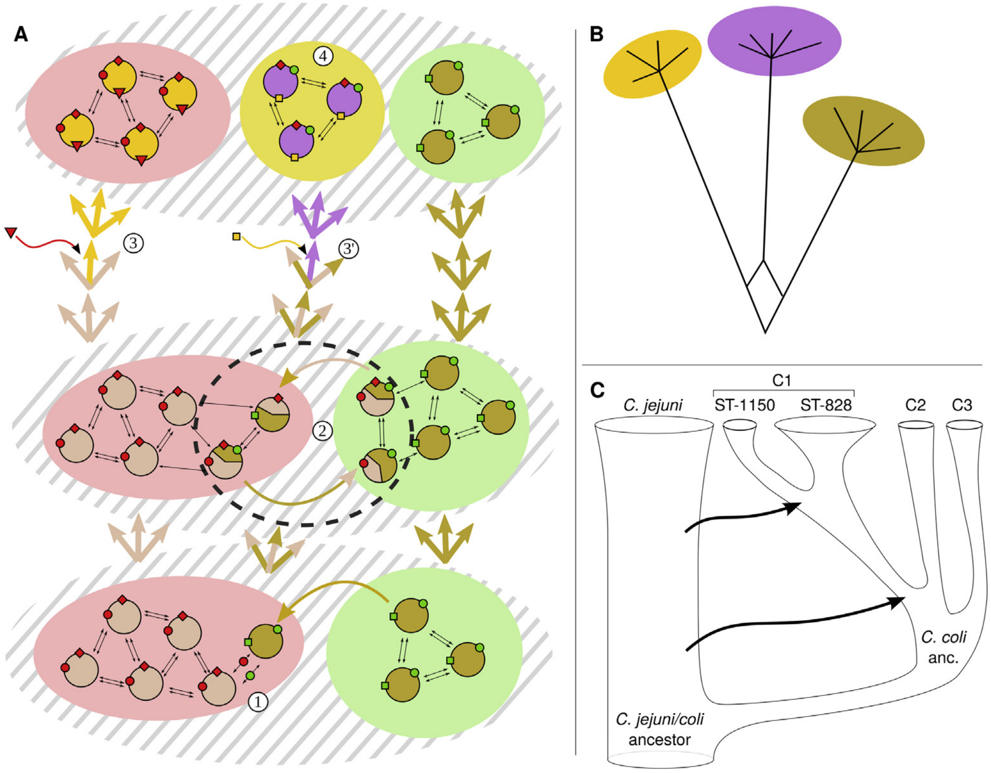

Fig. 2. Late hybrids can be maintained in the ecological speciation model. A. Two divergent ecotypes are adapted to fairly different niches. A rare event of gene exchange between ecotypes (1) leads to the emergence of a hybrid bearing partial sets of niche-specifying genes. (2) Frequent migration between niches and mating with resident ecotypes allows hybrid survival in the hybrid zone (discontinuous circ). (3 and 3′) Following new mutations that provide stronger niche-specific advantages, episodes of periodic selections (coloured arrows) leads to the purge of ecotype diversity. (4) Notably, fixation of synapomorphic adaptive mutations in the hybrid population makes it irremediably distinct from its parents in terms of both genome sequence and ecology. B. Hypothetical reticulated phylogeny obtained from neutral markers in extant ecotype genomes from scenario in (A), showing the early hybridisation event followed by the independent evolution of each three lineages. C. Schematic scenario of evolution of *Campylobacter jejuni/coli* (adapted from Sheppard et al., 2013 [44]). Two rounds of introgression of *C. jejuni* genetic material into *C. coli* Clade 1 (C1) ancestors (black arrows) leads to the emergence of hybrid lineages (ST-1150 and ST-828) adapted to the colonisation of agricultural animal hosts, while other *C. coli* lineages C2 and C3 mainly colonize wild animals. This was possibly due to the transfer of fucose metabolism genes mediating new interactions with host cells [44].

generalist that would outcompete them all [50]. On a larger evolutionary scale, however, pairs of clades with advanced divergence, such as between *C. coli* and *C. jejuni* (15% nucleotidic divergence over shared genes) [44], can hybridize without risking complete overlap of their ecological niches and exclusive competition. Rather, the hybrid will form a third lineage adapted to a new niche, which may differ from the former by an intermediate value of its key parameters – a situation that could still lead to ecological isolation under the nano-niche model – or more radically, owing to non-additive interactions between niche-specifying genes of mixed origins (Fig. 2A, step 4 and 2B).

5. Core genomes and core ecological niches

The major prediction of the ecotype model is that phylogenetic diversity should match ecological diversity. This principle was used to predict ecotype (or species) based on housekeeping gene sequence diversity in software such as EcotypeSimulation, BAPS, GMYC or PTP [36–39]. It was demonstrated at the finest scale of the ecotype emergence [52,53] and seems to hold as well at a much wider phylogenetic scale, with deep taxonomic ranks of bacterial diversity showing a hierarchical ecological structure [45,46], suggesting that ancient ecological speciation events might have left a durable imprint on bacterial genomes.

