RESEARCH ARTICLE

# Lineage-dependent ecological coherence in bacteria

Alexander F. Koeppel & Martin Wu

Department of Biology, University of Virginia, Charlottesville, VA, USA

**Correspondence:** Martin Wu, Department of Biology, University of Virginia, 404 Physical and Life Science Building, 90 Geldard Drive, Charlottesville, VA 22904-4328, USA. Tel.: +1 434 924 4518; fax: +1 434 982 5626; e-mail: mw4yv@virginia.edu

Received 28 October 2011; revised 28 March 2012; accepted 2 April 2012.

DOI: 10.1111/j.1574-6941.2012.01387.x

Editor: Julian Marchesi

**Keywords**
ecological coherence; human microbiota; taxonomic hierarchy.

## Abstract

Bacteria comprise an essential element of all ecosystems, including those present on and within the human body. Understanding bacterial diversity therefore offers enormous scientific and medical benefit, but significant questions remain regarding how best to characterize that diversity and organize it into biologically meaningful units. Bacterial communities are routinely characterized based on the relative abundances of taxa at the genus or even the phylum level, but the ecological coherence of these high-level taxonomic units is uncertain. Using human microbiota from the skin and gut as our model systems, we tested the ecological coherence of bacteria by investigating the habitat associations of bacteria at all levels of the taxonomic hierarchy. We observed four distinct taxonomic patterns of habitat association, reflecting different levels of ecological coherence among taxa. Our results support the hypothesis that deep-branch bacterial clades could be ecologically coherent and suggest that the phylogenetic depth of ecological coherence varies among the bacterial lineages and is an important factor to consider in studies of human microbiome associations.

## Introduction

Bacteria represent one of the greatest remaining frontiers of undiscovered and uncharacterized diversity on Earth. Most of bacterial diversity is known only from DNA sequences as the vast majority of bacteria have not been cultured (Giovannoni & Stingl, 2005) A key step in studying microbial ecology and evolution is developing methods to organize DNA sequences into biologically and ecologically meaningful taxonomic units. A critical question therefore is at what taxonomic level do bacteria display ecological coherence, i.e. share similar lifestyles or traits that distinguish them from members of other taxa (Philippot *et al.*, 2010).

Recent studies have shown evidence indicating ecological coherence among higher bacterial taxonomic ranks (Hackl *et al.*, 2004; Lozupone & Knight, 2005, 2007; Fierer *et al.*, 2007; von Mering *et al.*, 2007; Danon *et al.*, 2008; Fulthorpe *et al.*, 2008; Jones *et al.*, 2009; Philippot *et al.*, 2009; Pointing *et al.*, 2009). For instance, within the class *Alphaproteobacteria* the order *Rickettsiales* appears to be adapted to life inside animal host cells (Ettema & Andersson, 2009). The presence of the phylum *Acidobacteria* in soil is correlated with low pH (Jones *et al.*, 2009), indicating a possible phylum-wide adaptation to acidic soil conditions. Phylum-level bacterial spatial patterns have also been observed in soils subjected to different farming practices (Philippot *et al.*, 2009). There is also evidence of ecological associations with very deep branches of the bacterial lineage, including ancient divergences between organisms adapted to high and low levels of salinity, as well as between taxa inhabiting aqueous, soil and sediment habitats (Lozupone & Knight, 2005, 2007). A large-scale comparison of microbial sequences sampled from natural environments revealed statistically significant associations between habitats and taxonomic ranks up to the order level (von Mering *et al.*, 2007).

When studying higher order taxa in bacteria, it is not always clear whether their habitat associations reflect the ecology of most or even many of the species contained within them. It is possible that a few highly abundant species could make it appear that the taxon as a whole was strongly associated with a particular habitat, even if the majority of the species within it had no such association. Interpreting the results of habitat association studies requires a better understanding of how uniformly associations among higher order taxa are reflected among the member species. However, the ecological coherence of taxa above the species level has not been explicitly tested (Philippot *et al.*, 2010), leaving open the question of how





FEMS MICROBIOLOGY ECOLOGY

best to characterize and compare the diversity of bacterial communities.

Determining the depth of ecological coherence should also elucidate the origins of microbial ecological diversity. It has been established that not all ecological transitions are equally probable for bacterial populations (Lozupone & Knight, 2007; Cohan & Koeppel, 2008; Martiny *et al.*, 2009). Ecological coherence in deep lineages should indicate rare and early evolutionary transitions across habitats while ecological coherence at lower taxonomic levels should suggest more frequent and recent habitat transitions.

The primary aim of this study was to investigate the depth of ecological coherence in the bacterial taxonomic hierarchy. We chose to analyse sequence data from the human microbiome for the following reasons. First, examining the depth of ecological coherence in bacteria has clear implications for ongoing human microbiome research. A key focus of human microbiome studies is whether differences in the microbial communities can be related to health and disease states (Turnbaugh *et al.*, 2007). This is commonly done using association studies, in which the relative abundances of higher order taxa, such as genera, orders or phyla are compared between cases and healthy controls (Grice *et al.*, 2009; Hartman *et al.*, 2009; Turnbaugh *et al.*, 2009; Dominguez-Bello *et al.*, 2010). Association studies have shown, for example, that shifts in higher order community structure can be associated with important human health conditions, including diabetes (Giongo *et al.*, 2011), obesity (Ley, 2005; Ley *et al.*, 2006; Cani *et al.*, 2007), cancer (Turnbaugh *et al.*, 2007) and cardiovascular disease (Ordovas & Mooser, 2006). Our work addresses the following question: Do these putative higher order associations truly reflect taxon-wide ecological properties, or might they instead reflect the habitat associations of a few highly abundant lower level taxa contained within them? We hypothesize that the latter will be true for high-order associations in at least some taxa.

