

OPINION

The ecological coherence of high bacterial taxonomic ranks

Laurent Philippot, Siv G. E. Andersson, Tom J. Battin, James I. Prosser, Joshua P. Schimel, William B. Whitman and Sara Hallin

Abstract | The species is a fundamental unit of biological organization, but its relevance for Bacteria and Archaea is still hotly debated. Even more controversial is whether the deeper branches of the ribosomal RNA-derived phylogenetic tree, such as the phyla, have ecological importance. Here, we discuss the ecological coherence of high bacterial taxa in the light of genome analyses and present examples of niche differentiation between deeply diverging groups in terrestrial and aquatic systems. The ecological relevance of high bacterial taxa has implications for bacterial taxonomy, evolution and ecology.

If you have ever read a field guide for birds, mammals or plants, you will have noticed that it includes not only a description of each organism but also a wealth of information about their habitat, geographical distribution and ecological traits. Such information is available at the species level as well as at high taxonomic levels, such as the phylum. A field guide for bacteria might sound far-fetched, considering that the number of bacterial species that have been described is low (~7,000)¹ in relation to the millions of bacterial species that have been predicted to reside on Earth². However, the current limitations for such a guide are not only the enormous diversity of bacteria but also our lack of knowledge about the ecology of each of the known species. We do not know the distribution of most bacteria in nature, and we do not know what they are actually doing in their natural environments. In fact, we cannot even define the niche of the best studied bacterial species, *Escherichia coli*^{3,4}. These difficulties are intertwined with our struggle to apply the species concept to bacteria^{1,5-7}.

Ever since bacteria were discovered, the question of whether they form phylogenetically coherent units and the extent to which such units correlate with phenotypic characteristics and environmental adaptations have been debated. Carl Woese used

ribosomal RNA (rRNA) sequences to infer phylogenies, through which he proposed a natural system that sorted organisms into three domains of life: Bacteria, Archaea and Eukarya⁸. Attempts to attribute ecological traits to bacterial and archaeal groups that are defined at taxonomic ranks higher than species are rare, and the idea that the deep branches of the bacterial and archaeal trees could be ecologically coherent has seldom been considered⁹. There are several reasons for this. Genetic diversity is much greater in the domains Bacteria and Archaea than in the kingdom Metazoa. For example, the percentage of conserved genes between strains of the bacterial species *Ralstonia solanacearum* can be as low as 68%¹⁰, whereas 75% of human genes have homologues in the genome of the fish *Takifugu rubripes*¹¹. Likewise, there can be substantial physiological diversity within bacterial clades; for example, the phylum Proteobacteria includes heterotrophs, lithotrophs and phototrophs. In bacteria, ecological traits can be strain specific, and closely related bacteria can occupy distinct niches^{12,13}. Furthermore, horizontal gene transfer (HGT) can occur even between distantly related organisms¹⁴. In fact, some authors claim that HGT is so rampant that no natural classification system can be described for bacteria and that the Tree of

Life should instead be described as a Web of Life¹⁵. It has also been suggested that a microbial niche can be changed after a single HGT event¹⁶. However, there is a growing body of literature showing that HGT takes place mostly between closely related species because of reduced genetic barriers¹⁷. In addition, there is evidence that many high taxonomic levels of bacteria show 'ecological coherence' (REFS 18,19). Ecological coherence of a taxon means that the members of that taxon share general life strategies or traits that distinguish them from members of other taxa. Some obvious examples of shared life strategies among high taxonomic groups are found in the phylum Cyanobacteria, in which all members are capable of photoautotrophic growth, and the phylum Chlamydiae, in which all known members are obligate intracellular symbionts or parasites of eukaryotes. If we assume that adaptations to different ecological niches have shaped and differentiated bacterial genetic lineages during the course of evolution, then these divergences should be reflected not only at the strain or species level but also at higher taxonomic ranks. However, ecological coherence can be confounded by selective pressures acting on different niches, resulting in different genomes even for bacteria that share a common evolutionary history. Moreover, selective constraints can also drive convergent evolution, producing bacterial species that are functionally similar despite their different evolutionary histories. For these reasons, the question of whether ecological coherence exists at deeper evolutionary divergences also addresses the fundamental evolutionary processes that have shaped such lineages.

In this Opinion article, we examine whether there is ecological coherence of bacterial taxonomic ranks higher than the species level. First, we discuss the ecological coherence of various bacterial clades in the light of genome analyses. Next, we highlight studies that support the ecological relevance of high bacterial taxa in terrestrial and aquatic systems. Finally, we discuss the implications of these data for bacterial taxonomy, evolution and ecology. We focus here on the domain Bacteria, but the same proposal of ecological coherence of high taxonomical ranks can also be made for Archaea.

A genomic perspective

Currently, more than 1,000 bacterial genomes have been completely sequenced, and there are several hundred ongoing bacterial genome projects at various stages of completion (see Further information for links). Although the collection of species for which genomes are available is biased towards bacteria that are important to human health and agriculture, pilot projects that advocate the sequencing of genomes solely for their phylogenetic novelty, such as the Genomic Encyclopaedia of Bacteria and Archaea (GEBA) project, are in progress and promise to generate a more balanced and useful data set²⁰. This makes it timely to pose questions about the levels in the taxonomic hierarchy that display coherence in lifestyle and about whether such clades share unique ecological traits.

Genome sequence data enable the construction of sequence-based phylogenies for all genes in the genome and have led to the emergence of a new field of research: phylogenomics^{21,22}. Work in this field offers insights into the mechanisms by which bacterial genomes change and adapt to new environments. Core genes encoding proteins involved in basic information processes have been particularly useful in delineating genetic relationships as a complement to the more traditional 16S rRNA gene phylogenies. Although the divergences of deeper nodes are hard to resolve, the different approaches agree on the identification of many taxonomic clades. Indeed, the good correlation between the 16S rRNA gene tree

and gene content-based trees at high taxonomic ranks²³ suggests that a large fraction of bacterial genomes reflect an underlying pattern of vertical gene descent.