In fact, the major point of the ecotype model is that a niche-specifying mutation, such as a newly acquired gene, is transmitted vertically to the offspring and conserved in descendants due to periodic selection for the ecological property it provides. An ecological species can thus be defined as any clade of organisms that share an ecologically adaptive synapomorphic mutation. This definition stands at any level of diversity – for instance, the Firmicute phylum with its conserved Gram-positive cell wall allowing better resilience to osmotic stresses – and it is thus more appropriate to discuss ecologically relevant clades than ecological species. Potential transfer of the adaptive mutation to distant relatives or non-vertical transmission of the mutation within the clade would not invalidate this definition, because what matters is that the presence of the mutation in a globally similar genomic background leads to the expression of the same phenotypic trait by all clade members. This notably applies at the smallest level of bacterial variation, in the form of gene-specific sweeps within ecotypes (Box 1). In this way, and because the core genome of a clade necessarily encodes all the genetic determinants of the shared ecological properties of its members, it is theoretically possible to translate the core genome of a clade into its core ecological niche (Fig. 1, a grey core gene provides adaptation to a grey core niche). This principle can be applied recursively to all inclusive clades to define core ecologies of clades encoded by their core genome. The genomic variation marking evolutionary steps between every successive ancestor of inclusive clades should generally appear small compared to variation observed between individual strains, especially in terms of gene content. Indeed, the more recent the clade, the more likely the clade-specific variation will be neutral or deleterious, or only transiently adaptive. True adaptive mutations that may be stably retained are likely rare in the total amount of variation that each strain carries compared to the wealth of unselected strain-specific genes (ca. 30% of the gene content in strains of many taxa such as *Escherichia coli* or genomic species of *Agrobacterium* [17,54]). As time goes by, the deleterious mutations are purged by natural selection and neutral mutation, or else those providing only transient adaptations to rare conditions are eventually lost by drift. Hence, the more ancient a clade, the fewer conserved clade-specific mutations and the more likely these mutations would have contributed to long-standing and constant ecological adaptation. Therefore, the identification of core alleles or genes should help to pinpoint adaptive events that marked clade emergence and to locate genes that were under periodic selection during the entire period of its diversification. This latter point is of particular interest because exploration of the specific genomic properties of a clade can lead to the discovery of cryptic ecological properties conserved in a clade defined only by neutral traits, such as those used in standard multi-locus phylogeny. Indeed, it was possible to unravel unknown phenotypic traits common to *Agrobacterium fabrum* that likely contribute to the species' adaptation to life in the plant rhizosphere [54]. A set of ca. 200 genes were found to be specific to *A. fabrum*, far less than the ca. 1,800 genes specific to each strain in the same species; similar numbers of hundreds to thousands of strain-specific genes are observed in many taxa [17]. The strain-specific gene content must heavily impact the way each strain adapts to the environment — especially when it include genes that determine a pathogenic status. However, most of these strain-specific genomic and phenotypic features are transient, and conservation of ecological traits by members of a genomic species, while relatively rare, clearly marks the key features of a species' ecology.

A drawback of this approach could be that it relies on the sequencing of complete genomes, which today is only possible for cultured isolates. However, this limitation could soon be overcome by the advent of single-cell genome sequencing [55] and the assembly of pseudo-genomes from metagenomes [56–58] that can yield close-to-completion genomes for uncultivable strains, whose completeness and absence of chimerism can be validated using a robust statistical and phylogenetic framework [59].

Finally, one must take care to distinguish ancestral vs. core ecologies and genomes: core genomes (respectively ecologies) are the set of genes (resp. ecological properties) that remained conserved in all the extant descendants of an ancestor, and hence are included in, but not an exhaustive account of, the set of genes (resp. ecological properties) that made the ancestor's genome (resp. ecology). This distinction is crucial because some ancestral adaptations that were potentially important in the ancestor's niche definition and may have triggered key cladogenesis events could have been lost in the way. This may happen notably because of the appearance of trade-offs where new adaptations in a derived ecotype cause a decreased fitness in the ancestral niche, which may lead to later discarding of the ancestral niche-specifying genes. Such trade-offs have likely been occurring as frequently as cladogenesis events, because they ensure that a new ecotype will not immediately replace the ancestral population [60], from which another derived ecotype could later emerge, and descendants of both lineages may ultimately appear as sister clades (Fig. 1, step 1′). This leads to a more complex model where the derived niche is not simply an extension of the ancestral niche (in which case the derived ecotype could outcompete the ancestral one), but also induces modification of pre-existing adaptations. This phenomenon is well illustrated in a long-term evolutionary experiment where ecotypes specialised in the consumption of byproducts of glucose catabolism arose, but were less efficient growers on glucose than the ancestral ecotype, thus allowing cross-feeding of the former by the latter [61,62]. Similar dynamics may have led to the differentiation of S and L ecotypes of *V. cyclitrophicus*, with one having evolved better mobility for dispersion while discarding the possibility of thorough exploitation of food particles to which they attach [63].