The second advantage of using human microbiome data for studying ecological coherence is the availability of deeply sequenced datasets. Deep sequencing is critical for determining the taxonomic level of ecological coherence (Vogel *et al.*, 2009; Philippot *et al.*, 2010). As we divide up taxa into units of lower and lower ranks, the number of sequences that belong to each unit will keep diminishing. Deep sequencing is therefore necessary for meaningful statistical analysis of the minority and lower order taxa. Due to limited sampling (inadequate efforts or techniques) (Philippot *et al.*, 2010), some of the earlier habitat-association studies were limited either in scope, because only the dominant members were characterized, or in depth, because of lack of resolution at low taxonomic ranks. Consequently in these studies, the evidence for high-level associations is not conclusive.

Higher order habitat associations have been reported for human skin (Grice *et al.*, 2009) and gut (Ley *et al.*, 2006) microbiota. Grice *et al.* found that the abundance of higher order taxa varied with skin type. Specifically, they noted that different phyla and genera were in greatest abundance on moist skin, dry skin and sebaceous skin. Ley *et al.* observed an increase in the abundance of the phylum *Bacteroidetes* and a corresponding decrease in *Firmicutes* abundance as subjects lost weight. Both datasets offer deep sequencing of nearly full-length 16S rRNA genes, making it possible to characterize bacterial diversity at relatively fine scales. Using these two datasets, we investigated the depth of ecological coherence by testing whether patterns observed at higher taxonomic ranks are maintained at lower ranks. We performed a comprehensive test of the ecological coherence of taxa by comparing the habitat associations along the entire hierarchy, from the phylum level all the way to species [as approximated by operational taxonomic units (OTUs)].

## Materials and methods

### Sequence datasets

Two 16S rRNA gene sequence datasets were analysed in this study. The skin microbiome study (Grice *et al.*, 2009) sampled skin bacteria from 10 healthy volunteers, each of which were sampled at 21 different skin sites, including moist, dry and sebaceous skin. This set contained 116 391 sequences. The gut microbiome study (Ley *et al.*, 2006) sampled gut bacteria from 12 obese individuals and two lean controls over a time course of 52 weeks during which the obese subjects undertook one of two weight-loss regimens. This dataset contained 18 052 sequences. Both datasets consisted of Sanger sequenced, near full-length (minimum of 1250 bp) 16S rRNA gene sequences. The two datasets were retrieved from GenBank, along with the metadata for each sequence.

### Sequence alignment, classification and clustering

16S rRNA gene sequences were aligned using the Ribosomal Database Project (RDP) and classified to the genus level using RDP Classifier, version 2 (Wang *et al.*, 2007), set at the default parameters. Sequences from the same genera (as identified by RDP) were further subdivided into OTUs by complete-linkage clustering using MOTHUR (Schloss *et al.*, 2009). OTUs were generated using sequence identity cut-offs of 97% and 99%. We chose

these cut-offs because they are commonly used and accepted approximations of bacterial species (Stackebrandt & Ebers, 2006; Fierer *et al.*, 2008; Hamady & Knight, 2009).

### Analysis of habitat associations

Taxa at all levels were checked against the available metadata to determine their habitat associations. The skin data were tested for associations with a particular skin type (moist, dry or sebaceous skin; see Grice *et al.*, 2009 for classification of 21 skin sites). For each taxon, habitat associations were assessed using the indicator value index, using the IndVal metric (Dufrene & Legendre, 1997). IndVal is a powerful tool for detecting statistically significant associations between taxa and habitats, and has been used to detect such associations in microbial taxa (Hartmann *et al.*, 2009; Auguet *et al.*, 2010; Cardenas *et al.*, 2010). The inputs for the INDVAL software for a given taxon were the counts of sequences belonging to that taxon at each skin site on each subject. Counts were normalized by the total sequencing effort for each subject at each skin site. The value of the IndVal index is computed by factoring in both the specificity of a given taxon to each site group (i.e. how abundant the taxon is in the target group compared with other groups) and the fidelity of the taxon to the group (i.e. the fraction of samples within the group containing the taxon). The software then tests the significance of the IndVal by comparing the observed IndVal against a number of permutations (10 000 in our case) in which the samples are randomly reallocated among groups. Significance is determined both by a *t*-test comparing the observed result with the mean result from the permutations, and by the rank order of the observed IndVal among the permutation IndVals. Only associations that were significant at the $P < 0.05$ level by both tests were considered statistically significant in our study.