One approach of examining the depth of ecological coherence and the genomic changes associated with lifestyle shifts is to superimpose recently acquired functional traits and ecological adaptations onto a tree that displays vertical descent. The simplest expectation is that all members of ecologically coherent clades will share a unique set of genes²⁴ and, indeed, such genes have been identified for some taxonomic clades. For example, around 120 of the 1,054 core genes identified for the Cyanobacteria are specific to this phylum²⁵. Likewise, 233 and 51 signature genes have been reported for the phyla Actinobacteria and Chlorobi, respectively^{26,27}. For the phylum Bacteroidetes, the order Bacteroidales and the genus *Bacteroides*, 27, 38 and 185 signature genes, respectively, have been identified²⁷. The number of signature genes tends to be negatively correlated with taxonomic rank, as exemplified above, with fewer signature genes at the order level than at the genus level and even fewer at the phylum level. It is also likely that the number of signature genes present in all members of a taxon will decrease as more genomes are sequenced. From a conceptual standpoint, it can be argued that the coherence of high taxa could also result from a unique combination of genes, without any one gene being taxon specific on its own. For example, the presence of an almost complete set

of photosynthetic genes is easy to associate with the lifestyle of the Cyanobacteria. However, genomic signatures of more complex phenotypes such as virulence, thermophily or radiation resistance have been more difficult to delineate²⁸. Moreover, many signature genes encode proteins of unknown function, which might not necessarily be involved in environmental adaptation.

How much ecological diversity is there within a taxonomically defined bacterial lineage? The class Alphaproteobacteria, which has more than 100 sequenced representatives, is one of the most ecologically diverse classes and is therefore a particularly good model system to use to address this question²⁹. An alphaproteobacterial species tree has been inferred from concatenated data sets of 38 and 102 protein-coding genes^{30,31}. On the basis of this tree, a divergence pattern has been inferred for the major bacterial orders within the Alphaproteobacteria, including the Rickettsiales, Rhodospirillales, Sphingomonadales, Caulobacterales, Rhodobacterales and Rhizobiales. The habitats of a few species from each of these orders are shown in FIG. 1. Although many lifestyles and environments are represented, distinct patterns of ecological consistency at the level of genus and order are observed in several cases. This enables genomic diversity to be correlated with the ecological diversity in and between these groups. For the clades with intracellular habitats, the lower the taxonomic rank, the narrower the growth niche (as defined by the range of animal species infected). The identification of many diverse but ecologically consistent clades has revealed that adaptation to open and changing environments tends to be associated with genome size expansions and rapid changes in gene repertoires, whereas adaptation to isolated and stable environments is typically reflected in reductive evolution. Extensive gene loss during ecological shifts to restricted environments might also result in ecological coherence for high taxonomic ranks such as genera and phyla, just as novel gene acquisition might enable expansion into a novel niche, with subsequent diversification yielding ecological coherence of the lineage.

A striking observation is that differences in lifestyle correlate with variations in the genes encoding proteins that have functions associated with environmental interactions, such as regulation and transport³⁰. For example, the small genomes of obligate and facultative intracellular bacteria have only a few regulatory and transport genes, whereas the larger genomes of free-living

Glossary

Biogeography

The distribution of organisms over space and time, including where they live, at what abundance, and why; these data offer insights into the mechanisms that generate and maintain diversity.

Clade

A phylogenetically coherent bacterial group.

Deep sequencing

High-throughput sequencing designed to maximize the amount of sequence information that can be gained from an environmental sample, the ultimate aim being exhaustive sequencing.

High taxonomic level

A taxonomic level from genus to phylum.

Horizontal gene transfer

The acquisition of a new DNA fragment from another organism.

Ileostomy

A surgical opening between the abdominal wall and the ileum (small intestine).

Lineage

A group of taxa that are related by descent from a common ancestor.

Niche

The particular set of resources and environmental conditions that an individual species exploits.

Phylogeny

The inferred evolutionary relationships among a group of organisms (most often inferred on the basis of molecular data).

Phylotype

A taxon-neutral term referring to an organism with a unique genetic make-up according to its evolutionary relationship to other organisms.

Species

A generally accepted species definition has yet to be established for microorganisms. For the purposes of taxonomy, an *ad hoc* definition on the basis of DNA hybridization is widely used. This definition is based on degrees of relatedness, without an underlying biological principle.