6. Gene domestication and adaptation from the gene's standpoint

A new adaptive gene does not define a brand new ecology for its host. Rather, it adds a new dimension to the existing ecological niche, and the combination has to fit an evolutionarily stable strategy in order to continue. That is why most genes defining pathogenic behaviour are unlikely to integrate the core genome of a species, due to the intrinsic instability of the niche embodied by a sick host. Similarly, genes providing very high gain of fitness through strong ecological specialisation (e.g. heavy-metal resistance genes enabling colonisation of contaminated environments) may lead to evolutionary dead-ends if the targeted resources are too rare or the habitat unstable. Indeed, condition-specific genes are often seen to frequently transfer across the Tree of Life [64,65]—because they can (their associated fitness is disconnected from the genomic background) and because they must (those that do not transfer will go extinct with their host). In addition, the new gene must pass through a period of adaptation to the host genome environment or be lost. Higher deletion and substitution rates are observed for younger genes [66–68], a phenomenon interpreted by Hao and Golding [69] as “a mixture of directional selection to adapt in some genes and neutral mutation destroying function in others". Indeed, the successful integration of a gene is marked by its rapid evolution towards an adaptation to the norms of the genome in terms of codon usage [70,71] and integration into cellular networks of regulation, protein—protein interaction and metabolism [65,72,73]. Given the large majority of strain-specific orphan genes not conserved after acquisition [67,71,74], the norm is indeed to fail genomic integration or to lead to unstable evolutionary strategies (see ecotypes *sensu stricto* in Box 1). Sometimes however, genes with a nomadic past are domesticated by the core genome and these associations can yield successful specialists, as exemplified by *Salmonella* as an obligatory pathogen [75] or *Rhizobium sp*. NT-26 as a toxic mine resident [76].

In order to recognise a gene domestication event, one has to map the appearance of an adaptive mutation and recognise the clade or clades that acquired it. While generally satisfying, this framework may stumble on a notable issue: the capacity to define a phylogeny of organisms in which monophyly of groups of interest is reliable. This is a long-standing problem, due in part to the difference between the history of a genome and of its constituent genes (see Box 2). It could even prove to be completely wrong if the adaptive genes of interest co-evolved with a constant minority of the genome rather than with the bulk. This is the case for pathogenic or symbiotic genes borne by large mobile elements that are transmissible across species. It has thus been proposed, in the case of *Rhizobium leguminasorum*, host of the symbiotic plasmids pRL, that beyond the operational definition of the host species based on clusters of core genome diversity, one can define symbiovars based on specificity for the nodulated host plant. These symbiovars are a polyphyletic grouping of the host (bacterial) core genomes, but match clades of symbiotic genes [49]. Hence in this case, an ecological species concept could be applied when observing the nodulating rhizobia from the point of view of the plasmid, as has been proposed by

Box 2. Good species phylogeny for the recognition of clade-specific adaptations.

Obtaining reliable species phylogeny has long been known to be difficult, for several reasons: 1) phylogenetic reconstruction methods rely on wrong models of molecular evolution; or 2) they may lack the signal to distinguish hypotheses; and finally 3) because different genes in the genome have different histories, owing to homologous recombination, incomplete lineage sorting, lineage-specific duplication or loss and, most importantly in the case of prokaryotes, HGT [6]. Depending on which genomic locus one takes as a reference phylogenetic marker, the history of the mutation may appear different, with a number of introduction events (via de novo origination or HGT) varying widely.

One approach to dealing with this problem is to consider only the "vertically evolving" genes, meaning those whose common history is the most frequent in the genome. A first implementation consists in selecting widely conserved housekeeping genes, such as those used in MLST studies, and using concatenation or supertree methods to obtain an average history. Housekeeping genes are good phylogenetic markers because they are known to be less prone to duplication, loss and transfer through the Tree of Life [101]. This is likely due to the fitness cost of altering the function of genes highly integrated into the central cellular machinery when changing their copy number or replacing them with a distant version [65,73,102]. However, this assumption is not necessarily true at shorter evolutionary time scales, because highly-connected housekeeping genes have low evolutionary rates [103] and replacement by HGT between cognate species is likely to bring very few changes and hence be neutral. Another more agnostic procedure is to combine all conserved (core) gene families — presumed to be groups of orthologues — and to consider the emerging majority signal as the history of the genomes. The most satisfying methods are those aimed at reconciling histories of gene and genome. Among these, some perform explicit reconstruction of the recombination history in the core genome to extract the underlying clonal frame [104,105]. Other approaches reconstruct the many events of duplication and transfer in the entire pangenome [106,107], and actually use these events as informative characters to search the best species tree, for instance, the one maximizing the number of genes found with the same "vertical" signal [101,108].

When following the goal of identifying niche-specifying genes that marked the origin of a clade, one is tempted to look for genes that mirror clade diversity and to scrupulously follow the vertical history of this clade. However, it is expected that later adaptive allelic variants of adaptive genes to have spread repeatedly by homologous recombination within the ecotype. It is therefore more appropriate to look for candidate adaptive genes among those conserved in all clade members — irrespective of their histories — and the molecular evolution pattern of which is suggestive of periodic selection events.

Wiedenbeck and Cohan [60] under the ecotype model of "recurrent niche invasion".