The sequences of the gut dataset were tested for associations with both lean subjects and obese subjects. The obese group was defined as the set of 12 obese subjects at the starting time point (before any weight loss). The lean group consisted of the same 12 subjects sampled 52 weeks later (after weight loss), combined with all sequences sampled from the two lean control subjects. The same IndVal analysis was used to detect associations and determine their significance.

We tested the ecological coherence of taxa in each dataset by determining whether the habitat associations found for the entire taxon were shared by subtaxa within it. Taxa were analysed hierarchically from the phylum level down to the species level. To obtain meaningful statistical results, taxa at each taxonomic rank down to the genus level were analysed only if they contained at least 100 sequences. OTUs were analysed if they contained at least 50 sequences. High-level taxa (phylum, class, order and family) containing at least 500 sequences were subdivided to the next taxonomic rank. Only the most abundant genera in each phylum were further subdivided into OTUs.

## Results

### Habitat associations among phyla

As reported previously (Grice *et al.*, 2009) the skin microbiome was dominated by the phylum *Actinobacteria*, with the phyla *Firmicutes*, *Proteobacteria* and *Bacteroidetes* also showing high relative abundance (Fig. 1a). In agreement with previous results, our statistical analyses revealed significant associations between the phyla *Actinobacteria*, *Firmicutes* and *Proteobacteria* and sebaceous, moist and dry skin, respectively. Our phylum-level analysis of the gut dataset was also in agreement with the findings of the original study. The *Firmicutes* was the dominant phylum overall, and showed a statistically significant association with the obese group. The *Bacteroidetes* were also highly abundant, and this phylum was significantly associated with the lean group of subjects (Fig. 1b).

We then followed the phylum-level analyses by testing the associations of the lower taxa within each phylum. These hierarchical analyses allowed us to determine whether the phylum-level association persisted throughout each phylum, or instead reflected the underlying pattern of its dominant members. We were thereby able to determine the taxonomic depth of ecological coherence.

### Hierarchical analysis of *Actinobacteria* skin type associations

In the phylum *Actinobacteria*, the association of the entire phylum to sebaceous skin was maintained down through the rank of order (Fig. 2a, Supporting Information, Table S1). The phylum is represented by only a single class in this dataset (class *Actinobacteria*) and is dominated by a single order that contains 97.9% of the sequences in the phylum (order *Actinomycetales*). Not surprisingly, these taxa shared the whole-phylum association with sebaceous skin (colored in red). Interestingly, however, the pattern started to diverge at the family level. Two large families, the *Propionibacteriaceae* and *Corynebacteriaceae*, dominated *Actinomycetales*, comprising 47.0% and 51.3% of the sequences, respectively. *Propionibacteriaceae* shared the whole-phylum association with sebaceous skin, but

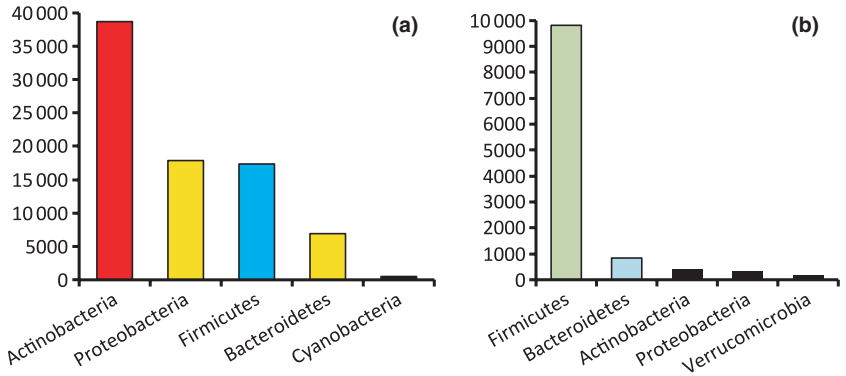


**Fig. 1.** Habitat associations of phyla. Rank abundance graphs of the dominant phyla in each dataset, color-coded by habitat associations. Bars were colored only if the association was statistically significant ($P < 0.05$). The skin dataset of Grice *et al.* (a) is coded as follows: red, significantly associated with sebaceous skin; blue, significantly associated with moist skin; yellow, significantly associated with dry skin; black, no significant associations. The gut dataset of Ley *et al.* (b) is coded as follows: light green, significantly associated with obese subjects; light blue, significantly associated with lean subjects; black, no significant associations.

strikingly *Corynebacteriaceae* showed a statistically significant association with moist skin (colored in blue). *Propionibacterium* and *Corynebacterium*, by far the most abundant genera both within their families and in the dataset as a whole, were significantly associated with sebaceous and moist skin, respectively. This result is in agreement with the findings of Grice *et al.* (2009), who observed that *Corynebacterium* was the most abundant genus on moist skin, and that *Propionibacterium* was the most abundant genus on sebaceous skin.

Greater variation in habitat associations was observed for 97% of the OTUs. Of 47 *Corynebacterium* OTUs containing 50 or more sequences, only six shared the whole-genus association with moist skin. For the majority of the 97% OTUs, no significant skin type association was detected (Fig. 2a). One minority member was actually associated with dry skin (see Fig. 2a, inset). The apparent habitat association of the whole genus largely reflected the patterns of a few abundant members. In the case of *Propionibacterium*, all but one member shared the association with sebaceous sites found for the genus as a whole (Fig. 2a), although there were only five OTUs containing more than 50 sequences in this genus.