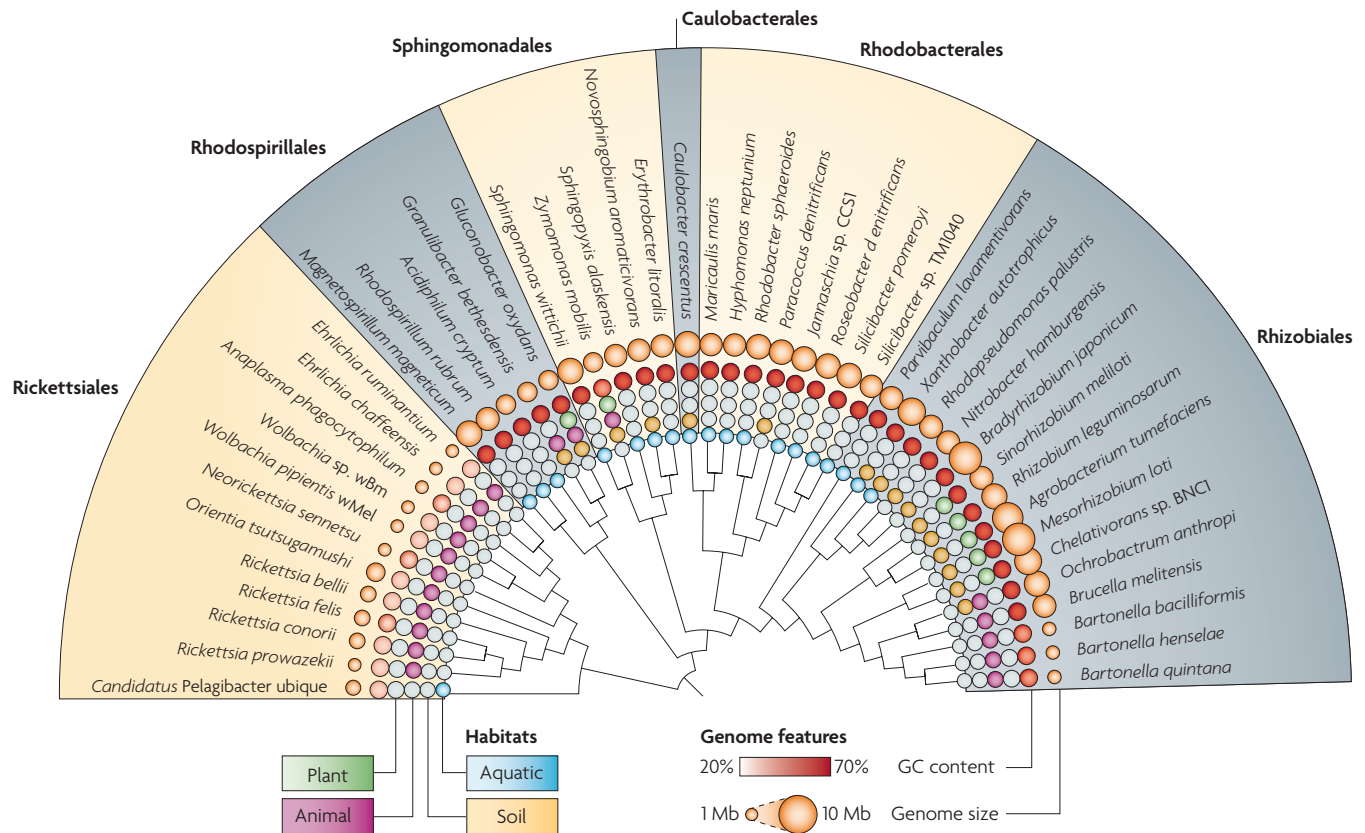


Figure 1 | Habitat–phylogeny associations for the Alphaproteobacteria. The growth habitats and genomic GC contents for the alphaproteobacterial species for which complete genome data are available. Ecological coherence in terms of habitat preference is manifested at different levels in the phylogeny. At the highest level, there is no

coherence in habitat preferences. At lower levels, patterns start to appear; for example, most members of the order Rickettsiales are adapted to the cytoplasm or nucleus of animal host cells. Figure is reproduced, with permission, from REF. 29 © (2009) Royal Society Publishing.

soil bacteria that alternate between environments of different nutritional status contain hundreds of such genes. Most are located in regions of the genome that are hot spots for mobility. By physically separating genes encoding proteins involved in core functions from those that are important for environmental interactions, and associating the latter with mobile elements, responses to environmental change can be achieved, even if the taxonomic distribution remains unaltered. Thus, the ecological coherence of a taxonomic clade could best be described at the genomic level by the presence of a unique and characteristic repertoire of gene families associated with environmental interactions, each of which contains individual genes that vary within the clade.

An environmental perspective

The idea of characterizing bacterial lineages according to life strategies has been suggested in several recent studies^{18,32–36}. In the most comprehensive work to date, Fierer *et al.*¹⁸ used both experimental and

meta-analysis approaches to classify soil bacteria. Their results led them to propose that certain bacterial classes or phyla can be differentiated into ecological categories on the basis of *r*-type or *K*-type life strategies, with the *r* types growing rapidly under conditions of high resource availability and the *K* types exhibiting lower growth rates but higher substrate affinities. The class Betaproteobacteria and the phylum Bacteroidetes were considered to be *r* strategists, whereas members of the phylum Acidobacteria were classed as *K* strategists¹⁸. As suggested by the phylum name, preference for an acidic pH was reported to be an important trait of several subdivisions in the Acidobacteria. In the environment, the relative abundance of the Acidobacteria in the total bacterial community correlated negatively with soil pH, in contrast to the observations for several other taxa^{37–39}. Accordingly, the recent sequencing of the complete genomes of three Acidobacteria strains revealed candidate genes indicative of a mechanism for tolerance to acidic conditions⁴⁰. Another example of a bacterial clade that shares life strategies

is SAR11 (in the order Rickettsiales), which consists of very small, heterotrophic, aquatic alphaproteobacteria and accounts for around 35% of the bacterioplankton communities in marine surface waters^{41,42}. The microbial-ecology literature reveals many other studies reporting either site- or time-specific community structures at high taxonomic levels^{43–46}, as well as differential responses of bacterial lineages to changing environmental factors^{44,47–51}. Because shifts in community structure can be detectable at the phylum level, it has been suggested that certain bacterial phyla, such as the Acidobacteria and Proteobacteria, can be used as indicators of nutrient status owing to differences in their lifestyles^{32,34,39}.

If deeply branching bacterial lineages have ecological relevance, then we would expect there to be evidence for biogeography at high taxonomic levels. In fact, there are cases of such patterns being seen. In a recent study using metagenomic data for taxonomic profiling of microbial communities from four different environments — agricultural soil, ocean surface water, acid mine drainage