However, it would be untrue to say that the association of these large mobile elements with their host genome is unstable and completely uncoupled from the host genome's ecology: they indeed have been domesticated, but are only fit in a frequency-dependent manner. Many pathogenic islands and other elements like the SXT ICEs in *Vibrio cholerae* or the PAI islands in *E. coli* are specific to their host, though only present in epidemic strains and in a minor environmental reservoir. Legume nodulation genes (*nod*) are present quite heterogeneously in the Proteobacterial tree, but have been tamed a few times in Alpha and Beta subdivisions [77], with the notorious event of association with symbiotic megaplasmids in the Rhizobiaceae family [78]. These genes did coevolve with their host and may have driven the shape of their core ecology. Ti plasmids, causing *Agrobacterium* strains to induce tumors on plants, use molecular signalling to initiate the plant colonisation that was likely an exaptation from intimate interactions that agrobacteria had previously developed with plants [79]. This may have led certain agrobacteria to maintain molecular systems on their core genome that are advantageous in the presence of the Ti plasmid: for instance, genes domesticated by *A. fabrum* that allow detoxification and use of phenolic acids as carbon sources [54,80] target the same compounds for which the Ti plasmid provides a positive tropism [81]. Hence, these large mobile elements have the ambivalent status of parasites that manipulate their host and of beneficial symbionts that may have contributed to the success of their host genomes over the long term. Following the "selfish operon" principle [40], a group of co-functioning genes promotes its survival through the maintenance of its co-transmission, which can be through common transfer, as efficiently promoted by ICEs and conjugative plasmids, or through vertical descent, which cannot be better achieved than by providing a fitness advantage to the host genome.

7. Adaptive genomic islands and incipient speciation

A comparative study conducted on all genomic species of the *Agrobacterium tumefaciens* species complex revealed sets of genes specific to all inclusive clades, and provided them with unique phenotypic traits and probably specific ecological abilities [54]. Interestingly, most clade-specific genes are located in several genomic islands in the genome, including those bringing clearer gains of function, such as catabolism of phenolics and synthesis of a siderophore [54]. Such patterns of conserved islands of adaptation are common in Prokaryotes, as exemplified by heavy-metal resistance island in *Thiomonas arsenivorans* [82] or islands involved in biosynthesis of outer membrane structures in *Alteromonas macleodii* and *Prochlorococcus* [55,83]. Similarly, several studies comparing related genomes showed that genomic island variation occurred at a few defined loci [15,17,84], where homologous recombination of surrounding core genes can promote the non-homologous replacement of genes [83,85]. Acquisition of gene cassettes, integrated plasmids and other mobile elements is often seen as characteristic of strain-specific acquisition of new ecological traits, with short-term conservation expectancy matching the duration of an extreme ecological niche such as non-obligatory pathogeny or symbiosis [60]. However, environmental lineages acquiring functions inducing a less dramatic change in ecology may fix such mobile elements and stably incorporate the ecological trait they encode into their niche. In the context of the very large and actively recombining metapopulation of *Prochlorococcus*, distinct genomic backbones were found to be statistically associated with niche-specifying genomic islands, seemingly restricted to clades by a dynamic balance of selection over recombination [55]. Recognition of such fluidly domesticated genes relies on good coverage of extant genomic diversity to avoid confusion with nomadic genes displaying similar intermediate frequencies, but no relation to population structure [86,87]. An interesting feature of such domesticated genomic islands is that they may retain the genomic plasticity associated with mobile elements (e.g. by conserving the flanking recombinogenic integrases or ISs), allowing the genetic basis for the selected trait to evolve more dynamically. For instance, accumulation of functionally related genes in an integron can build up the organism's fitness relative to a specific ecological aspect [76,88]. Inversely, several cases have been reported of frequent replacement of genomic islands encoding structures displayed at the cellular surface by non-homologous counterparts through homologous recombination at the islands' extremities between closely related but ecologically differentiated lineages living in sympatry. This can allow evasion of phage predation [89] or immune system pressure [90] while maintaining function, events which may contribute to the blurring of their ecological differentiation.

8. Tracking niche-specifying genes in genomes: methodologies

Clade-specific genes (i.e. those present in all members of a clade and absent from closest relatives) could be used to identify potential determinants of past adaptations under the postulate that genes that supported the adaptation were gained by the ancestor of this clade and conserved afterwards through continued periodic selection [54]. This method enables robust characterization of genes that were most likely under strong selective pressure since their acquisition, and hence participated significantly in the ecological adaptation of the lineage. However, this approach discards many other genes that may have participated in the emancipation of the lineage, via several possible scenarios: for instance, genes already present in the ancestral population but changing function in the derived ecotype through a point mutation may be conserved in both the ecotype and sister lineages, though under different selective pressures. Alternatively, genes which triggered the ecological speciation of a clade may have been lost, for instance because subsequent fixation of secondary niche-specifying mutations may have led to the decreased importance of primary niche-specifying mutations in their contribution to niche adaptation. Simple analysis of gene distribution in clades (i.e. phylogenetic profiling) would not be able to spot such cases. In addition, non-adaptive processes impacting nucleotide composition [91], organisational constraints upon genome expression at the scale of operons, topological domains or replichores [92,93] or acquisition of selfish addictive genes such as toxin-antitoxin modules can lead to the conservation of genomic traits that are not ecologically adaptive and may be mistaken as such.

An alternative way to characterize events of ecological adaptation could be the characterisation of changes in selective pressure regimes, as proposed by Vos [94] in the so-called "adaptive divergence species concept". This concept, in fact, derives directly from the stable ecotype model, as it relies on the signature left by periodic selection in a genome and is reminiscent of Lawrence and Retchless's early attempt to characterize speciation genes based on codon usage adaptation [32]. As the ecotype diverges and specializes into occupying its new niche, adaptive changes accumulate through time in the ecotype lineage and are maintained afterwards at high frequency by periodic selection, while the resulting purge of the ecotype diversity erases traces of most neutral changes that had occurred during the same period. This contrast can be captured by the MacDonald–Kreitman test when comparing a pair of populations (e.g. an ecotype and its parent) or in a phylogenetic framework through dN/dS-related tests [95]. Adaptive divergence thus provides an operational basis for a species concept that does not rely on thresholds of divergence levels and thus can be applied to any system, irrespective of its evolutionary dynamics [94]. Contrary to the approach consisting of detecting clade-specific gene gains or losses, i.e. variation in the accessory genome, this kind of approach has the advantage of applying to conserved (core-genome) genes and detecting point-mutation-based adaptations. In fact, dN/dS-based tests could be applied systematically to all genes in a pangenome, for instance by conducting branch-specific tests on reconciled gene trees – gene trees whose branches are systematically mapped to a species tree – to provide a generalized means of characterising adaptive mutations occurring anywhere in the genome, core or accessory. However, current implementations [96] only consider protein-coded adaptations and ignore changes in the accessory genome – though it is the most likely site of adaptive changes – and hence can only characterize relatively long-standing adaptive divergence of lineages, such as between genomic species. This may be due to the difficulty in accounting for biases due to differences in size of gene families inducing heterogeneous power to detect selection across the pangenome. Also, in such an approach relying on characterisation of molecular processes of evolution, it would be straightforward to integrate model parameters that account for non-adaptive processes biasing the molecular signature of natural selection [91,97].

Additional predictions of the ecotype (sensu lato) model (see Box 1) could similarly be tested, including the fact that periodic selection events are likely to propagate fitter alleles of the niche-specifying genes through homologous recombination. Continuous adaptation to a niche should thus lead to frequent purging of diversity at the niche-specifying loci, while other loci, unlinked by recombination, are more likely to conserve their standing diversity. Hence, sequences at selected loci may result from histories denser in recombination events, which may be best diagnosed by long branches leading to clusters with rare diversity (clades with star topologies in gene phylogenies). This consideration also applies to non-core genes that provide adaptation to a clade under transient conditions, like genomic islands. It has been observed that flexible genomic islands providing frequency-dependent, condition-specific advantages, like the staphylococcal cassette chromosome (SCC) in the pathogen *Staphylococcus aureus*, or flexible genomic islands coding surface proteins targeted by phages in the marine bacterium *A. macleodii*, were more easily transmitted through homologous recombination of flanking core genes [85,89]. In the case of the flagellum glycosylation locus of *A. macleodii*, it is interesting to note that recombination promoted diversification of surrounding genes while enforcing monomorphy of the adaptive flexible locus. Such a variable signature may be a landmark of adaptation of a species to a secondary ecological niche.

9. Conclusion

The advent of genomics provides us with the means to finely describe the diversity of bacteria at the finer within-population to the coarser inter-species scale. Model-based evolutionary studies geared to reconstructing the process of bacterial diversification enable the identification the forces – adaptive and non-adaptive – that have generated the observed diversity. These approaches allow us to evaluate the ambiguous role of genetic exchange between closely related ecotypes living in sympatry in driving or preventing speciation. Indeed, recombination can maintain the genomic cohesion of ecologically diverse meta-populations while favouring the partition of ecological niches, by combining ecological traits from diverged populations. As a trait may be shared by different clades via vertical descent (forming a dimension of their core ecological niche) or by late transfer (forming, in different combination with other traits, the specific dimension of each derived niche), it is crucial to reconstruct the history of acquisition of niche-specifying mutations in order to understand the role of ecology in driving bacterial cladogenesis. Studying core-genomes of inclusive clades enables us to reconstruct the incremental process of adaptation of lineages to fractal definitions of ecological niches.

The reverse ecology approach – learning ecological properties from genomes – is attractively remote from the sometimes tedious task of culturing bacteria and characterising their phenotype in the lab. However, this practice could lead to the partial and abstract definition of niches based on functions barely related to the physical world. It is thus crucial to confirm the ecological predictions of reverse ecology using functional assays [54,80] and to complement it with systematic characterisation of the biochemical properties of the ecological niche of microbes [98].

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

FL was supported by European Research Council (ERC) grant BIG_IDEA 260801 (http://erc.europa.eu). XN was funded by the French National Research Agency (ANR) grant ECOGENOME (ANR-BLAN-08-0090) and AGROMICS grant from the ENVIROMICS challenge of the Interdisciplinary Mission of the French National Centre for Scientific Research (CNRS). The authors thank FM Cohan and another anonymous reviewer for constructive comments, and Benjamin Sobkowiak for his help proofreading the manuscript.

References

[1] Achtman M, Wagner M. Microbial diversity and the genetic nature of microbial species. Nat Rev Micro 2008;6:431–40.

[2] Konstantinidis KT, Tiedje JM. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr Opin Microbiol 2007;10:504–9.

[3] Caro-Quintero A, Konstantinidis KT. Bacterial species may exist, metagenomics reveal. Environ Microbiol 2012;14:347–55.

[4] Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kämpfer P, Maiden MCJ, et al. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 2002;52:1043–7.

[5] Mayr E. Systematics and the origin of species, from the viewpoint of a zoologist. Harvard University Press; 1942.

[6] Doolittle WF, Zhaxybayeva O. On the origin of prokaryotic species. Genome Res 2009;19:744–56.

[7] Cohan FM. What are bacterial species? Annu Rev Microbiol 2002;56:457–87.

[8] Kopac S, Wang Z, Wiedenbeck J, Sherry J, Wu M, Cohan FM. Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. Appl Environ Microbiol 2014;80:4842–53.

[9] Doolittle WF. Microbial evolution: stalking the wild bacterial species. Curr Biol 2008;18:R565–7.

[10] Haggerty LS, Jachiet P-A, Hanage WP, Fitzpatrick DA, Lopez P, O'Connell MJ, et al. A pluralistic account of homology: adapting the models to the data. Mol Biol Evol 2014;31:501–16.

[11] Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. Curr Biol CB 2007;17:R373–86.

[12] Fraser C, Hanage WP, Spratt BG. Recombination and the nature of bacterial speciation. Science 2007;315:476–80.

[13] Shapiro BJ, Polz MF. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol 2014;22:235–47.

[14] Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci 2006;103:15611–6.

[15] Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, et al. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet 2007;3:e231.

[16] Lefébure T, Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol 2007;8:R71.

[17] Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet 2009;5:e1000344.

[18] Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. Proc Natl Acad Sci 2011;108:7200–5.

[19] Williams D, Gogarten JP, Papke RT. Quantifying homologous replacement of loci between haloarchaeal species. Genome Biol Evol 2012;4:1223–44.

[20] Cohan FM. Sexual isolation and speciation in bacteria. Genetica 2002;116:359–70.

[21] Roberts MS, Cohan FM. The effect of DNA sequence divergence on sexual isolation in *Bacillus*. Genetics 1993;134:401–8.

[22] Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. ISME J 2008;3:199–208.

[23] Cohan FM. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am Nat 1994;143:965–86.

[24] Cohan FM. Bacterial species and speciation. Syst Biol 2001;50:513–24.

[25] Cadillo-Quiroz H, Didelot X, Held NL, Herrera A, Darling A, Reno ML, et al. Patterns of gene flow define species of *Thermophilic archaea*. PLoS Biol 2012;10:e1001265.

[26] Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, et al. Population genomics of early events in the ecological differentiation of bacteria. Science 2012;336:48–51.

[27] Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pašić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. Nat Rev Microbiol 2009;7:828–36.

[28] Cordero OX, Wildschutte H, Kirkup B, Proehl S, Ngo L, Hussain F, et al. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. Science 2012;337:1228–31.

[29] Majewski J, Zawadzki P, Pickerill P, Cohan FM, Dowson CG. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J Bacteriol 2000;182:1016–23.

[30] Doroghazi JR, Buckley DH. A model for the effect of homologous recombination on microbial diversification. Genome Biol Evol 2011;3:1349–56.

[31] Retchless AC, Lawrence JG. Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. Proc Natl Acad Sci U S A 2010;107:11453–8.

[32] Retchless AC, Lawrence JG. Temporal fragmentation of speciation in bacteria. Science 2007;317:1093–6.

[33] Engel P, Stepanauskas R, Moran NA. Hidden diversity in honey bee gut symbionts detected by single-cell genomics. PLoS Genet 2014;10:e1004596.

[34] Lawrence JG, Retchless AC. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol Clifton NJ 2009;532:29–53.

[35] Vetsigian K, Goldenfeld N. Global divergence of microbial genome sequences mediated by propagating fronts. Proc Natl Acad Sci U S A 2005;102:7332–7.

[36] Koeppel A, Perry EB, Sikorski J, Krizanc D, Warner A, Ward DM, et al. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc Natl Acad Sci 2008;105:2504–9.

[37] Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinform 2008;9:539.

[38] Fujisawa T, Barraclough TG. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol 2013;62:707–24.

[39] Zhang J, Kapli P, Pavlidis P, Stamatakis A. A general species delimitation method with applications to phylogenetic placements. Bioinformatics 2013;29:2869–76.

[40] Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 1996;143:1843–60.

[41] Hanage WP, Fraser C, Spratt BG. Fuzzy species among recombinogenic bacteria. BMC Biol 2005;3:6.

[42] Corander J, Connor TR, O'Dwyer CA, Kroll JS, Hanage WP. Population structure in the *Neisseria*, and the biological significance of fuzzy species. J R Soc Interface 2012;9:1208–15.

[43] De Queiroz K. Ernst Mayr and the modern concept of species. Proc Natl Acad Sci 2005;102:6600–7.

[44] Sheppard SK, Didelot X, Jolley KA, Darling AE, Pascoe B, Meric G, et al. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol Ecol 2013;22:1051–64.

[45] Philippot L, Andersson SGE, Battin TJ, Prosser JI, Schimel JP, Whitman WB, et al. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Micro 2010;8:523–9.

[46] Koeppel AF, Wu M. Lineage-dependent ecological coherence in bacteria. FEMS Microbiol Ecol 2012;81:574–82.

[47] Vogel J, Normand P, Thioulouse J, Nesme X, Grundmann GL. Relationship between spatial and genetic distance in *Agrobacterium spp*. in 1 cubic centimeter of soil. Appl Environ Microbiol 2003;69:1482–7.

[48] Shams M, Vial L, Chapulliot D, Nesme X, Lavire C. Rapid and accurate species and genomic species identification and exhaustive population diversity assessment of *Agrobacterium spp*. using recA-based PCR. Syst Appl Microbiol 2013;36:351–8.

[49] Kumar N, Lad G, Giuntini E, Kaye ME, Udomwong P, Shamsani NJ, et al. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. Open Biol 2015;5:140133.

[50] Cohan FM, Koeppel AF. The origins of ecological diversity in prokaryotes. Curr Biol CB 2008;18:R1024–34.

[51] Hardin G. The competitive exclusion principle. Science 1960;131:1292–7.

[52] Sikorski J, Nevo E. Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at “Evolution Canyons” I and II, Israel. Proc Natl Acad Sci U. S. A 2005;102:15924–9.

[53] Smith NH, Gordon SV, de la Rua-Domenech R, Clifton-Hadley RS, Hewinson RG. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat Rev Microbiol 2006;4:670–81.

[54] Lassalle F, Campillo T, Vial L, Baude J, Costechareyre D, Chapulliot D, et al. Genomic species are ecological species as revealed by comparative genomics in *Agrobacterium tumefaciens*. Genome Biol Evol 2011;3:762–81.

[55] Kashtan N, Roggensack SE, Rodrigue S, Thompson JW, Biller SJ, Coe A, et al. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 2014;344:416–20.

[56] Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat Biotechnol 2014;32:822–8.

[57] Kang DD, Froula J, Egan R, Wang Z. A robust statistical framework for reconstructing genomes from metagenomic data. bioRxiv 2014:011460.

[58] Imelfort M, Parks D, Woodcroft BJ, Dennis P, Hugenholtz P, Tyson GW. GroopM: an automated tool for the recovery of population genomes from related metagenomes. PeerJ 2014;2:e603.

[59] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 2015;25:1043–55.

[60] Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiol Rev 2011;35:957–76.

[61] Rozen DE, Schneider D, Lenski RE. Long-term experimental evolution in *Escherichia coli*. XIII. Phylogenetic history of a balanced polymorphism. J Mol Evol 2005;61:171–80.

[62] Le Gac M, Plucain J, Hindré T, Lenski RE, Schneider D. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. Proc Natl Acad Sci 2012;109:9487–92.

[63] Yawata Y, Cordero OX, Menolascina F, Hehemann J-H, Polz MF, Stocker R. Competition–dispersal tradeoff ecologically differentiates recently speciated marine bacterioplankton populations. Proc Natl Acad Sci 2014;111:5622–7.

[64] Daubin V, Lerat E, Perrière G. The source of laterally transferred genes in bacterial genomes. Genome Biol 2003;4:R57.

[65] Pál C, Papp B, Lercher MJ. Horizontal gene transfer depends on gene content of the host. Bioinformatics 2005;21(Suppl. 2):ii222–3.

[66] Marri PR, Hao W, Golding GB. The role of laterally transferred genes in adaptive evolution. BMC Evol Biol 2007;7:S8.

[67] Van Passel MWJ, Marri PR, Ochman H. The emergence and fate of horizontally acquired genes in *Escherichia coli*. PLoS Comput Biol 2008;4:e1000059.

[68] Daubin V, Ochman H. Bacterial genomes as new gene homes: the genealogy of ORFans in *E. coli*. Genome Res 2004;14:1036–42.

[69] Hao W, Golding GB. The fate of laterally transferred genes: life in the fast lane to adaptation or death. Genome Res 2006;16:636–43.

[70] Lawrence JG, Ochman H. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol 1997;44:383–97.

[71] Daubin V, Perrière G. G+C3 structuring along the genome: a common feature in prokaryotes. Mol Biol Evol 2003;20:471–83.

[72] Lercher MJ, Pal C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol Biol Evol 2008;25:559–67.

[73] Davids W, Zhang Z. The impact of horizontal gene transfer in shaping operons and protein interaction networks – direct evidence of preferential attachment. BMC Evol Biol 2008;8:23.

[74] Lerat E, Daubin V, Ochman H, Moran NA. Evolutionary origins of genomic repertoires in bacteria. PLoS Biol 2005;3:e130.

[75] Porwollik S, Wong RM-Y, McClelland M. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. Proc Natl Acad Sci U. S. A 2002;99:8956–61.

[76] Andres J, Arsène-Ploetze F, Barbe V, Brochier-Armanet C, Cleiss-Arnold J, Coppée JY, et al. Life in an arsenic-containing gold mine: genome and physiology of the autotrophic arsenite-oxidizing bacterium *Rhizobium sp*. NT-26. Genome Biol Evol 2013;5:934–53.

[77] Tian CF, Zhou YJ, Zhang YM, Li QQ, Zhang YZ, Li DF, et al. Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations. Proc Natl Acad Sci 2012;109:8629–34.

[78] Gonzalez V, Acosta JL, Santamaria RI, Bustos P, Fernandez JL, Hernandez Gonzalez IL, et al. Conserved symbiotic plasmid DNA sequences in the multireplicon pangenomic structure of *Rhizobium etli*. Appl Env Microbiol 2010;76:1604–14.

[79] Vaudequin-Dransart V, Petit A, Poncet C, Ponsonnet C, Nesme X, Jones JB, et al. Novel Ti plasmids in agrobacterium strains isolated from fig tree and chrysanthemum tumors and their opinelike molecules. Mol Plant-Microbe Interact MPMI 1995;8:311–21.

[80] Campillo T, Renoud S, Kerzaon I, Vial L, Baude J, Gaillard V, et al. Analysis of hydroxycinnamic acid degradation in *Agrobacterium fabrum* reveals a coenzyme A-dependent, beta-oxidative deacetylation pathway. Appl Environ Microbiol 2014;80:3341–9.

[81] Lee YW, Jin S, Sim WS, Nester EW. Genetic evidence for direct sensing of phenolic compounds by the VirA protein of *Agrobacterium tumefaciens*. Proc Natl Acad Sci U S A 1995;92:12245–9.

[82] Arsène-Ploetze F, Koechler S, Marchal M, Coppée J-Y, Chandler M, Bonnefoy V, et al. Structure, function, and evolution of the Thiomonas spp. genome. PLoS Genet 2010;6:e1000859.

[83] López-Pérez M, Gonzaga A, Rodriguez-Valera F. Genomic diversity ofnbsp;“Deep ecotype” *Alteromonas macleodii* isolates: evidence for pan-mediterranean clonal frames. Genome Biol Evol 2013;5:1220–32.

[84] Gonzaga A, Martin-Cuadrado A-B, López-Pérez M, Mizuno CM, García-Heredia I, Kimes NE, et al. Polyclonality of concurrent natural populations of *Alteromonas macleodii*. Genome Biol Evol 2012;4:1360–74.

[85] Everitt RG, Didelot X, Batty EM, Miller RR, Knox K, Young BC, et al. Mobile elements drive recombination hotspots in the core genome of *Staphylococcus aureus*. Nat Commun 2014;5.

[86] Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. Science 2006;311:1768–70.

[87] Cordero OX, Polz MF. Explaining microbial genomic diversity in light of evolutionary ecology. Nat Rev Microbiol 2014;12:263–73.

[88] Muller D, Médigue C, Koechler S, Barbe V, Barakat M, Talla E, et al. A tale of two oxidation states: bacterial colonisation of arsenic-rich environments. PLoS Genet 2007;3:e53.

[89] López-Pérez M, Martin-Cuadrado A-B, Rodriguez-Valera F. Homologous recombination is involved in the diversity of replacement flexible genomic islands in aquatic prokaryotes. Evol Popul Genet 2014;5:147.

[90] Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet 2013;45:656–63.

[91] Lassalle F, Périan S, Bataillon T, Nesme X, Duret L, Daubin V. GC-content evolution in bacterial genomes: the biased gene conversion hypothesis expands. PLoS Genet 2015;11:e1004941.

[92] Abby S, Daubin V. Comparative genomics and the evolution of prokaryotes. Trends Microbiol 2007;15:135–41.

[93] Lassalle F, Daubin V. Evolution of prokaryotic pangenomes. In: Francino MP, editor. Horizontal gene transfer in microorganisms. Norfolk, UK: Caister Academic Press; 2012. p. 23–31.

[94] Vos M. A species concept for bacteria based on adaptive divergence. Trends Microbiol 2011;19:1–7.

[95] Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol 2011;28:3033–43.

[96] Vos M, te Beek TAH, van Driel MA, Huynen MA, Eyre-Walker A, van Passel MWJ. ODoSE: a webserver for genome-wide calculation of adaptive divergence in prokaryotes. PLoS One 2013;8:e62447.

[97] Dutheil J, Boussau B. Non-homogeneous models of sequence evolution in the Bio++ suite of libraries and programs. BMC Evol Biol 2008;8:255.

[98] Materna AC, Friedman J, Bauer C, David C, Chen S, Huang IB, et al. Shape and evolution of the fundamental niche in marine *Vibrio*. ISME J 2012;6:2168–77.

[99] Hutchinson GE. Concluding remarks. Cold Spring Harb Symp Quant Biol 1957;22:415–27.

[100] Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nat Rev Genet 2013;14:827–39.

[101] Abby SS, Tannier E, Gouy M, Daubin V. Lateral gene transfer as a support for the tree of life. Proc Natl Acad Sci 2012;109:4962–7.

[102] Cohen O, Gophna U, Pupko T. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. Mol Biol Evol 2011;28:1481–9.

[103] Aris-Brosou S. Determinants of adaptive evolution at the molecular level: the extended complexity hypothesis. Mol Biol Evol 2005;22:200–9.

[104] Didelot X, Falush D. Inference of bacterial microevolution using multilocus sequence data. Genetics 2007;175:1251–66.

[105] Didelot X, Lawson D, Darling A, Falush D. Inference of homologous recombination in bacteria using whole genome sequences. Genetics 2010;186:1435–49.

[106] Szöllősi GJ, Boussau B, Abby SS, Tannier E, Daubin V. Phylogenetic modeling of lateral gene transfer reconstructs the pattern and relative timing of speciations. Proc Natl Acad Sci 2012;109:17513–8.

[107] Szöllősi GJ, Tannier E, Lartillot N, Daubin V. Lateral gene transfer from the dead. Syst Biol 2013;62:386–97.

[108] Boussau B, Szöllősi GJ, Duret L, Gouy M, Tannier E, Daubin V. Genome-scale coestimation of species and gene trees. Genome Res 2013;23:323–30.