To investigate the ecological coherence of the taxonomic unit at a finer level, we further divided each 97% OTU into its constituent 99% OTUs (Fig. 2, bottom layer,). When subdivided, the two most abundant *Corynebacterium* 97% OTUs were sorted into heterogeneous groups. While some of the more abundant 99% OTUs shared their parent taxon's association patterns, many OTUs showed no association. For *Propionibacterium*, the association with sebaceous skin was more consistent among 99% OTUs. Ecological coherence at the finer taxonomic scale appears to be lineage-dependent.

## Hierarchical analysis of *Proteobacteria* and *Firmicutes* skin type associations

The hierarchical analysis of *Proteobacteria* returned very different results. In this phylum, the association with dry skin appeared to derive from the highly abundant order *Burkholderiales*, in which the association was maintained all the way to the 99% OTU level (Fig. 2b, Table S2). The only other abundant order in this phylum was *Neisseriales*, which showed no significant association with any skin type, although the genus *Neisseria* within that order was associated with dry skin.

*Firmicutes* showed another distinct result. The phylum as a whole showed an association with moist skin, and *Bacilli*, the phylum's most abundant class, shared the phylum-level association with moist skin. Intriguingly, the subtaxa within *Bacilli* lacked association with any skin type all the way to the genus level. However, several 97% OTUs within *Staphylococcus* showed associations to either moist or sebaceous skin (Fig. 2c, Table S3). In this case the mix of different associations at the 97% OTU level explains the lack of association at the higher levels. The hierarchical analysis of 99% OTUs in *Staphylococcus* returned similar results to those of *Corynebacterium*, with some taxa predicting the associations of their subgroups but others not. These examples further reinforce our interpretation of the *Actinobacteria* results, i.e. that the taxonomic depth of ecological coherence with respect to these particular habitats is lineage-specific.

## Hierarchical analysis of obese and lean gut associations

We performed the same hierarchical analysis on the gut microorganism dataset. Nearly all the *Firmicutes* sequences