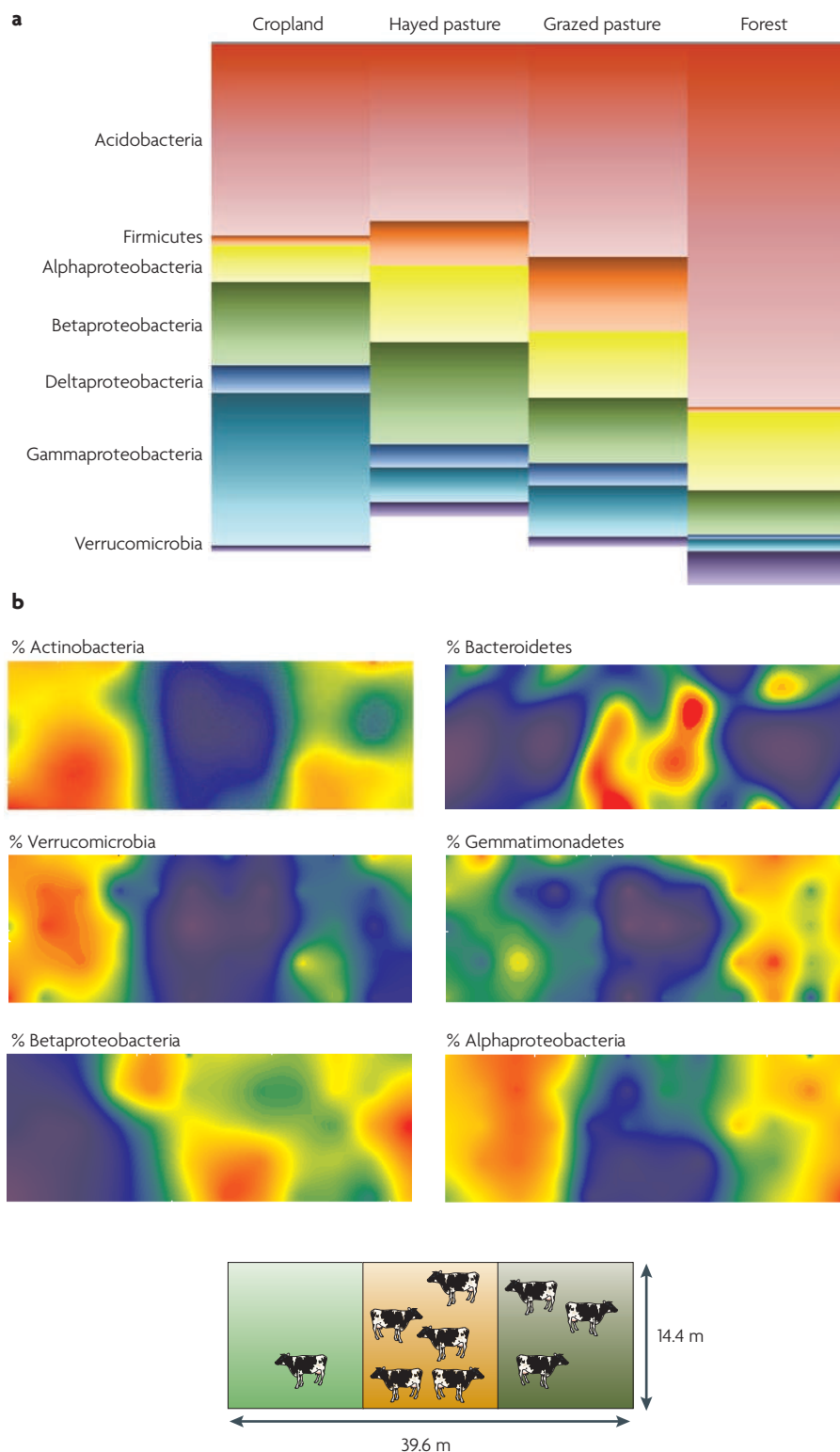


Figure 2 | Niche differentiation at the phylum and class levels. a | The relative abundance of bacterial taxa in bacterial communities in soil under different land management regimes⁴⁹. Only those taxa for which the abundance changed significantly with management are shown. **b** | The spatial distribution of the relative abundance of targeted bacterial classes or phyla within the total bacterial community of a pasture subjected to three levels of cattle impact, as shown at the bottom. Field-scale variations of the relative abundance of the different taxa are indicated by the different colours, from blue (lower relative abundance) to red (higher relative abundance). Part **a** courtesy of K. Jangid, University of Georgia, Athens, USA. Part **b** modified, with permission, from REF. 19 © (2009) Society for Applied Microbiology and Blackwell Publishing.

and deep-sea whale bone — strong associations were reported between habitat and phylotype⁵². For example, relatives from the order Rickettsiales were found mostly in surface water, whereas relatives from the orders Rhizobiales and Burkholderiales were found mostly in soil. In aquatic ecosystems, the composition of bacterial communities showed limited overlap between marine and freshwater environments, and a few high taxa, such as the Betaproteobacteria (which are nearly absent in oceans), were typical of fresh waters^{53,54}. In estuaries, which are the mixing zones between both ecosystems and, thus, extraordinary field laboratories to study ecological niches of bacterial lineages, a clear dominance was found for the Alphaproteobacteria in the lower, saltier water and for the Betaproteobacteria in the upper, fresher water, whereas members of the Flavobacteria prevailed in the turbidity maximum of the estuary⁵⁵. In terrestrial ecosystems, which have greater spatial heterogeneity than water columns, ecological niches and spatial patterns of microbial communities can develop at much smaller scales. For example, it was found that changes in land management, such as fertilizer treatment and cropping, led to large differences in the abundances of many classes and phyla in the soils of adjacent fields⁴⁹ (FIG. 2a). Likewise, the relative abundances of several classes and phyla were shown to display spatial patterns at the metre scale¹⁹; these patterns were taxon specific and correlated to soil properties in pastures that were differently affected by cattle (FIG. 2b).

In addition to spatial patterns, temporal patterns have been reported and are exemplified by successional transitions of bacterial populations. During aging of compost, different high-rank phylogenetic groups were identified at different stages⁴³, including increases in the phylum Actinobacteria during the later stages⁵⁶. Another example of succession at high taxonomic rank is provided by a recent analysis of the microbial community of the human small-bowel lumen, which revealed a dramatic shift from the orders Clostridiales and Bacteroidales to the orders Lactobacillales and Enterobacteriales after an ileostomy⁵⁷. Following surgical closure of the ileostomy, the community composition is similar to that of the normal microbiota, suggesting that the introduction of oxygen into the otherwise anaerobic ileum was responsible for the shift in the dominant bacterial orders (FIG. 3). Clear seasonal signatures in the abundance of major bacterial lineages have been observed in soil^{32,33} and in marine ecosystems⁵⁸. In aquatic ecosystems, temporal changes have also been reported at much

shorter timescales, with variations in the representation of the major classes and phyla over the course of a 12-hour tidal cycle in an estuary⁵⁹.

These environmental studies support the idea of ecological coherence of bacterial groups at taxonomic ranks higher than the species. Being based on a range of independent approaches, including metagenomics, fluorescent *in situ* hybridization, quantitative PCR, and PhyloChip and phospholipid fatty acid analyses, these conclusions are robust. However, owing to the limitations of these current methods for grasping the entire diversity of a microbial community, even in a single environmental sample, the patterns observed mainly reflect variations in the dominant members of the bacterial communities.

Such methodological limitations highlight important difficulties. First, the patterns described above might reflect underlying patterns occurring in nature at low taxonomic levels, because some taxa in a lineage were very dominant, whereas others were either absent or present at very low abundance in the studied environment. A good example is the highly abundant SAR11 clade in ocean surface water, which results in the dominance of the Alphaproteobacteria in this environment⁴¹. Therefore, deep sequencing, as suggested by Vogel *et al.*⁶⁰, and comparisons across environments are required to identify the exact taxonomic level at which bacterial lineages are ecologically coherent. More extensive metagenomic studies will provide unique opportunities for microbiologists to fill in the gaps in the phylogenetic tree with genetic and ecological data, enabling inferences of the environment from which any given clade emerged and how that clade has since dispersed. However, information on the habitat preferences of 15,350 cultured microorganisms showed that strains related at the genus, family or order levels occur in the same environment more frequently than do random pairs of isolates⁵² (FIG. 4). This suggests that it is possible to distinguish ecological coherence up to the order level using only a few keywords to describe the habitat. It should be emphasized here that it is unlikely that taxon-specific traits will be uniformly shared by all members of the taxon. Moreover, ecological coherence will probably not be observed at the same level of organization among all bacterial lineages owing to the extensive variation in sequence divergence between lineages. It is therefore important to accept imperfection when recognizing the ecological cohesiveness of high bacterial

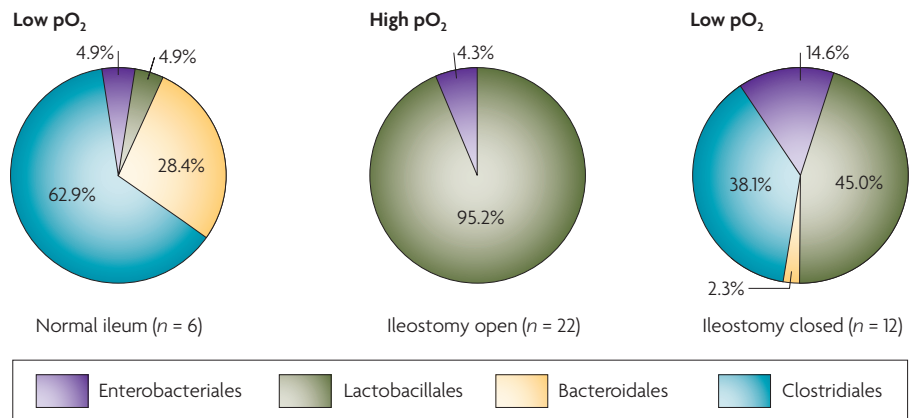


Figure 3 | **Microbial succession at the order level in the human small-bowel lumen.** The relative abundance of the four dominant bacterial taxa in ileal effluents of a normal ileum, after an ileostomy and after surgical closure of the ileostomy. Figure is modified from REF. 57 © (2009) National Academy of Sciences USA.

taxa and to focus on unifying rather than on universal traits, as biologists frequently do in animal and plant ecology. This approach is perfectly reasonable when we acknowledge that the traits that unify high taxa might be advantageous in a variety of otherwise unrelated habitats.

Implications

Our proposal that high bacterial taxa, as delineated by phylogenetic analyses of 16S rRNA and other highly conserved genes, could be ecologically meaningful has numerous implications.

Let us first consider the implications for microbial taxonomy. It is difficult for systematists to identify ecologically meaningful units of bacterial diversity⁶¹, and phylogenetic analysis of conserved genes has been used to facilitate the interpretation of frequently conflicting phenotypic data and to delineate taxonomic ranks from the genus up to the class or phylum levels. The ecological coherence of high bacterial taxa provides evidence that these ranks have a biological or physiological basis. An increasing body of literature proposes the incorporation of ecology into the definition of bacterial species^{1,62–64}. We argue that this approach should not be restricted to the species level but should be extended to higher taxonomic ranks when appropriate. We acknowledge that this might not be easy for taxonomists, but hopefully, thanks to the recent developments of metagenomics and other high-throughput technologies, valuable data sets will soon be available to link bacterial ecology and evolution. Metagenomics will also provide a better estimate of the sequence diversity that is present in different habitats at different time points, and mapping this diversity back onto the species tree will

present new opportunities for characterizing the extent of coherence versus the spread of ecological units for different lineages and environments.

As bacterial taxa at high ranks are defined in part by the phylogeny of the 16S rRNA and other highly conserved genes, their ecological coherence also has implications for bacterial evolution, providing an independent validation of deep nodes of the bacterial phylogenetic tree. However, even if ecological features that distinguish deeply diverging clades can be identified, this, per se, does not provide proof that we have found and recognized the tree of life in nature⁶⁵. Without knowing the hierarchical relatedness of high taxa, their ecological coherence provides no evidence for the actual topology of the tree. Instead, whenever phylogenetic and ecological clusters match, deeply diverging clades can be considered as a unit in evolution. Indeed, the use of high taxa as evolutionary units has previously been suggested by Valentine, who stated that “there does not seem to be anything special about higher Linnaean taxa, as compared to species, that would prevent them from being macroevolutionary units, provided that they have heritable features than can undergo sorting” (REF. 66). Correlating the presence or absence of patterns of traits with the presence or absence of patterns of genes might help functional annotations and pathway predictions, and may provide information about the selective forces driving the appearance and disappearance of genes throughout bacterial evolution. The ecological coherence of deep nodes also supports recent work suggesting that there are barriers to HGT between distantly related bacteria and that a cohesive force holds members of high taxa together¹⁷.