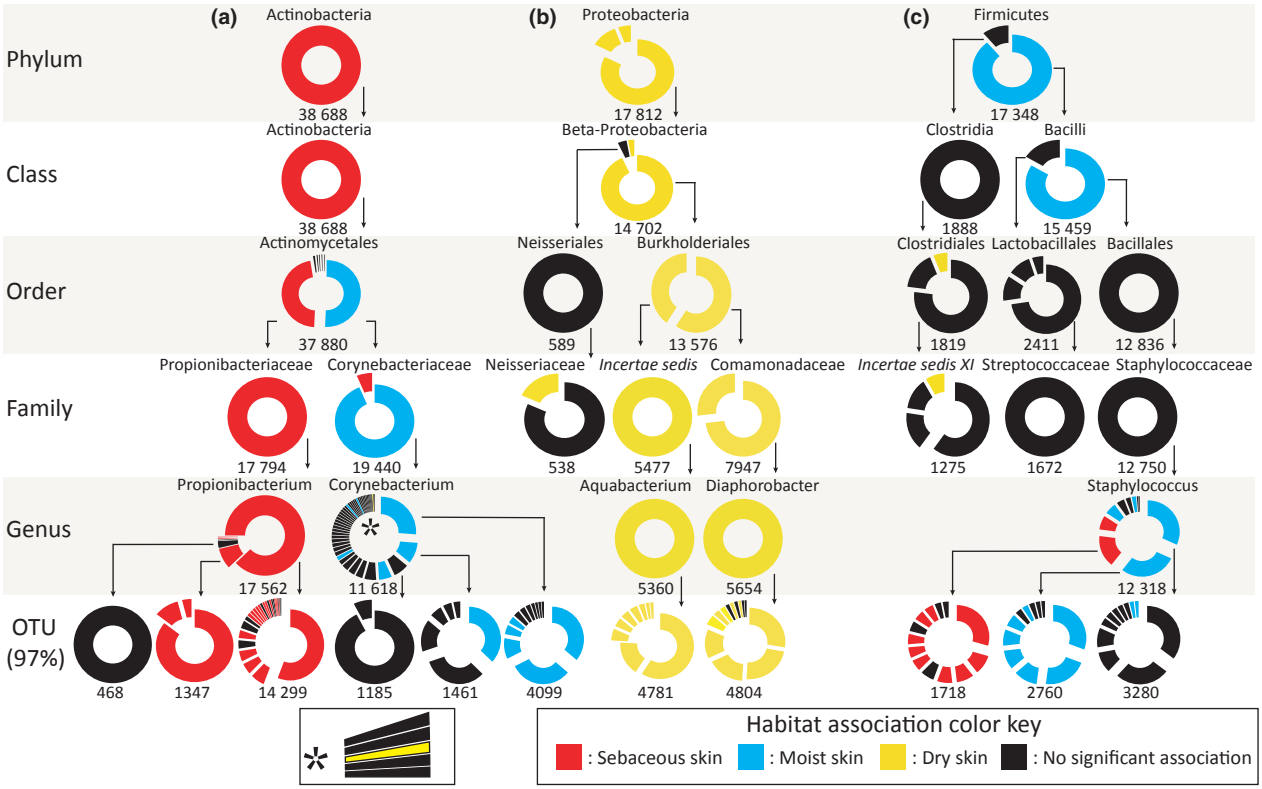


**Fig. 2.** Hierarchical analyses of skin habitat associations. Each doughnut graph represents a taxon of the rank specified for its row. Each of the wedges in the doughnut represents an abundant (containing at least 100 sequences for higher taxa, at least 50 in the case of OTUs) subtaxon, with the wedge size corresponding to that subtaxon's relative abundance. Subtaxa are sorted clockwise by their relative abundance. Example: in (a), the graph for the whole phylum is an unbroken circle because all of the sequences in the phylum belong to a single class. The graph for the order *Actinomycetales* is split into seven wedges because there are seven abundant families in that order. Two of those families (the *Corynebacteriaceae* and *Propionibacteriaceae*) are much more abundant than the other five and so are represented by larger wedges. Numbers below each graph display the number of sequences represented in the graph, not the total number of sequences classified to that taxon (i.e. taxa with fewer that 100 sequences are not in the graph, and so are not included in the total). (a) *Actinobacteria*, (b) *Proteobacteria*, (c) *Firmicutes*. Graphs in the top row represent higher order taxa, while the bottom row shows their division into taxa of the next lower rank. The wedges of all graphs are color coded to reflect the habitat associations of the taxa represented (see key). The asterisk in Fig. 1a denotes the location represented by the insert at the bottom of the figure, showing that one OTU in the genus *Corynebacterium* was associated with dry skin.

in this dataset (96.36%) belonged to the order *Clostridiales* (Fig. 3a). As the class *Bacilli*, the only other abundant taxon, lacked an association, we conclude that it is in fact the order *Clostridiales* and not the entire phylum *Firmicutes* that is associated with the obese gut environment. The association with obese subjects seems to disappear below the rank of order. None of the individual families showed a statistically significant difference in relative abundance between the lean and obese groups (Fig. 3a, Table S4). Of the abundant genera in this phylum, only the genus *Dorea* showed a significant association with the obese group. As this genus is not particularly abundant, however, it is unlikely to explain the apparent association of the entire order. Among *Bacteroidetes*, all abundant taxa from phylum down to family showed a strong association with the lean group (Fig. 3b, Table S4). The genus *Bacteroides* shared the association with the lean group, but the genus *Prevotella* showed no association with either the lean or the obese group. No individual 97% or 99% OTUs from this phylum were abundant enough to meet our criteria for statistical analysis, and so analysis was stopped at the genus level.

## Discussion

Our hierarchical analysis of habitat associations revealed four broad patterns, described below, and illustrated in Fig. 4:

**(1)** Habitat association of a taxon is maintained among all or nearly all of its constituent subtaxa (for example,

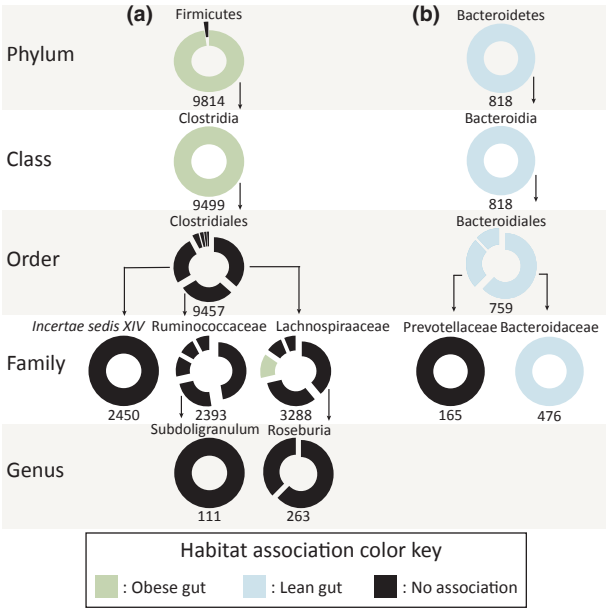


**Fig. 3.** Hierarchical analyses of gut habitat associations. Graphs are formatted exactly as in Fig. 2, but with different color-coding to reflect gut habitat associations (see key).

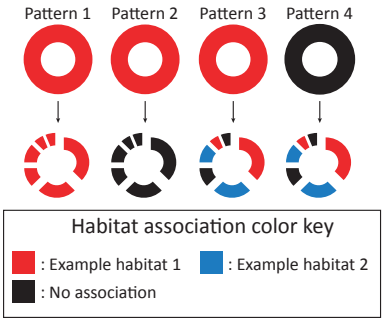


**Fig. 4.** Patterns of habitat association. A schematic diagram of the four overall patterns of habitat associations observed among these taxa. Graphs in the top row represent higher order taxa, while the bottom row shows their division into taxa of the next lower rank.

the associations of the phylum *Bacteroidetes* with lean subjects, the order *Burkholderiales* with dry skin and the family *Propionibacteriaceae* with sebaceous skin).