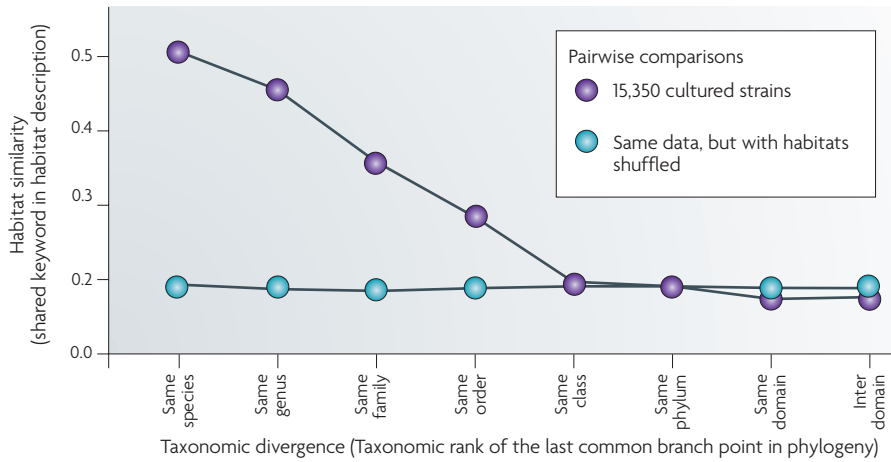


Figure 4 | Habitat–phylogeny associations. A comparison of cultured microbial strains, plotting habitat similarity against level of strain relatedness according to the National Center for Biotechnology Information (NCBI) taxonomy database. For the order level and for all taxonomic ranks lower than this, the difference is highly significant ($P < 10^{-6}$). The similarity of the annotated habitats, measured by automated keyword comparisons between any two microbial isolates, was strongly correlated to the evolutionary relatedness of the strains. Figure is modified, with permission, from REF. 52 © (2007) American Association for the Advancement of Science.

Possibly the greatest repercussions of the concept of the coherence of high taxa will be in microbial ecology. Microbial ecologists are still struggling with both species definitions and the huge number of bacterial species that are present in the environment. Circumventing such bottlenecks would facilitate the development of generic concepts and theories in microbial ecology⁶ and aid in understanding coexistence and community assembly, which are central issues in microbial ecology. We think that dissecting the immense amount of bacterial diversity down to the species level is not always necessary in order to understand bacterial community functions and interactions. The ecological coherence of high taxa suggests that deeper clades of various ranks could be used as alternative ecologically meaningful units, although we do not claim that using higher levels of organization will explain everything. Rather, different biological patterns can be observed at different levels of organization as the result of the constraints imposed by higher levels upon lower levels. Therefore, even if the species is a fundamental category of biological organization, it is important to understand ecology at every ecologically meaningful level of organization; in the words of Mayr, we need to obtain “an interpretation of the role of lower and higher taxa in the economy of nature and in evolutionary history” (REF. 67).

In addition, collapsing the immense bacterial diversity at the species level by integrating it into high levels of organization

would provide a framework and the operational tools to study bacterial assemblages in a more practical way. This is shown by a model of the temporal dynamics of the gut microbiota of human infants⁶⁸. This dynamic interaction model for the Proteobacteria, Firmicutes and Bacteroidetes showed that the tendency of the microbiota to converge on a common profile might be explained by a small number of simple phylum-level interactions.

Understanding the diversity of living organisms is not an esoteric exercise, and the relationship between biodiversity and ecosystem functioning has emerged as a major concern owing to the potential impact of the diversity losses that could be caused by human activities⁶⁹. Interest in this issue rapidly extended to microbiology, and the importance of microbial diversity for ecosystem function is now being widely investigated. By working with life history traits that are coherent at high levels of bacterial taxonomy, we can scale up this question and address the potential role of bacterial communities at high taxonomic ranks for ecosystem functioning. Promising results in this direction from the gut microbiome show that obesity is associated with changes in the relative abundance of the two dominant bacterial groups defined at the phylum level, the Bacteroidetes and the Firmicutes^{70,71}. The ecological coherence of high bacterial taxonomic ranks provides hope that bacterial community composition can be used to predict ecosystem functioning.

Conclusions

Data from genome analyses and ecological studies support the hypothesis that some high bacterial taxa are ecologically coherent. Such coherence is not categorical but is rather a continuous variable across taxonomic ranks within a lineage, with a negative correlation between the level of ecological coherence and the taxonomic rank. From a practical perspective, all taxonomic levels of organization are of interest, as each could provide different pieces of information on the underlying mechanisms driving the formation of bacterial assemblages. Thus, using the allegory of a ‘field guide’ for bacteria, it might not be necessary to wait until we can describe the ecology of every single species; instead, such a guide could describe unifying ecological traits at higher ranks, with the number and complexity of traits increasing as we move down the rank levels.

Laurent Philippot is at the Department of Soil and Environmental Microbiology, Institut National de la Recherche Agronomique, University of Burgundy, BP 86510, 21065 Dijon, France.

Siv G.E. Andersson is at the Department of Molecular Evolution, Evolutionary Biology Center, Uppsala University, 752 36 Uppsala, Sweden.

Tom J. Battin is at the Department of Limnology, University of Vienna, 1090 Vienna, Austria.

James I. Prosser is at the Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 3UU, Scotland, UK.

Joshua P. Schimel is at the Department of Ecology, Evolution & Marine Biology, University of California, Santa Barbara, California 93106-9620, USA.

William B. Whitman is at the Department of Microbiology, University of Georgia, Athens, Georgia 30602-2605, USA.

Sara Hallin is at the Department of Microbiology, Swedish University of Agricultural Sciences, 750 07 Uppsala, Sweden.

Correspondence to L.P.
e-mail: Laurent.Phillippot@dijon.inra.fr

doi:10.1038/nrmicro2367

Published online 7 June 2010

1. Achtman, M. & Wagner, M. Microbial diversity and the genetic nature of microbial species. *Nature Rev. Microbiol.* **6**, 431–440 (2008).
2. Gans, J., Wolinsky, M. & Dunbar, J. Computational improvements reveal great microbial diversity and high metal toxicity in soil. *Science* **309**, 1387–1390 (2005).
3. Byappanahalli, M. N., Whitman, R. L., Shively, D. A., Sadowsky, M. J. & Ishii, S. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ. Microbiol.* **8**, 504–513 (2006).
4. Ishii, S., Ksoll, W. B., Hicks, R. E. & Sadowsky, M. J. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl. Environ. Microbiol.* **72**, 612–621 (2006).
5. Gevers, D. *et al.* Re-evaluating prokaryotic species. *Nature Rev. Microbiol.* **3**, 733–739 (2005).
6. Prosser, J. I. *et al.* The role of ecological theory in microbial ecology. *Nature Rev. Microbiol.* **5**, 384–392 (2007).

7. Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G. & Hanage, W. P. The bacterial species challenge: making sense of genetic and ecological diversity. *Science* **323**, 741–746 (2009).
8. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domain Archaea, Bacteria and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
9. Doolittle, W. F. The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis mean for both of them. *Phil. Trans. R. Soc. Lond. B* **364**, 2221–2228 (2009).
10. Guidot, A. *et al.* Genetic structure and phylogeny of the plant pathogen *Ralstonia solanacearum* inferred from gene distribution analysis. *J. Bacteriol.* **189**, 377–387 (2007).
11. Aparicio, S. *et al.* Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**, 1301–1310 (2002).
12. Johnson, Z. I. *et al.* Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science* **311**, 1737–1740 (2006).
13. Hunt, D. E. *et al.* Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**, 1081–1085 (2008).
14. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502 (1997).
15. Doolittle, W. F. Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128 (1999).
16. Gogarten, J. P., Doolittle, W. F. & Lawrence, J. G. Prokaryotic evolution in light of gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
17. Kurland, C. G., Canback, B. & Berg, O. G. Horizontal gene transfer: a critical view. *Proc. Natl Acad. Sci. USA* **19**, 9658–9662 (2003).
18. Fierer, N., Bradford, M. A. & Jackson, R. B. Toward an ecological classification of soil bacteria. *Ecology* **88**, 1354–1364 (2007).
19. Philippot, L. *et al.* Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ. Microbiol.* **11**, 1518–1526 (2009).
20. Wu, D. *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056–1060 (2009).
21. Eisen, J. A. & Fraser, C. M. Phylogenomics: Intersection of evolution and genomics. *Science* **300**, 1706–1707 (2003).
22. Sicheritz-Ponten, T. & Andersson, S. G. E. A phylogenomic approach to microbial evolution. *Nucleic Acids Res.* **29**, 545–552 (2001).
23. Snel, B., Bork, P. & Huynen, M. A. Genome phylogeny based on gene content. *Nature Genet.* **21**, 108–110 (1999).
24. Dutilh, B. E., Snel, B., Ettema, T. J. G. & Huynen, M. A. Signature genes as a phylogenomic tool. *Mol. Biol. Evol.* **25**, 1659–1667 (2008).
25. Mulikjanian, A. Y. *et al.* The cyanobacterial genome core and the origin of photosynthesis. *Proc. Natl Acad. Sci. USA* **103**, 13126–13131 (2006).
26. Gao, B., Paramanathan, R. & Gupta, R. S. Signature proteins that are distinctive characteristics of Actinobacteria and their subgroups. *Antonie Van Leeuwenhoek* **90**, 69–91 (2006).
27. Gupta, R. & Lorenzini, E. Phylogeny and molecular signatures (conserved proteins and indels) that are specific for the Bacteroidetes and Chlorobi species. *BMC Evol. Biol.* **7**, 71 (2007).
28. Koonin, E. G. & Wolf, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008).
29. Ettema, J. G. & Andersson, S. G. E. The α -proteobacteria: the Darwin finches of the bacterial world. *Biol. Lett.* **5**, 429–432 (2009).
30. Bousseau, B., Karlberg, E. O., Frank, A. C., Legault, B. A. & Andersson, S. G. E. Computational inference of scenarios for α -proteobacterial genome evolution. *Proc. Natl Acad. Sci. USA* **101**, 9722–9727 (2004).
31. Williams, K. P., Sobral, B. W. & Dickerman, A. W. A robust species tree for the Alphaproteobacteria. *J. Bacteriol.* **189**, 4578–4586 (2007).
32. Smit, E. *et al.* Diversity and seasonal fluctuations of the dominant members of the bacterial soil community in a wheat field as determined by cultivation and molecular methods. *Appl. Environ. Microbiol.* **67**, 2284–2291 (2001).
33. Cruz-Martinez, K. *et al.* Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *ISME J.* **3**, 738–744 (2009).
34. Thomson, B. C. *et al.* Vegetation affects the relative abundances of dominant soil bacterial taxa and soil respiration rates in an upland grassland soil. *Microb. Ecol.* **59**, 335–343 (2009).
35. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
36. Wallenstein, M. D., McMahon, S. & Schimel, J. Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. *FEMS Microbiol. Ecol.* **59**, 428–435 (2007).
37. Jones, R. T. *et al.* A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* **3**, 442–453 (2009).
38. Männistö, M. K., Tiirola, M. & Häggblom, M. Bacterial communities in Arctic fields of Finnish Lapland are stable but highly pH-dependent. *FEMS Microbiol. Ecol.* **59**, 452–465 (2007).
39. Hartman, W. H., Richardson, C. J., Vilgalys, R. & Bruland, G. L. Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proc. Natl Acad. Sci. USA* **105**, 17842–17847 (2008).
40. Ward, N. L. *et al.* Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* **75**, 2046–2056 (2009).
41. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810 (2002).
42. Rusch, D. B. *et al.* The *Socerer II* Global Ocean Sampling expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
43. Danon, M., Franke-Whittle, I. H., Insam, H., Chen, Y. & Hadar, Y. Molecular analysis of bacterial community succession during prolonged compost curing. *FEMS Microbiol. Ecol.* **65**, 133–144 (2008).
44. Fulthorpe, R. R., Roesch, L. F. W., Riva, A. & Triplett, E. W. Distantly sampled soils carry few species in common. *ISME J.* **2**, 901–910 (2008).
45. Hackl, E., Zechmeister-Boltenstern, S., Bodrossy, L. & Sessitsch, A. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. *Appl. Environ. Microbiol.* **70**, 5057–5065 (2004).
46. Pointing, S. B. *et al.* Highly specialized microbial diversity in hyper-arid polar desert. *Proc. Natl Acad. Sci. USA* **106**, 19964–19969 (2009).
47. Buckley, D. H. & Schmidt, T. M. The structure of microbial communities in soil and the lasting impact of cultivation. *Microb. Ecol.* **42**, 11–21 (2001).
48. Nemerugut, D. R. *et al.* The effects of chronic nitrogen fertilization on alpine tundra soil microbial communities: implications for carbon and nitrogen cycling. *Environ. Microbiol.* **10**, 3093–3105 (2008).
49. Jangid, K. *et al.* Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil Biol. Biochem.* **40**, 2843–2853 (2008).
50. Mou, X. Z., Sun, S. L., Edwards, R. A., Hodson, R. E. & Moran, M. A. Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**, 708–711 (2008).
51. Goffredi, S. K. & Orphan, V. J. Bacterial community shifts in taxa and diversity in response to localized organic loading in the deep sea. *Environ. Microbiol.* **12**, 344–363 (2010).
52. Mering, C. V. *et al.* Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**, 1126–1130 (2007).
53. Glockner, F. O., Fuchs, B. M. & Amann, R. Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl. Environ. Microbiol.* **65**, 3721–3726 (1999).
54. Zwart, G., Crump, B. C., Agetvereld, M., Hagen, F. & Han, S. K. Typical freshwater bacteria: an analysis of available 16S rRNA gene sequences from plankton of lakes and rivers. *Aquat. Microb. Ecol.* **28**, 141–155 (2002).
55. Bouvier, T. C. & del Giorgio, P. A. Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries. *Limnol. Oceanogr.* **47**, 453–470 (2002).
56. Steger, K., Jarvis, A., Vasara, T., Romantschuk, M. & Sundh, I. Effects of differing temperature management on development of Actinobacteria populations during composting. *Res. Microbiol.* **158**, 617–624 (2007).
57. Hartman, A. L. *et al.* Human gut microbiome adopts an alternative state following small bowel transplantation. *Proc. Natl Acad. Sci. USA* **106**, 17187–17192 (2009).
58. Gilbert, J. A. *et al.* The seasonal structure of microbial communities in the Western English Channel. *Environ. Microbiol.* **11**, 3132–3139 (2009).
59. Chauhan, A., Cherrier, J. & Williams, H. N. Impact of sideways and bottom-up control factors on bacterial community succession over a tidal cycle. *Proc. Natl Acad. Sci. USA* **106**, 4301–4306 (2009).
60. Vogel, T. M. *et al.* TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Rev. Microbiol.* **7**, 252 (2009).
61. Cohan, F. M. & Perry, E. B. A systematics for discovering the fundamental units of bacterial diversity. *Curr. Microbiol.* **17**, R373–R386 (2007).
62. Konstantinidis, K. T. & Tiedje, J. M. Genomic insights that advance the species definition for prokaryotes. *Proc. Natl Acad. Sci. USA* **102**, 2567–2572 (2005).
63. Ward, D. M. *et al.* Genomics, environmental genomics and the issue of microbial species. *Hereditas* **100**, 207–219 (2008).
64. Cohan, F. M. & Koeppl, A. F. The origins of ecological diversity in prokaryotes. *Curr. Biol.* **18**, R1024–R1034 (2008).
65. Baldauf, S. L. The tree of life is a tree (more or less). *Trends Ecol. Evol.* **17**, 450–451 (2002).
66. Valentine, J. W. *On the Origin of Phyla* (University of Chicago Press, Chicago, 2004).
67. Mayr, E. *This is Biology: the Science of the Living World* (Harvard University Press, Cambridge, USA, 1997).
68. Trosvik, P., Stenseth, N. C. & Rudi, K. Convergent temporal dynamics of the human infant gut microbiota. *ISME J.* **3**, 151–158 (2009).
69. Loreau, M. *et al.* Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* **294**, 804–808 (2001).
70. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: Human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
71. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).