**(2)** Parent taxon has a significant habitat association but few or none of its subtaxa have any associations at all (for example, the association of the order *Clostridiales* with obesity).

**(3)** Habitat association of a taxon is shared by only a few of its constituent subtaxa. Most subtaxa have either no association or their associations are different from that of the parent taxon (for example, the association of the genus *Corynebacterium* with moist skin).

**(4)** Parent taxon has no habitat association but many of its subtaxa do have significant habitat associations (for example, the genus *Staphylococcus*, which itself has no association with any skin type, but contains numerous OTUs that are associated with either moist skin, or sebaceous skin).

It has been shown that in macroorganisms both spatial and taxonomic scales influence the observed phylogenetic structure of community assemblages (Keddy & Weiher, 1999; Silvertown *et al.*, 2001, 2006; Cavender-Bares *et al.*, 2004, 2006). Habitat filtering and competitive exclusion are two important processes central to the assembly of communities. They can occur simultaneously in real communities but their relative predominance depends on the scales of the communities under study. Habitat filtering, in which closely related species share habitats due to inherited ecological similarity, is thought to play an important role in community assembly (Weiher & Keddy, 1995; Webb *et al.*, 2002) and tends to dominate on broader spatial and phylogenetic scales whereas competitive exclusion is more important in local communities of closely related species (Cavender-Bares *et al.*, 2006). Our analyses revealed similar trends in the phylogenetic structures of bacterial communities. For example, at higher taxonomic ranks both pattern-1 and pattern-2 taxa show habitat associations. This is consistent with the idea that they may possess traits originating deep in the history of their lineage that confer similar ecological properties, resulting in niche conservation and habitat filtering, as observed previously (Horner-Devine & Bohannan, 2006; Newton *et al.*, 2007; Bryant *et al.*, 2008; Auguet *et al.*, 2010; Barberán *et al.*, 2012). Pattern-1 taxa meet our definition of ecological coherence, because associations are maintained through the taxonomic hierarchy. However, there are also reasons to consider pattern-2 taxa as ecologically coherent. Pattern-2 taxa could reflect an overall preference of a taxon for a particular habitat due to habitat filtering, but one that, although significant, is not strong enough to be detectable among its subtaxa, and so is only apparent when taxa are viewed in aggregate.

It is by no means our intention to suggest that these pattern-1 and pattern-2 taxa are homogeneous over all aspects of their ecological niche, only that they display coherence with respect to certain specific ecological parameters examined in this study. Ecological coherence is a relative term when applied to higher order taxa. In most cases, a taxon will be coherent with respect to some ecological factors but not to others. In datasets with multiple ecological factors to consider, each could be assessed independently. Alternatively, information from multiple ecological factors could be incorporated into more complex definitions of habitat using software such as ADAPTML (Hunt *et al.*, 2008). Taxa could then be

tested for their associations with these aggregate habitats.

We argue that pattern-3 or pattern-4 taxa are not ecologically coherent with respect to the ecological parameters represented by these habitats because the habitat associations of the subtaxa are heterogeneous. Within these taxa, transitions between habitats among species seem to have taken place relatively recently. In pattern-3, the parent taxon's habitat association reflects those of a few highly abundant subtaxa, and not necessarily the majority of its members. Although heterogeneity was observed at different levels of organization, in both high- and low-level taxa, it is of particular interest to note that habitat associations appear to diversify at the 97% OTU level. For example, both the genera *Corynebacterium* and *Staphylococcus* contained members with significant associations distinct from those of the whole genus. Previously reported results (Ward *et al.*, 2006; Goris *et al.*, 2007; Hunt *et al.*, 2008; Koeppel *et al.*, 2008; Connor *et al.*, 2010) have shown that ecologically important diversity can also exist below the level of traditionally defined species.

The observation of pattern-3 or pattern-4 for a taxon may indicate that the taxon is evolutionarily labile with respect to the habitat under study. For example, our results for the class *Bacilli* suggest that transitions between moist and sebaceous skin habitats are routine in that taxon. Transitions between these habitats appear to have occurred in the recent past, because many of the 97% OTUs showed different associations, even within the same genus. By contrast, in the class *Actinobacteria*, the same transition appears to be more difficult, with one major shift from sebaceous skin to moist skin at the family level. Past studies of bacterial systems have shown similar results. For example, in marine *Prochlorococcus* it has been shown that the transition between high-light and low-light habitats occurs at greater phylogenetic depth than the transition between different temperature optima, which in turn occurs at greater depth than transitions between habitats with different nitrate concentrations (Martiny *et al.*, 2009).

Our results go further in one respect, by showing that the taxonomic depth of ecological coherence for the same classes of habitat is lineage-specific. For some lineages, deeper clades of various ranks possess ecological coherence, but for others, even genera appear to lack ecological coherence. We found evidence of coherence above the species rank, at the phylum (e.g. *Bacteroidetes*), order (e.g. *Burkholderiales*) and family (e.g. *Propionibacteriaceae*) levels. Our study supports the notion that high-level bacterial taxa can be ecologically relevant (Philippot *et al.*, 2010). In fact, for pattern-2 taxa, because the ecologically meaningful patterns are discernible only at high taxonomic ranks, we argue that analysis of high-level bacterial community structure is not only useful but also necessary for predicting ecosystem function. While in other cases (e.g. pattern-3 and pattern-4), the apparent ecological associations of a higher taxon are not always predictive of the associations of the subtaxa within them. One explanation for this observation may be that the criteria for defining higher order taxa are themselves arbitrary. What is designated a genus or family within one lineage might encompass as much diversity as a family or order in another. Regardless of the reason, it is clear that association studies that use only high-order taxa risk missing the most important features of a community's diversity and can even be misleading.

These findings challenge the methods by which communities of human microbiota are currently characterized in association studies. Such studies generally compare communities based on the relative abundances of taxa at one level, usually at or above the genus. Our results suggest that the choice of taxonomic level used to define the operational unit is a very important factor to consider. To truly understand whether and how the composition of the microbiome affects human health, it is necessary to dissect the biological patterns along the entire taxonomic hierarchy, including the very fine scales of microdiversity. This will not only help us tease biological meaningful taxa apart from the 'hitchhikers', but also elucidate the phylogenetic depth where the ecological selection is acting.

The hierarchical analysis proposed here requires sequences to be sampled from a reasonable number of individuals and also with a reasonable depth of coverage. This can be readily achieved with next-generation sequencing technology (Kuczynski *et al.*, 2010). Another important consideration is the amount of metadata to collect. Our statistical analyses of the skin data showed that although many taxa are associated with particular skin types, many others showed no associations. This is probably because skin moisture level is only one of many potential ecological parameters on human skin (Grice & Segre, 2011). Having access to more ecological data will certainly refine our ability to determine the ecological coherence of taxa.

In conclusion, our results demonstrate that the taxonomic depth of ecological coherence with respect to any particular trait can vary greatly between lineages. To get a complete picture of the structure of a bacterial community, it is necessary to perform comprehensive taxonomic analyses as we have done here. Demonstrating associations of higher order taxa alone may be informative depending on whether that taxon is an ecologically coherent unit with respect to the environment being studied.

## Acknowledgements

This research was funded by the Gordon and Betty Moore Foundation (grant no. 1660). We also thank Frederick M. Cohan for feedback on the manuscript.

## References

Auguet J, Barberan A & Casamayor EO (2010) Global ecological patterns in uncultured Archaea. *ISME J* **4**: 182–190.

Barberán A, Bates ST, Casamayor EO & Fierer N (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343–351.

Bryant JA, Lamanna C, Morlon H, Kerkhoff AJ, Enquist BJ & Green JL (2008) Colloquium paper: microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *P Natl Acad Sci USA* **105** (suppl 1): 11505–11511.

Cani PD, Amar J, Iglesias MA *et al.* (2007) Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**: 1761–1772.

Cardenas E, Wu WM, Leigh MB *et al.* (2010) A combined massively parallel sequencing – indicator species approach revealed significant association between sulfate-reducing bacteria and uranium-reducing microbial communities. *Appl Environ Microbiol* **76**: 6778–6786.

Cavender-Bares J, Keen A & Miles B (2006) Phylogenetic structure of Floridian plant communities depends on taxonomic and spatial scale. *Ecology* **87**: 109–122.

Cavender-Bares J, Ackerly DD, Baum DA & Bazzaz FA (2004) Phylogenetic overdispersion in Floridian oak communities. *Am Nat* **163**: 823–843.

Cohan FM & Koeppel AF (2008) The origins of ecological diversity in prokaryotes. *Curr Biol* **18**: R1024–R1034.

Connor N, Sikorski J, Rooney AP *et al.* (2010) Ecology of speciation in the genus *Bacillus*. *Appl Environ Microbiol* **76**: 1349–1358.

Danon M, Franke-Whittle IH, Insam H, Chen Y & Hadar Y (2008) Molecular analysis of bacterial community succession during prolonged compost curing. *FEMS Microbiol Ecol* **65**: 133–144.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N & Knight R (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *P Natl Acad Sci USA* **107**: 11971–11975.

Dufrene M & Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345.

Ettema TJ & Andersson SG (2009) The alpha-proteobacteria: the Darwin finches of the bacterial world. *Biol Lett* **5**: 429–432.

Fierer N, Bradford MA & Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology* **88**: 1354–1364.

Fierer N, Hamady M, Lauber CL & Knight R (2008) The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *P Natl Acad Sci USA* **105**: 17994–17999.

Fulthorpe RR, Roesch LF, Riva A & Triplett EW (2008) Distantly sampled soils carry few species in common. *ISME J* **2**: 901–910.

Giongo A, Gano KA, Crabb DB *et al.* (2011) Toward defining the autoimmune microbiome for type 1 diabetes. *ISME J* **5**: 82–91.

Giovannoni SJ & Stingl U (2005) Molecular diversity and ecology of microbial plankton. *Nature* **437**: 343–348.

Goris J, Konstantinidis K, Klappenbach J, Coenye T, Vandamme P & Tiedje J (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Grice EA & Segre JA (2011) The skin microbiome. *Nat Rev Microbiol* **9**: 244–253.

Grice EA, Kong HH, Conlan S *et al.* (2009) Topographical and temporal diversity of the human skin microbiome. *Science* **324**: 1190–1192.

Hackl E, Zechmeister-Boltenstern S, Bodrossy L & Sessitsch A (2004) Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. *Appl Environ Microbiol* **70**: 5057–5065.

Hamady M & Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* **19**: 1141–1152.

Hartman AL, Lough DM, Barupal DK, Fiehn O, Fishbein T, Zasloff M & Eisen JA (2009) Human gut microbiome adopts an alternative state following small bowel transplantation. *P Natl Acad Sci USA* **106**: 17187–17192.

Hartmann M, Lee S, Hallam SJ & Mohn WW (2009) Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. *Environ Microbiol* **11**: 3045–3062.

Horner-Devine MC & Bohannan BJM (2006) Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**: 100–108.

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ & Polz MF (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.

Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R & Fierer N (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J* **3**: 442–453.

Keddy PA & Weiher E (1999) Introduction: the scope and goals of research on assembly rules. (Keddy PA & Weiher E, eds). *Ecological Assembly Rules: Perspectives, Advances, Retreats.* Cambridge University Press, Cambridge, UK, pp. 1–22.

Koeppel A, Perry EB, Sikorski J *et al.* (2008) Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. *P Natl Acad Sci USA* **105**: 2504–2509.

Kuczynski J, Elizabeth CK, Nemergut DR *et al.* (2010) Direct sequencing of the human microbiome readily reveals community differences. *Genome Biol* **11**: 210.

Ley RE (2005) Obesity alters gut microbial ecology. *P Natl Acad Sci USA* **102**: 11070–11075.

Ley RE, Turnbaugh PJ, Klein S & Gordon JI (2006) Microbial ecology: human gut microbes associated with obesity. *Nature* **444**: 1022–1023.

Lozupone C & Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Lozupone CA & Knight R (2007) Global patterns in bacterial diversity. *P Natl Acad Sci USA* **104**: 11436–11440.

Martiny AC, Tai APK, Veneziano D, Primeau F & Chisholm SW (2009) Taxonomic resolution, ecotypes and the biogeography of *Prochlorococcus*. *Environ Microbiol* **11**: 823–832.

Newton RJ, Jones SE, Helmus MR & McMahon KD (2007) Phylogenetic ecology of the freshwater *Actinobacteria* acI lineage. *Appl Environ Microbiol* **73**: 7169–7176.

Ordovas JM & Mooser V (2006) Metagenomics: the role of the microbiome in cardiovascular diseases. *Curr Opin Lipidol* **17**: 157–161.

Philippot L, Bru D, Saby NPA, Čuhel J, Arrouays D, Šimek M & Hallin S (2009) Spatial patterns of bacterial taxa in nature reflect ecological traits of deep branches of the 16S rRNA bacterial tree. *Environ Microbiol* **11**: 3096–3104.

Philippot L, Andersson SG, Battin TJ, Prosser JI, Schimel JP, Whitman WB & Hallin S (2010) The ecological coherence of high bacterial taxonomic ranks. *Nat Rev Microbiol* **8**: 523–529.

Pointing SB, Chan Y, Lacap DC, Lau MC, Jurgens JA & Farrell RL (2009) Highly specialized microbial diversity in hyper-arid polar desert. *P Natl Acad Sci USA* **106**: 19964–19969.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537–7541.

Silvertown J, Dodd M & Gowing D (2001) Phylogeny and the niche structure of meadow plant communities. *J Ecol* **89**: 428–435.

Silvertown J, Dodd M, Gowing D, Lawson C & McConway K (2006) Phylogeny and the hierarchical organization of plant diversity. *Ecology* **87**: 39–49.

Stackebrandt E & Ebers J (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* **33**: 152–155.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R & Gordon JI (2007) The human microbiome project. *Nature* **449**: 804–810.

Turnbaugh PJ, Hamady M, Yatsunenko T *et al.* (2009) A core gut microbiome in obese and lean twins. *Nature* **457**: 480–484.

Vogel TM, Simonet P, Jansson JK, Hirsch PR, Tiedje JM, van Elsas JD, Bailey MJ, Nalin R & Philippot L (2009) TerraGenome: a consortium for the sequencing of a soil metagenome. *Nat Rev Microbiol* **7**: 252.

von Mering C, Hugenholtz P, Raes J, Tringe SG, Doerks T, Jensen LJ, Ward N & Bork P (2007) Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**: 1126–1130.

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Ward DM, Bateson MM, Ferris MJ, Kuhl M, Wieland A, Koeppel A & Cohan FM (2006) Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philos Trans R Soc Lond B Biol Sci* **361**: 1997–2008.

Webb CO, Ackerly DD, McPeek MA & Donoghue MJ (2002) Phylogenies and community ecology. *Annu Rev Ecol Syst* **33**: 475–505.

Weiher E & Keddy PA (1995) The assembly of experimental wetland plant communities. *Oikos* **73**: 323.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** IndVal results for taxa of the phylum *Actinobacteria*.

**Table S2.** IndVal results for taxa of the phylum *Proteobacteria*.

**Table S3.** IndVal results for taxa of the phylum *Firmicutes*.

**Table S4.** IndVal results for taxa of the obesity dataset.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.