Acknowledgements

L.P. gratefully acknowledges the Environment and Agronomy research division and the International Relations Department of the Institut National de la Recherche Agronomique, France, and the Department of Microbiology of the Swedish University of Agricultural Sciences in Uppsala for supporting and hosting his sabbatical in Sweden. T.J.B. was supported by the Lake Ecosystems Response to Environmental Change (LEREC) programme (which is funded by the Swedish Research Council Formas) while at the University of Uppsala. S.H. is supported by the Swedish Research Council and the Swedish Research Council Formas, which finance the Uppsala Monitoring Centre, and S.G.E.A. is supported by the Swedish Research Council, the Göran Gustafsson Foundation, the Swedish Foundation for Strategic Research and the Knut and Alice Wallenberg Foundation. J.P. acknowledges support from the Royal Society of Edinburgh, and W.B.W. from the US Department of Agriculture and the US National Science Foundation.

Competing interests statement

The authors declare no competing financial interests.

DATABASES

Entrez Genome Project: <http://www.ncbi.nlm.nih.gov/genome/prj>
[Escherichia coli](http://www.ncbi.nlm.nih.gov/genome/prj/Escherichia_coli) | [Ralstonia solanacearum](http://www.ncbi.nlm.nih.gov/genome/prj/Ralstonia_solanacearum) | [Takifugu rubripes](http://www.ncbi.nlm.nih.gov/genome/prj/Takifugu_rubripes)

FURTHER INFORMATION

Laurent Philippot's homepage: <http://www2.dijon.inra.fr/umrme/spip.php?article30>
 GEBA: <http://www.jgi.doe.gov/programs/GEBA>
 NCBI Microbial Genomes: http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html
 Genomes OnLine: <http://www.genomesonline.org>
 JCVI Comprehensive Microbial Resource: <http://cmr.jvri.org/tigr-scripts/CMR/CMRHomePage.cgi>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF