Phylogenetic Measures of the Core Microbiome

Sharon Anne Bewick[1]*, Benjamin Thomas Camper[1]

[1]Department of Biological Sciences, Clemson University, Clemson, South Carolina 29631, USA

*sbewick@clemson.edu

Abstract

Background: A fundamental concept in microbial ecology is the 'core microbiome.' Typically, core microbiomes are defined as the microbial taxa, genes, or functions shared by a threshold number of microbiome samples from a particular type of habitat (e.g., a particular type of host or a particular type of environment/ecosystem). In defining the core microbiome, the goal is to capture the portion of the microbial community that is conserved across samples from the focal habitat. Recently, there has been growing interest in developing methods to better characterize core microbiomes. As a result, numerous occurrence- and abundance-based measures have been defined. However, none have included phylogenetically aware metrics for analyzing core microbiomes.

Results: In this paper, we develop the concept of the 'core community phylogeny' – a phylogeny where branches are selected based on their presence in multiple samples from a single type of habitat. We then use the core community phylogeny to define phylogenetic metrics for the diversity of core microbiomes from a single type of habitat, for turnover between core microbiomes from two different types of habitat, and for shared diversity across core microbiomes from two or more different types of habitat. As compared to non-phylogenetic metrics, our phylogenetic metrics show greater consistency across taxonomic rank and less

sensitivity to strain variation across microbiome samples. Thus, our metrics address key challenges in the interpretation of core microbiomes.

Conclusions: We provide a phylogenetic framework for characterizing and comparing core microbiomes. Importantly, the methods that we propose allows integration of microbiome properties across taxonomic rank. Ultimately, this will provide both a more consistent picture of the core microbiome, as well as novel biological insight into the conserved components of microbial communities.

Introduction

Core microbiomes have become increasingly important as a dimension for studying and characterizing sets of microbial communities.[1,2] Broadly speaking, a core microbiome is the group of microbial taxa, genes, or functions shared by a large number of microbiome samples from a specific type of habitat. Often, the habitat is defined based on a host species[3–5] (e.g., the core microbiome of the human gut[6]), or a specific type of ecosystem (e.g., the core microbiome of a particular soil type[7]). However, much broader[8] (e.g., the core microbiome of ruminant guts[9]) and much narrower (e.g., the core microbiome of a particular host population[10] or host genotype[11]) habitat classes can be used as well. Because core microbial taxa are commonly, if not always, associated with a specific habitat, they are thought to be important for identifying microbial persistence requirements within the habitat, predicting ecosystem function[12–15] or, in the case of host-associated microbiomes, understanding symbiotic relationships essential to host performance.[16–19] By contrast, it is generally assumed that microbial taxa occurring only sporadically in a given habitat are more likely to be transients or other community members that do not contribute substantially to host or ecosystem function.

While the concept of a core microbiome is both simple and appealing, how to best define the core microbiome of any given habitat remains unclear.[1,2] In general, three strategies have been proposed[1] based on 1) the frequency of occurrence (i.e., incidence) of microbial taxa across samples, 2) the abundance of microbial taxa in individual samples or across a set of samples, or 3) a combination of occurrence and abundance criteria. Within these three strategies, however, there are a wide range of parameters that can influence interpretation. One obvious example is the threshold occurrence rate or abundance that a microbial taxon must achieve to be considered part of the core. In some cases, microbial taxa are included in the core if they are found in at least two samples from a given habitat. In other cases, however, much higher core thresholds have been used, such as requiring that a microbial taxon be present in half or even all of the samples from a particular habitat. Some studies even present analyses across a range of core thresholds so that results can be interpreted in context.[20,21]

Beyond the abundance or occurrence threshold, another decision that can influence identification of core microbial taxa is the taxonomic rank at which the core microbiome is defined. Studies exist spanning a wide range of taxonomic ranks from phyla to amplicon sequence variants (ASVs).[1] Not surprisingly, examining the core microbiome at different taxonomic scales often results in different conclusions. Higher taxonomic ranks like phyla and order tend to be more conserved across samples, resulting in a larger proportion of microbial taxa being identified as part of the core. However, neither ecology nor function are widely conserved at these higher taxonomic ranks, making the core microbiome less useful for informing host or ecosystem ecology. By contrast, at very low taxonomic ranks, it is unclear whether variation between

70 microbial taxa has significant ecological or functional consequences. ASVs, for example, differ
71 by one base pair in the portion of the marker gene used for sequencing.[22] However, in microbial
72 taxa with multiple copies of the 16S rRNA marker gene, individual gene copies can vary
73 intragenomically by >1%.[23,24] Thus, ASVs may be much more diverse than the populations or
74 functions that they are being used to represent.[25] This can impact the core microbiome in one of
75 two ways. First, if multiple closely related ASVs are present in the threshold number of samples,
76 they may all be counted individually as part of the core, artificially inflating core diversity.
77 Second, if individual samples each contain unique but closely related ASVs, none of the ASVs
78 may occur in sufficient samples to be included in the core, artificially deflating core diversity.
79 Perhaps because of the challenges associated with defining core microbiomes at very high and
80 very low taxonomic ranks, most analyses of core microbiomes have used intermediate taxonomic
81 ranks, for example 97% operational taxonomic units (OTUs) or genera.[1]
82
83 The importance of taxonomic scale, and the apparent variation in study outcomes as a function of
84 taxonomic rank are not unique to analysis of core microbiomes.[1] Indeed, these same challenges
85 occur in more general types of ecological analyses of both microbial and macrobial systems, for
86 example, calculation of α- and β-diversity in and between ecological communities. This has led
87 to the development of different phylogenetically aware metrics including Faith's phylogenetic
88 diversity[26] (PD), Rao's quadratic entropy[27] and phylogenetic entropy[28] for α-diversity, and
89 UniFrac distances[29] for β-diversity. Phylogenetically aware metrics are advantageous because
90 they integrate information across taxonomic ranks based on the branch lengths in a phylogenetic
91 tree. This removes the need to select an arbitrary taxonomic rank for any particular ecological
92 analysis. Further, because some phylogenetic properties can serve as indicators of both

93 functional diversity[30] and evolutionary potential[31] (i.e., capacity to evolve in response to a
94 perturbation, for example environmental change[32–34]), phylogenetically aware metrics are
95 thought to provide more informative comparisons of properties like diversity than do analyses at
96 a single taxonomic scale.
97
98 Despite the many benefits of phylogenetically aware ecological analyses, phylogenetically aware
99 metrics for characterizing the core microbiome are currently lacking. In this paper, we develop a
100 framework for analyzing core microbiomes using phylogenetic methods. To do this, we first
101 develop the concept of a 'core community phylogeny' - a phylogeny where branches are defined
102 based on being present in multiple samples from a single type of host or environment. We then
103 use this phylogeny to examine common measures of core communities, including Venn
104 diagrams, as well as metrics of within-habitat and between-habitat diversity.
105
106
107 **Core Community Phylogenies**
108
109 All phylogenetically aware metrics of ecological communities, for example Faith's phylogenetic
110 diversity (PD) and UniFrac distance, are based on an underlying phylogenetic tree comprising
111 taxa present in the community. Logically, the analogous phylogenetic tree for core microbiomes
112 should comprise all microbial taxa that are conserved across microbiomes from a particular type
113 of habitat. However, for core microbiomes, there are two possible ways of defining such a tree: a
114 'tip-based' approach, and a 'branch-based' approach.
115

116 ***Tip-based phylogeny:*** One way to build a core community phylogeny is to first identify core
117 microbial taxa based on their occurrence across a threshold number of microbiome samples from
118 a particular type of habitat. Once these microbial taxa have been identified, a phylogenetic tree
119 can be reconstructed based on the sequences (or other traits) of the conserved taxa. Notably,
120 conserved microbial taxa can be identified at any pre-selected taxonomic rank (e.g., phyla,
121 genera, 97% OTUs, ASVs, etc.) provided that a phylogenetic tree can be built connecting the
122 various microbial lineages.
123
124 ***Branch-based phylogeny:*** An alternative way to build a core community phylogeny is to first
125 build a phylogeny of all microbial taxa present in a set of microbiome samples from a particular
126 type of habitat. This phylogenetic tree can then be examined, branch-by-branch, to determine
127 which branches are present on/in a threshold number of microbiome samples. Branches shared
128 by a threshold number of samples can then be retained, while branches not shared by the
129 threshold number of samples can be discarded. Note that, at least for a rooted tree, this results in
130 discarding shallow branches connected to deeper branches that have also been discarded, such
131 that the remaining phylogeny is fully connected (but see below for issues that arise with unrooted
132 trees). Like the tip-based phylogeny, the branch-based phylogeny can have leaf nodes at any
133 taxonomic rank. However, because the branch-based phylogeny integrates information on shared
134 microbial lineages across all depths of the phylogeny both for defining the core and calculating
135 core metrics, there is less reason to use high taxonomic levels (see 'Examples' below). Thus, we
136 recommend ASVs as the default choice.
137

Importantly, the tip-based phylogeny and the branch-based phylogeny for a given system will not always be identical. Whereas the branch-based phylogeny will always have all of the branches present in the tip-based phylogeny, the tip-based phylogeny will often be missing some of the branches present in the branch-based phylogeny (see Figure 1). More specifically, the tip-based phylogeny will be missing branches whenever a set of microbiome samples share closely related, but not identical, taxa. In this case, none of the shallow branches will be considered core for either the tip-based or the branch-based phylogeny. However, the deeper branches from which all of the related tips are derived may be present on the threshold number of samples and thus may be retained in the branch-based phylogeny. By contrast, these branches will be discarded in the tip-based phylogeny, since definition of core is solely based on shared leaf nodes. This fundamental difference is what makes the branch-based approach less sensitive to variations in the precise taxonomic scale considered, at least when making relative comparisons between two or more core communities (see 'Examples' below).

Figure 1 shows the construction of both a tip-based phylogeny (A) and a branch-based phylogeny (B) from the same set of microbiome samples. In Figure 1, the non-core branches of the microbial phylogeny are shown in black, while the core branches are shown in red. Notice that there are differences between the tip-based phylogeny and the branch-based phylogeny (see starred branches). In particular, the branch-based phylogeny includes additional internal branches that are missing from the tip-based phylogeny. This occurs because some clades are represented in the threshold number of samples, and thus are part of the branch-based core microbiome, even though the individual taxa within the clades are not present in the threshold number of samples.

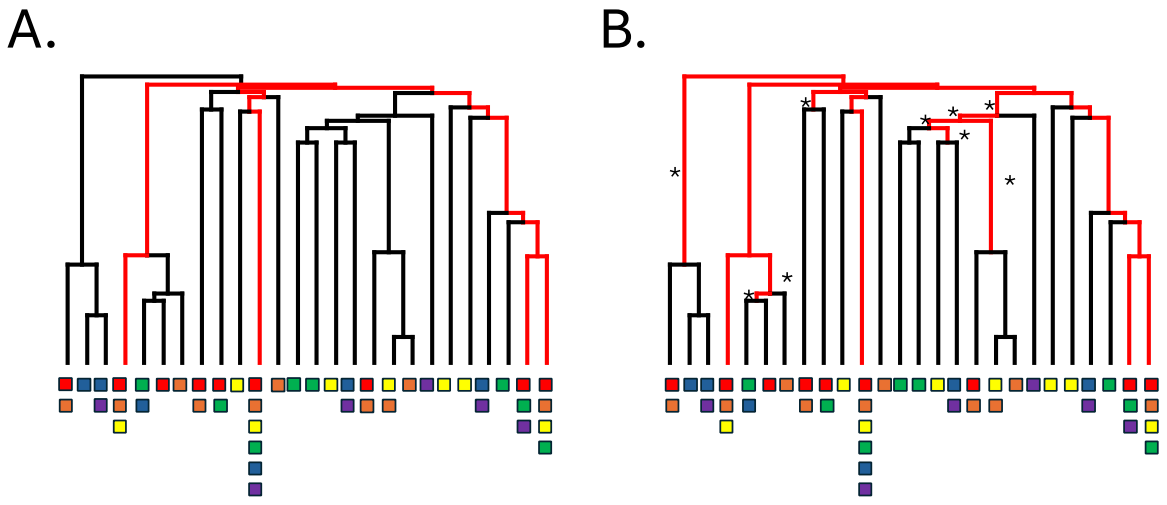


Figure 1. Examples of a tip-based phylogeny (A) and a branch-based phylogeny (B) for the same microbial phylogeny. Branches of the microbial phylogeny that are not part of the core community phylogeny are shown in black, while branches that are part of the core community phylogeny are shown in red. For this example dataset, there are 6 samples (red, orange, yellow, green, blue, and purple squares) and the core threshold is set at 0.5. Thu,s being part of the core requires being present in 3 samples. Colored boxes at the end of each terminal branch indicate the samples that a particular microbial taxon is present on. Stars are used to mark branches that are included in the branch-based phylogeny but not in the tip-based phylogeny.

Another consideration in defining the core community phylogeny is whether to assume that the root node is the most recent common ancestor (MRCA) of all taxa in the sample. The root node should only be used as the MRCA when a rooted tree is provided. Otherwise, the root node is arbitrary, and there is no reason to believe that the root node is any more basal than the other nodes in the tree. Fortunately, microbial phylogenies generated using pipelines like Qiime2[35] usually have the option for rooting trees (although this is typically done through midpoint rooting, which involves multiple assumptions). Similar to UniFrac, we strongly encourage using a rooted tree and forcing the root node to be the MRCA. This is particularly true for branch-based core community phylogenies. The reason is as follows: using the branch-based approach, branches are counted based on their inclusion in individual phylogenies from each sample. When

182 a root node is defined, these individual sample phylogenies necessarily include the root. In
183 absence of a root, however, individual sample phylogenies only span the taxa present in each
184 sample (core or otherwise). Consequently, basal branches may not be counted if entire clades are
185 differentially distributed across subsets of samples. When this happens, distinct clades are not
186 necessarily connected in the phylogeny that emerges. To remedy this, we connect the final core
187 community phylogeny based on the MRCAs for all of the distinct clades. This, however, does
188 not guarantee that branches will not be lost in the process. For tip-based phylogenies, the choice
189 to use an unrooted tree is less problematic because the branches of the core community
190 phylogeny are selected based on the occurrence rate of microbial taxa (tip nodes), not on the
191 occurrence rate of the branches themselves. Even for tip-based trees, however, we recommend
192 using a rooted tree and assuming that the root node is the MRCA.
193
194 **Phylogenetic Metrics for Core Microbiomes**
195 Once a core community phylogeny has been constructed, it can be used to perform phylogenetic
196 analyses of microbiome samples from different types of habitats. Here we consider several
197 common analysis methods, including Venn diagrams, a core microbiome equivalent of Faith's
198 phylogenetic diversity (α-diversity), and a core microbiome equivalent of UniFrac distance (β-
199 diversity).
200
201 ***Venn Diagrams*:** Venn diagrams are one of the most common methods for visualizing core
202 microbiomes across habitats. Non-phylogenetic Venn diagrams show counts of the core
203 microbial taxa shared by different combinations of habitat types. This can be extended to a
204 phylogenetic framework by replacing counts of shared microbial taxa with shared phylogenetic

205 branch lengths. Briefly, the core community phylogenies from different habitat types can be
206 compared, branch-by-branch, to determine which branches appear in which habitat
207 combinations. The total lengths of these shared branch segments can then be calculated and
208 reported either as an absolute value or as percentages of the total branch length spanning the core
209 community phylogenies of all habitats in the analysis.
210
211 ***Core Faith's Phylogenetic Diversity* (PD):** The phylogenetic equivalent to richness (a count of
212 microbial taxa present in a community), Faith's PD is the sum of the branch lengths of all
213 branches spanning the taxa present in a particular microbiome sample or ecological community.
214 This definition can easily be extended to core microbiomes by summing the branch lengths of all
215 branches in the core community phylogeny. Doing so gives a measure of how much
216 phylogenetic diversity is conserved across a threshold number of microbiome samples from a
217 particular type of habitat.
218
219 ***Core UniFrac Distances*:** The phylogenetic equivalent to Jaccard's index[36] (the fraction of
220 unique microbial taxa between two communities), the UniFrac distance[29,37] between two
221 communities is the sum of the branch lengths that are unique to each microbiome sample or
222 ecological community divided by the sum of the branch lengths of all branches spanning taxa
223 present in both microbiome samples or ecological communities. Again, this definition can be
224 extended to core microbiomes by dividing the branch lengths unique to each core community
225 phylogeny by the total branch lengths of the two core community phylogenies combined. Doing
226 so gives a measure of how phylogenetically distinct the core microbiomes are from two different
227 types of habitats.

228 **Examples**

229 Additional file 1 outlines the calculation of Venn diagrams, core Faith's PD, and core UniFrac
230 distances using a set of highly simplified core community phylogenies. Specifically, we show
231 differences between tip- and branch-based approaches, and outline calculations with both rooted-
232 and unrooted trees. In the remainder of the main text, we apply our methods to more realistic
233 systems using rooted trees. More specifically, we consider two datasets - one comprising the skin
234 microbiota from four lizard populations and the other comprising the gut microbiota from the
235 same four lizard populations (see Materials and Methods). In doing so, we show how our
236 phylogenetic metrics provide additional and/or alternate information compared to existing
237 metrics. We further demonstrate how our approach resolves some of the outstanding issues that
238 have been raised regarding choice of taxonomic rank for defining the core microbiome.

239

240 Figure 2 shows standard Venn diagrams based on ASV counts (A, D) along with our new
241 phylogenetic Venn diagrams for tip-based (B, E) and branch-based (C, F) trees. As compared to
242 our phylogenetic Venn diagrams, non-phylogenetic Venn diagrams (A, D) show the lowest
243 percentage of system diversity in the compartment representing intersection of all four lizard
244 populations (21.4% for skin and 0% for gut) and the highest percentage of system diversity in the
245 four compartments unique to single lizard populations (42.5% for skin and 65% for gut). This is
246 not surprising. Using a non-phylogenetic approach, microbial taxa from different host
247 populations must be identical to be counted in intersection compartments. This biases against
248 finding shared microbial diversity, particularly when using ASVs, because microbial lineages
249 with small differences in 16S rRNA gene sequences (e.g., one base pair substitutions) are

deemed unique. By contrast, with a phylogenetic approach, even if different host populations

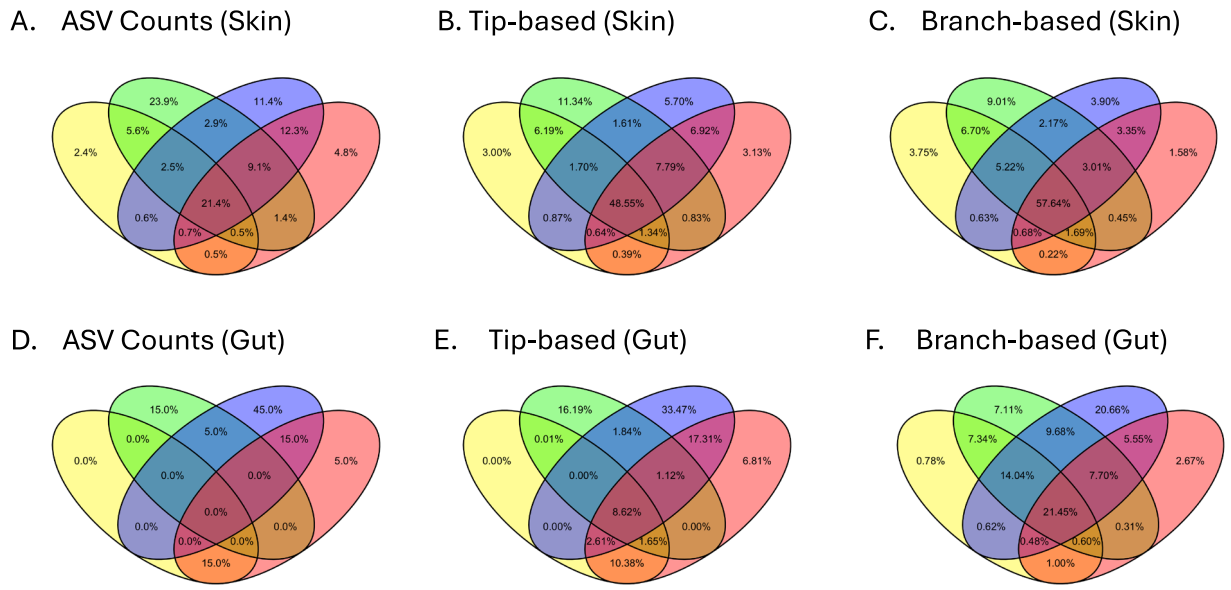


Figure 2. Venn diagrams for lizard skin (A-C) and gut (D-F) microbiota showing the percentage of the total ASV counts shared by the core microbiota from each lizard population (A, D) and the phylogenetic branch length shared by the core microbiota from each lizard population using either tip-based (B, E) or branch-based (C, F) phylogenies. In all Venn diagrams, lizard populations are as follows: *A. arizonae* (yellow), *A. neomexicanus* from SBluG (green), *A. neomexicanus* from SNW (blue) and *A. marmoratus* (red).

have slightly different ASVs, this only contributes a small amount of diversity to the unique compartments of the Venn diagram. Meanwhile, the majority of the diversity is allotted to intersection compartments because the ASVs from the different host populations share most of their phylogenetic history. Comparing branch-based and tip-based phylogenetic approaches, branch-based phylogenies show a higher percentage of system diversity in the compartment representing intersection across all four populations (57.64% for skin and 21.45% for gut) and a lower percentage of system diversity in the four compartments unique to single lizard populations (18.24% for skin and 31.22% for gut). This reflects the inclusion of additional, deep internal branches in branch-based phylogenies that are dropped from tip-based phylogenies (see

269 Figure 1). These basal branches are more likely to be conserved across different host populations
270 (i.e., phyla are more likely to be shared than ASVs), inflating diversity in shared compartments
271 of the Venn diagram.
272
273 Figure 3 (A, C) shows the tip-based and branch-based core Faith's PD, along with core richness
274 (count of microbial taxa) as a function of taxonomic rank for lizard skin and gut microbiota. For
275 the sake of comparison across metrics and taxonomic rank, we have plotted the core Faith's PD
276 of three lizard populations (*A. marmoratus* at site SNW as red, *A. neomexicanus* at SNW as blue,
277 and *A. neomexicanus* at SBluG as green) relative to the fourth lizard population (*A. arizonae* at
278 SBluG; see Materials and Methods). Figure 3 (B, D) shows the range of values obtained for
279 relative branch-based and tip-based core Faith's PD and core richness across all taxonomic ranks
280 considered. Importantly, for both skin and gut microbiota, phylogenetic measures, and the
281 branch-based core Faith's PD in particular, show much less variation across taxonomic ranks.
282 For example, whereas richness indicates that the core gut microbiota of the *A. neomexicanus*
283 population at SNW (blue) is anywhere from 2.28 (microbial order) to 9.25 (microbial species)
284 times as diverse as the core gut microbiota of the *A. arizonae* population at SBluG, branch-based
285 core Faith's PD suggests a much narrower range from 1.73 (microbial species) to 1.80 (microbial
286 family). Not all differences in metric stability across taxonomic rank are as dramatic. Still, the
287 branch-based core Faith's PD typically shows the least variation across taxonomic ranks, the tip-
288 based core Faith's PD the second least variation, and richness the most variation.

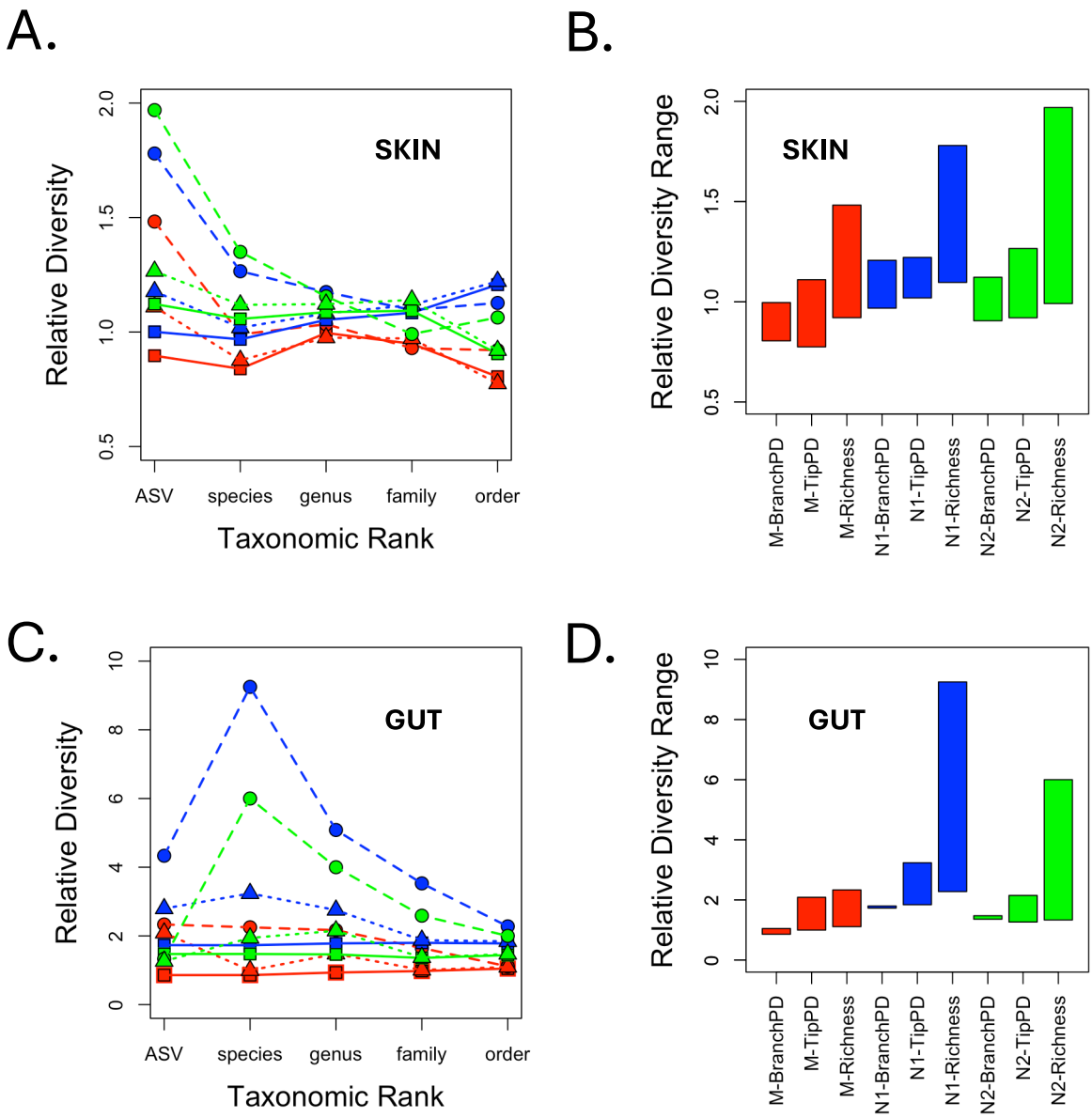


Figure 3. Comparison of alpha diversity across taxonomic ranks (A, C) for the branch-based Core Faith's PD (solid, squares), the tip-based Core Faith's PD (dotted, triangles), and richness (dashed, circles). In both panels, we show the diversity of the focal population relative to the diversity of the *A. arizonae* population. Focal populations are *A. marmoratus* (red), *A. neomexicanus* at SNW (blue) and *A. neomexicanus* at SBluG (green). Bar graphs (B, D) showing minimum and maximum values of alpha diversity across taxonomic rank for *A. marmoratus* (red), *A. neomexicanus* at SNW (blue), and *A. neomexicanus* at SBluG (green) relative to *A. arizonae*. Diversity metrics are indicated as follows: 'BranchPD' is the branch-based phylogeny approach, 'TipPD' is the tip-based phylogeny approach, and 'Richness' is a count of taxa for each taxonomic rank.

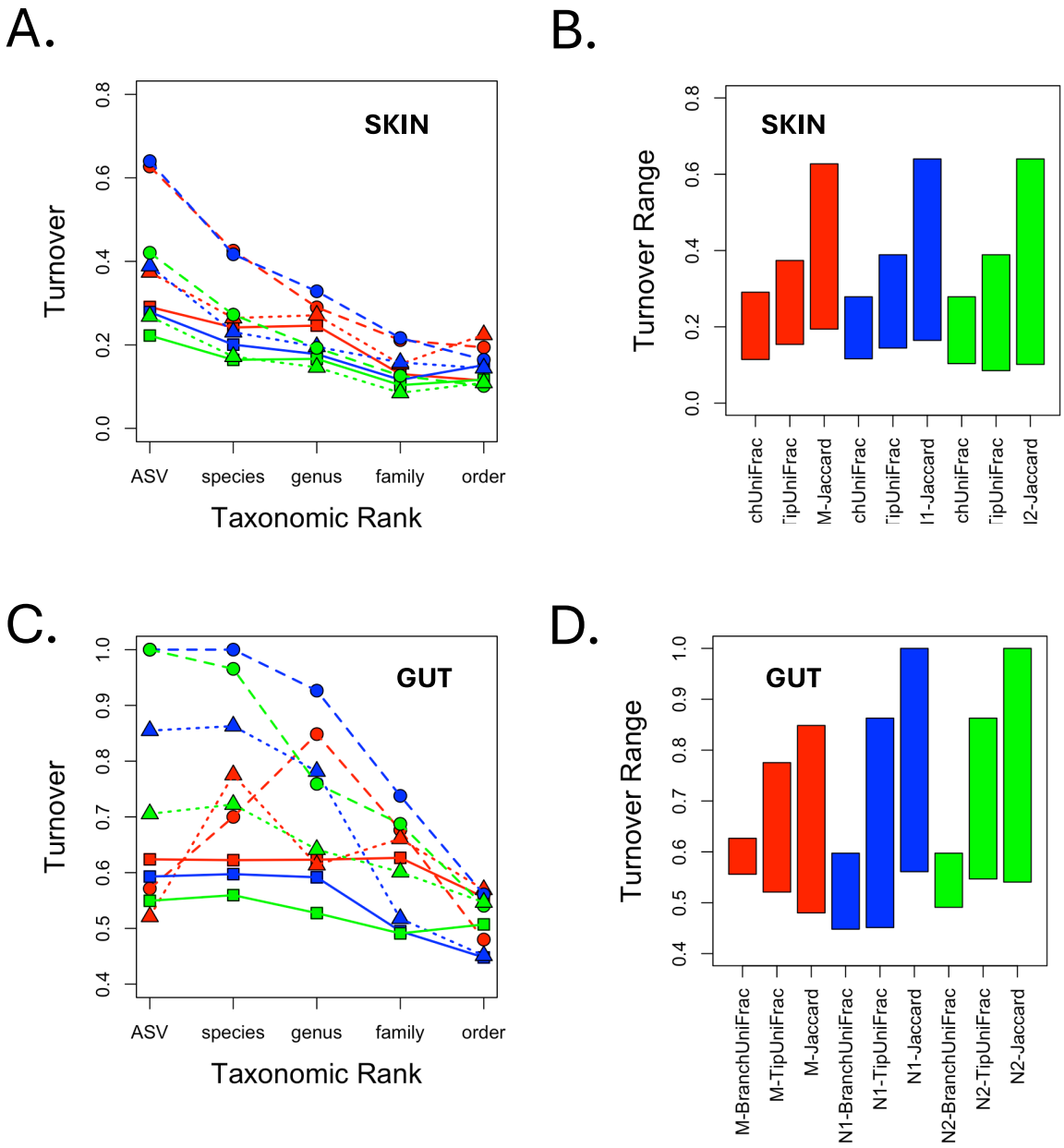


Figure 4. Comparison of beta diversity across taxonomic ranks (A, C) for the branch-based Core UniFrac distance (solid, squares), the tip-based Core UniFrac distance (dotted, triangles), and Jaccard distance (dashed, circles). In both panels, we show turnover of the focal population with respect to the *A. arizonae* population. Focal populations are *A. marmoratus* (red), *A. neomexicanus* at SNW (blue), and *A. neomexicanus* at SBluG (green). Bar graphs (B, D) showing minimum and maximum values of beta diversity (turnover or distance) across taxonomic rank between *A. arizonae* and *A. marmoratus* (red), *A. neomexicanus* at SNW (blue), and *A. neomexicanus* at SBluG (green) with respect to *A. arizonae*. Diversity metrics are as follows: where 'BranchUniFrac' is the branch-based phylogeny approach, 'TipUniFrac' is the tip-based phylogeny approach, and 'Jaccard' is a count of taxa for each taxonomic rank.

312 Figure 4 (A, C) shows the tip-based and branch-based UniFrac distances, along with the Jaccard
313 distance (turnover) between the core skin and gut microbiota for *A. arizonae* at SBluG and the
314 core skin and gut microbiota for all three other whiptail populations (see 'Materials and
315 Methods') as a function of the taxonomic rank. Figure 4 (B, D) shows the range of values
316 obtained for tip-based and branch-based UniFrac distances as well as Jaccard distances across all
317 taxonomic ranks considered. Broadly speaking, the non-phylogenetic approach (i.e., Jaccard
318 dissimilarity) gives the largest dissimilarities/distances, while the branch-based phylogenetic
319 approach gives the smallest dissimilarities. With respect to variation in dissimilarity across
320 taxonomic rank, we again see that the branch-based phylogenetic distance measure gives the
321 least variation, while the tip-based distance gives intermediate variation and the Jaccard distance
322 gives the most variation.
323

324 **The holobiont R package**

325 To facilitate the use of our new phylogenetic measures of core microbiomes, we developed a
326 new R package, holobiont, that includes functions for identifying the core community
327 phylogeny in any microbiome, drawing phylogenetic Venn diagrams, calculating the core Faith's
328 PD for a set of communities, and calculating the core UniFrac distance between two sets of
329 communities. Briefly, the coretree function takes a phyloseq object[38] and a value for the core
330 threshold and identifies the core community phylogeny, coloring the branches of the core
331 community phylogeny red, while leaving the remaining branches of the community tree black.
332 The corePhyloVenn function takes a phyloseq object, a vector defining the habitat class for
333 each microbiome sample, and a core threshold and draws a phylogenetic Venn diagram of either
334 branch lengths or percentages of branch lengths shared by the core community phylogenies of

335 different combinations of habitat classes. The coreVennTree function takes the same input as
336 the corePhyloVenn function and colors the branches of the phylogenetic tree based on the
337 combination of habitat class core community phylogenies that they belong to. The coreFaithsPD
338 function takes a phyloseq object and a core threshold and calculates the core Faith's PD. Finally,
339 the coreUniFrac function takes a phyloseq object and a vector specifying which samples belong
340 to which habitat class and calculates the UniFrac distance between the two core community
341 phylogenies. Additional details outlining the methods and use of our R package can be found at
342 https://bewicklab.weebly.com/databases--packages.html.
343

344 **Discussion**

345 In this paper, we propose a framework for performing phylogenetically aware analyses of core
346 microbiomes. More specifically, we develop the concept of a core community phylogeny - a
347 phylogeny based on the microbial lineages that are core to a specific habitat class. We then
348 propose two different methods for building this tree - a tip-based approach and a branch-based
349 approach. For the tip-based approach, core membership is decided based on occurrence of the
350 leaf (tip) nodes. This is akin to the standard, non-phylogenetic method for defining a core
351 microbiome. However, once members of the core microbiome have been identified, a
352 phylogenetic tree is constructed. This allows for calculation of phylogenetically aware metrics
353 that leverage the relatedness of the leaf nodes to determine overall properties (e.g., diversity) of
354 the core microbiome. For the branch-based approach, core membership is decided independently
355 for each branch of the phylogenetic tree. Therefore, in the branch-based approach, phylogeny
356 informs not only the calculation of metrics but also, the identification of the core microbiome
357 itself. The downside of the branch-based approach, however, is that the 'core microbiome' often

358 includes some internal or ancestral nodes without their associated leaf nodes. Thus, metrics
359 calculated using the branch-based approach are not necessarily based on a set of extant taxa.
360 Rather, they are based on a set of complete and partial microbial lineages. Still, because it uses
361 phylogenetic information to both define the core microbiome and to calculate core metrics, we
362 suggest using the branch-based approach except when there is a strong desire to identify specific
363 microbial taxa within the core (e.g., to select for culturing experiments).
364
365 As compared to existing metrics, the primary advantage of our phylogenetic metrics is that they
366 integrate core microbiome properties (i.e., richness, turnover, shared taxa) across taxonomic
367 scales. Thus, our phylogenetic metrics are more likely to provide biologically relevant
368 information. This is especially true when non-phylogenetic metrics are calculated at taxonomic
369 ranks that fail to capture meaningful biological variation.[39] If, for example, the chosen taxonomic
370 rank is too low, substantial biological variation may be identified that has no functional or
371 ecological consequences. Alternatively, if the chosen taxonomic rank is too high, large amounts
372 of meaningful biological variation may be pooled into the same taxonomic unit. Depending on
373 the system, the type of variation considered and the metric calculated, it can be difficult to *a*
374 *priori* determine (or even *a posteriori* assess) which taxonomic ranks are more or less likely to
375 be biologically informative. Further, in certain instances multiple taxonomic ranks exhibit
376 substantial meaningful biological variation. In this case, a non-phylogenetic metric at any single
377 taxonomic rank - even a well-chosen taxonomic rank - cannot fully capture the underlying
378 signal. Phylogenetic metrics circumvent these challenges by summarizing the full spectrum of
379 biological variation across all taxonomic ranks, with variation at lower taxonomic ranks
380 weighted less than variation at higher taxonomic ranks.

381

382 Another benefit of our branch-based approach (but not our tip-based approach) is that it
383 overcomes challenges associated with defining the core microbiome in the face of high levels of
384 strain variation. This is common in certain microbiome systems, including many host-associated
385 microbiomes. The pervasive human skin isolate, *Staphylococcus epidermidis*, for example,
386 evolves on individual humans over time,[40,41] resulting in strain level variation across the human
387 population. Another common human skin isolate, *Cutibacterium acnes*, exhibits even more
388 dramatic strain-level variation, showing differences in sequences across pores from a single
389 human individual.[42] Indeed, in-depth profiling of the human microbiome has shown that there
390 are numerous personalized microbial taxa that are found on only one or several individuals.[41,43]
391 Depending on the taxonomic rank at which the core microbiome is defined and the degree to
392 which any single microbial strain is shared among individual hosts, natural strain-level variation
393 can cause broadly shared microbial clades to be obscured from the non-phylogenetic core
394 microbiome. By contrast, if we apply our branch-based approach to these same scenarios, only
395 the terminal branches of the tree are omitted. Meanwhile, the more basal branches of the clades
396 are retained because they are shared across all strains.

397

398 Because our phylogenetic metrics summarize microbiome measures across taxonomic rank and
399 because the definition of the core microbiome - at least when using the branch-based approach -
400 is less sensitive to strain-level variation, our phylogenetic metrics tend to be more consistent than
401 comparable non-phylogenetic metrics when calculated at different taxonomic ranks. This is true
402 both in a relative (i.e., comparison between microbiome samples) and an absolute sense. To see
403 why phylogenetic metrics are more consistent, consider a habitat with hundreds of closely related

404 core microbial strains within a single phylum. Using a non-phylogenetic approach this system
405 will have a very high α-diversity at the ASV-level, but a very low α-diversity at the phylum-
406 level. By contrast, using our phylogenetic approach, α-diversity will be low regardless of
407 whether the leaf nodes are defined at the ASV- or phylum-level. This is because, even when the
408 leaf nodes are ASVs all of the different ASV strains contribute very short branch lengths.
409 Consequently, they do not substantially inflate α-diversity beyond the diversity of the single
410 phylum. The greater consistency of our phylogenetic metrics across taxonomic ranks means that
411 it is less important to choose the correct taxonomic rank, and that it is rarely disadvantageous to
412 use ASVs. This addresses one of the challenges that has been raised about interpreting core
413 microbiomes when outcomes are variable across taxonomic rank.[1]
414
415 The primary goal of this paper is to introduce a framework for phylogenetic analysis of core
416 microbiomes. For this reason, we have focused on the concepts required to define a core
417 community phylogeny and have outlined the application of a core community phylogeny to three
418 types of microbiome analyses - Venn diagrams, α-diversity (Faith's PD), and β-diversity
419 (UniFrac). Our broad approach, however, is not limited to these particular types of analyses.
420 Indeed, additional analyses could be developed within the same framework. A straightforward
421 example would be to use UniFrac distances between core microbiomes to project into ordination
422 space (e.g., Principal Coordinate Analysis or Nonmetric Multidimensional Scaling). This would
423 provide phylogenetically informed visualization of the relatedness of different core microbiomes.
424 A more challenging extension would be to develop a method for applying our phylogenetic
425 framework to core microbiomes based on abundance thresholds[44] or even to develop metrics

426 where phylogenetic branches are weighted based on average abundances or occurrence rates in
427 the samples across which the core microbiome is defined.
428
429 While initially developed for microbiome analysis, beyond microbiomes, our community
430 phylogenetics approach and associated R package may ultimately have wide application to
431 macrobial systems as well. Currently, very few macrobial studies attempt to identify core
432 communities (but see Murdock et al.[45]). This is likely because, at least historically, macrobial
433 systems have not been studied with many replicates. Indeed, even when replicate macrobial
434 communities are characterized, the effort required to identify all taxa in a single community
435 means that replicate numbers are low. This makes it difficult to identify a core community.
436 Further, in species poor macrobial systems where characterization of replicate communities is
437 possible, the core community concept is not as necessary or useful because the systems are
438 simple enough to analyze in their entirety. As sequencing techniques improve, however, and
439 more and more macrobial systems are characterized in a manner similar to microbiomes (e.g.,
440 through the use of eDNA[46,47]), it seems likely that core communities will emerge as a common
441 form of analysis for species rich macrobial systems (e.g. tropical rainforests, litter insects) as
442 well. There is no reason why our proposed framework cannot be applied to macrobial
443 communities. In fact, our framework can be applied to any system including prokaryotic and
444 eukaryotic communities, host-associated and free-living communities, and microbial (virus,
445 bacteria, fungi, etc.) and macrobial communities. It can also be applied to systems characterized
446 through a wide variety of methods, even including non-sequencing techniques. The only
447 requirements for our approach are tables of presences/absences of different taxa across different
448 communities, along with a phylogenetic tree comprising all taxa present across all communities.

449 Thus, occurrence data can be based on sequencing, but can also be based on any other form of
450 observation, including transect or plot surveys. Likewise, phylogenies can be constructed using
451 sequence data, including sequencing of any kind (single marker genes, combinations of marker
452 genes, whole genomes), but can also be generated based on morphology or other traits
453 commonly used to build trees (though the validity of any core community analysis will depend
454 on the accuracy of both the occurrence table and the phylogeny used).
455
456 The identification and characterization of core microbiomes has become a hallmark of
457 microbiome analysis. However, to date, most analyses of core microbiomes have focused on
458 taxonomy at specified taxonomic ranks. By contrast, phylogenetically aware metrics for
459 characterizing core microbiomes are largely lacking. Phylogenetically aware metrics have the
460 benefit of integrating information across taxonomic ranks. This can be particularly helpful when
461 analysis at different taxonomic ranks yields different outcomes - a common problem in ecology
462 and evolution[48–52] including the study of core microbiomes.[1] Our phylogenetic framework
463 overcomes this challenge. Indeed, even when our analysis is performed at different taxonomic
464 ranks, it typically yields more consistent results than non-phylogenetic methods. With the
465 development of our new framework and associated R package, we provide a mechanism to bring
466 phylogenetically aware community analyses to the study of core microbiomes. Ultimately, this
467 will allow for novel insight into the conserved components of microbial communities both within
468 and between different types of habitats.
469
470 **Methods**
471 ***Data collection***

472 We sampled the microbiota from several whiptail lizard species (genus *Aspidoscelis*, family
473 *Teiidae*) at two locations (SBluG and SNW; see Camper et al.[53] for more site details) within
474 Sevilleta National Wildlife Refuge. Each location represents a site of syntopy between hybrid *A.*
475 *neomexicanus* and a progenitor species, *A. marmoratus* (♀) and *A. arizonae* (♂, recently split
476 from *A. inornatus*[54]). Following taxonomic guidelines by Tucker et al.[55] and Walker et al.,[56] we
477 refer to *Aspidoscelis* spp. using the masculine epithet. We captured lizards using a rapidly
478 deployable drift fence trapping array lined with box funnel traps, further outlined in Camper et
479 al.,[53] and by hand or lizard noose when possible. Upon capture, we sampled the gut (=cloacal)
480 and skin microbiota of each species following a swabbing protocol in Camper et al.[20] We
481 sampled each location until we captured at least 15 female lizards of each whiptail species. All
482 samples were sent to ZymoBIOMICS Targeted Sequencing Service for Microbiota Analysis
483 (Zymo Research) using 16S rRNA gene sequencing. Additional details are available from
484 https://www.zymoresearch.com/pages/16s-its-amplicon-sequencing and Camper et al.[20] FASTQ
485 files for lizard microbiomes, as well as environmental and sequencing controls, are available at
486 (These will be uploaded to NCBI upon publication).
487
488 ***Data analysis***
489 Using the .biom files supplied by Zymo, we first removed 16S rRNA gene sequences that could
490 not be identified as Bacteria or Archaea based on the proprietary Zymo database. Additionally,
491 we removed all samples with a read depth lower than 10,000 reads. Finally, we rarefied all
492 samples to the read depth of the lowest sample from each particular body site using the
493 `rarefy_even_depth` function from the `phyloseq` package.[38] All analyses were performed using
494 the R programming language[57] (version 4.2.1) available on our Github page.

For Venn diagrams, we used the corePhyloVenn function in our holobiont package, selecting habitat classes based on species x site. We generated Venn diagrams separately for skin and gut microbiota. In all cases, we reported percentages of the total branch length and used rooted phylogenies. To compare standard Venn diagrams to our phylogenetic Venn diagrams, we counted the number of core taxa shared by each species x site combination and used these counts (richness) to generate a standard core microbiome Venn diagram. All Venn diagrams were calculated based on ASVs using a core threshold of 0.5 (i.e., a microbial taxon or branch must be present in 50% of the samples from a particular lizard species x site habitat class).

For Faith's PD, we used the coreFaithsPD function in our holobiont package. To compare across taxonomic ranks, we took the original ASV .biom file supplied by Zymo to generate an ASV phyloseq object.[38] We then used the tax_glom function from the phyloseq package to generate OTU tables and phylogenies at taxonomic ranks from species ('Rank7') through to order ('Rank4'). This allowed us to calculate Faith's PD at all ranks, as well as the related non-phylogenetic richness. For the latter, we summed up the number of microbial taxa at any given taxonomic rank that were present in the core microbiome of the focal habitat. All Faith's PD and richness calculations were performed on rooted trees assuming a core threshold of 0.5.

For Unifrac distances, we used the coreUniFrac function in our holobiont package. Comparisons across taxonomic ranks were performed as described for Faith's PD. To compare to the analogous non-phylogenetic metric, we calculated the Jaccard distance based on taxonomic co-occurrence patterns between pairs of core microbiomes from different habitats. This functionality is included as the coreJaccard function in our holobiont package. Again, all

UniFrac and Jaccard calculations were performed on rooted trees assuming a core threshold of 0.5.

All data and code necessary to recalculate our results are available through GitHub at: https://github.com/bewicklab/CoreTrees

Declarations

Ethics approval and consent to participate

All research was approved by Clemson University under IACUC protocol numbers #2020-015 and #2021-047. We completed this work under the Sevilleta National Wildlife Refuge Special Use Permit #SEV_Bewick_Camper_2022_59 and the New Mexico Department of Game and Fish permit authorization #3772.

Consent for publication

not applicable

Funding

This study was funded by NSF awards #2105604 and #2025541, a Clemson University Support for Early Exploration and Development (CUSEED) Grant, and the Clemson University Creative Inquiry (CI) Program.

541 **Availability of data and materials**

542 All sequence data will be deposited in the National Center for Biotechnology Information
543 (NCBI) database upon acceptance. BIOM and metadata files are included in Additional file 2.
544 Additionally, BIOM files, metadata files, and all code necessary for the analyses presented in
545 this manuscript are available at https://github.com/bewicklab/CoreTrees. The holobiont R
546 package is available at https://bewicklab.weebly.com/databases--packages.html.

547

548 **Competing interests**

549 The authors declare that they have no competing interests.

550

551 **Authors' contributions**

552 BTC and SAB conceived the idea and wrote the first draft of the manuscript. BTC performed all
553 related fieldwork. SAB developed the code and data analysis. All authors contributed
554 substantially to the content of the manuscript.

555

556 **Acknowledgements**

557 We thank Thomas Dempster, August Spencer, Eva Purcell, Riley Manuel, Drew Kanes and
558 Zachary Laughlin for their field assistance in New Mexico as well as Lily Margeson, Simon
559 Dunn, Georgianna Bellinger, Henry Egloff, Kaila Hodges, Camryn Lachica, and Savannah Utz
560 for their assistance assembling drift fence trapping arrays.

561

562

563

References

1. Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy Sciences*. 2021;118(51):1-10. doi:10.1073/pnas.2104429118/-/DCSupplemental
2. Custer GF, Gans M, van Diepen LTA, Dini-Andreote F, Buerkle CA. Comparative Analysis of Core Microbiome Assignments: Implications for Ecological Synthesis. *mSystems*. 2023;8(1):1-14. doi:10.1128/msystems.01066-22
3. Huse SM, Ye Y, Zhou Y, Fodor AA. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS One*. 2012;7(6):1-12. doi:10.1371/journal.pone.0034242
4. Zaura E, Keijser BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of oral microbial communities. *BMC Microbiol*. 2009;9(259):1-12. doi:10.1186/1471-2180-9-259
5. Muletz Wolz CR, Yarwood SA, Campbell Grant EH, Fleischer RC, Lips KR. Effects of host species and environment on the skin microbiome of Plethodontid salamanders. *Journal of Animal Ecology*. 2018;87(2):341-353. doi:10.1111/1365-2656.12726
6. Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. *Journal of Physiology*. 2009;587(17):4153-4158. doi:10.1113/jphysiol.2009.174136
7. Pershina E V., Ivanova EA, Korvigo IO, et al. Investigation of the core microbiome in main soil types from the East European plain. *Science of the Total Environment*. 2018;631-632:1421-1430. doi:10.1016/j.scitotenv.2018.03.136
8. Salzman S, Whitaker M, Pierce NE. Cycad-feeding insects share a core gut microbiome. *Biological Journal of the Linnean Society*. 2018;123:728-738. https://academic.oup.com/biolinnean/article/123/4/728/4920834
9. Henderson G, Cox F, Ganesh S, et al. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep*. 2015;5:1-15. doi:10.1038/srep14567
10. Hamady M, Knight R. Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Res*. 2009;19(7):1141-1152. doi:10.1101/gr.085464.108
11. Hernandez-Teran A, Navarro-Diaz M, Benitez M, Lira R, Wegier A, Escalante AE. Host genotype explains rhizospheric microbial community composition: The case of wild cotton metapopulations (Gossypium hirsutum L.) in Mexico. *FEMS Microbiol Ecol*. 2020;96(8):1-11. doi:10.1093/femsec/fiaa109
12. Fan K, Chu H, Eldridge DJ, et al. Soil biodiversity supports the delivery of multiple ecosystem functions in urban greenspaces. *Nat Ecol Evol*. 2023;7:113-126.
13. Sun L, Li G, Zhao J, Zhang T, Liu J, Zhang J. Core microbiota drive multi-functionality of the soil microbiome in the Cinnamomum camphora coppice planting. *BMC Microbiol*. 2024;24(1). doi:10.1186/s12866-023-03170-8

602 14. Björk JR, O'hara RB, Ribes M, Coma R, Montoya JM. The dynamic core microbiome:
603 structure, stability and resistance. *bioRxiv*. Published online 2017:1-50.
604 doi:10.1101/137885
605 15. De Filippis F, La Storia A, Stellato G, Gatti M, Ercolini D. A selected core microbiome
606 drives the early stages of three popular Italian cheese manufactures. *PLoS One*.
607 2014;9(2):1-8. doi:10.1371/journal.pone.0089680
608 16. Risely A. Applying the core microbiome to understand host–microbe systems. *Journal of*
609 *Animal Ecology*. 2020;89(7):1549-1558. doi:10.1111/1365-2656.13229
610 17. Zaura E, Keijser BJ, Huse SM, Crielaard W. Defining the healthy "core microbiome" of
611 oral microbial communities. *BMC Microbiol*. 2009;9(259):1-12. doi:10.1186/1471-2180-
612 9-259
613 18. Ganz HH, Jospin G, Rojas CA, et al. The Kitty Microbiome Project: Defining the Healthy
614 Fecal "Core Microbiome" in Pet Domestic Cats. *Vet Sci*. 2022;9(635):1-21.
615 doi:10.3390/vetsci9110635
616 19. Holyoak GR, Premathilake HU, Lyman CC, et al. The healthy equine uterus harbors a
617 distinct core microbiome plus a rich and diverse microbiome that varies with geographical
618 location. *Sci Rep*. 2022;12(14790):1-14. doi:10.1038/s41598-022-18971-6
619 20. Camper BT, Kanes AS, Laughlin ZT, Manuel RT, Bewick SA. Transgressive Hybrids as
620 Hopeful Holobionts. *bioRxiv*. Published online 2024:1-68.
621 doi:10.1101/2024.08.17.607925
622 21. Astudillo-García C, Bell JJ, Webster NS, et al. Evaluating the core microbiota in complex
623 communities: A systematic investigation. *Environ Microbiol*. 2017;19(4):1450-1462.
624 doi:10.1111/1462-2920.13647
625 22. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace
626 operational taxonomic units in marker-gene data analysis. *ISME Journal*.
627 2017;11(12):2639-2643. doi:10.1038/ismej.2017.119
628 23. Větrovský T, Baldrian P. The Variability of the 16S rRNA Gene in Bacterial Genomes
629 and Its Consequences for Bacterial Community Analyses. *PLoS One*. 2013;8(2):1-10.
630 doi:10.1371/journal.pone.0057923
631 24. Coenye T, Vandamme P. Intragenomic heterogeneity between multiple 16S ribosomal
632 RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett*. 2003;228(1):45-49.
633 doi:10.1016/S0378-1097(03)00717-1
634 25. Schloss PD. Amplicon Sequence Variants Artificially Split Bacterial Genomes into
635 Separate Clusters. *mSphere*. 2021;6(4):1-6. doi:10.1128/msphere.00191-21
636 26. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv*. 1992;61:1-
637 10.
638 27. Rao CR. Diversity and Dissimilarity Coefficients: A Unified Approach. *Theor Popul Biol*.
639 1982;21:24-43.

640 28. Allen B, Kon M, Bar-Yam Y. A new phylogenetic diversity measure generalizing the
641 shannon index and its application to phyllostomid bats. *American Naturalist*.
642 2009;174(2):236-243. doi:10.1086/600101
643 29. Lozupone C, Knight R. UniFrac: A new phylogenetic method for comparing microbial
644 communities. *Appl Environ Microbiol*. 2005;71(12):8228-8235.
645 doi:10.1128/AEM.71.12.8228-8235.2005
646 30. Tucker CM, Davies TJ, Cadotte MW, Pearse WD. On the relationship between
647 phylogenetic diversity and trait diversity. *Ecology*. 2018;99(6):1473-1479.
648 doi:10.1002/ecy.2349
649 31. Winter M, Devictor V, Schweiger O. Phylogenetic diversity and nature conservation:
650 Where are we? *Trends Ecol Evol*. 2013;28(4):199-204. doi:10.1016/j.tree.2012.10.015
651 32. Mouquet N, Devictor V, Meynard CN, et al. Ecophylogenetics: Advances and
652 perspectives. *Biological Reviews*. 2012;87(4):769-785. doi:10.1111/j.1469-
653 185X.2012.00224.x
654 33. Lankau R, Jørgensen PS, Harris DJ, Sih A. Incorporating evolutionary principles into
655 environmental management and policy. *Evol Appl*. 2011;4(2):315-325.
656 doi:10.1111/j.1752-4571.2010.00171.x
657 34. Lavergne S, Mouquet N, Thuiller W, Ronce O. Biodiversity and climate change:
658 Integrating evolutionary and ecological responses of species and communities. *Annu Rev*
659 *Ecol Evol Syst*. 2010;41:321-350. doi:10.1146/annurev-ecolsys-102209-144628
660 35. Estaki M, Jiang L, Bokulich NA, et al. QIIME 2 Enables Comprehensive End-to-End
661 Analysis of Diverse Microbiome Data and Comparative Studies with Publicly Available
662 Data. *Curr Protoc Bioinformatics*. 2020;70(1):1-46. doi:10.1002/cpbi.100
663 36. Jaccard P. Distribution de la flore alpine dans le bassin des Dranses et dans quelques
664 régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*. 1901;37:241-
665 272.
666 37. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: An effective
667 distance metric for microbial community comparison. *ISME Journal*. 2011;5(2):169-172.
668 doi:10.1038/ismej.2010.133
669 38. McMurdie PJ, Holmes S. phyloseq: An R package for reproducible interactive analysis
670 and graphics of microbiome census data. *PLoS One*. 2013;8(4).
671 39. Dubois A. Phylogeny, taxonomy and nomenclature: The problem of taxonomic categories
672 and of nomenclatural ranks. *Zootaxa*. 2007;(1519):27-68. doi:10.11646/zootaxa.1519.1.3
673 40. Zhou W, Spoto M, Hardy R, et al. Host-Specific Evolutionary and Transmission
674 Dynamics Shape the Functional Diversification of Staphylococcus epidermidis in Human
675 Skin. *Cell*. 2020;180(3):454-470. doi:10.1016/j.cell.2020.01.006
676 41. Both A, Huang J, Qi M, et al. Distinct clonal lineages and within-host diversification
677 shape invasive staphylococcus epidermidis populations. *PLoS Pathog*. 2021;17(2):1-39.
678 doi:10.1371/JOURNAL.PPAT.1009304

42. Conwill A, Kuan AC, Damerla R, et al. Anatomy promotes neutral coexistence of strains in the human skin microbiome. *Cell Host Microbe*. 2022;30(2):171-182. doi:10.1016/j.chom.2021.12.007
43. Lloyd-Price J, Mahurkar A, Rahnavard G, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550(7674):61-66. doi:10.1038/nature23889
44. Shade A, Stopnisek N. Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr Opin Microbiol*. 2019;49:50-58. doi:10.1016/j.mib.2019.09.008
45. Murdock SA, Tunnicliffe V, Boschen-Rose RE, Juniper SK. Emergent "core communities" of microbes, meiofauna and macrofauna at hydrothermal vents. *ISME Communications*. 2021;1(27):1-13. doi:10.1038/s43705-021-00031-1
46. García-Machado E, Laporte M, Normandeau E, et al. Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. *Environmental DNA*. 2022;4(1):117-134. doi:10.1002/edn3.221
47. Gelis ERE, Kamal MM, Subhan B, Bachtiar I, Sani LMI, Madduppa H. Environmental biomonitoring of reef fish community structure with eDNA metabarcoding in the Coral Triangle. *Environ Biol Fishes*. 2021;104(8):887-903. doi:10.1007/s10641-021-01118-3
48. Cavender-Bares J, Kozak KH, Fine PVA, Kembel SW. The merging of community ecology and phylogenetic biology. *Ecol Lett*. 2009;12(7):693-715. doi:10.1111/j.1461-0248.2009.01314.x
49. Graham CH, Storch D, Machac A. Phylogenetic scale in ecology and evolution. *Global Ecology and Biogeography*. 2018;27(2):175-187. doi:10.1111/geb.12686
50. Wanntorp HE, Brooks DR, Nilsson T, et al. Phylogenetic approaches in ecology. *Oikos*. 1990;57(1):119-132. https://www.jstor.org/stable/3565745
51. Mouquet N, Devictor V, Meynard CN, et al. Ecophylogenetics: Advances and perspectives. *Biological Reviews*. 2012;87(4):769-785. doi:10.1111/j.1469-185X.2012.00224.x
52. Hernández CE, Rodríguez-Serrano E, Avaria-Llautureo J, et al. Using phylogenetic information and the comparative method to evaluate hypotheses in macroecology. *Methods Ecol Evol*. 2013;4(5):401-415. doi:10.1111/2041-210X.12033
53. Camper BT, Laughlin ZT, Kanes AS, Manuel RT. A portable and wind resistant drift fence array for arid environments. *bioRxiv*. Published online 2023:1-25. doi:10.1101/2023.09.28.560034
54. Barley AJ, Reeder TW, Nieto-Montes De Oca A, Cole CJ, Thomson RC. A New Diploid Parthenogenetic Whiptail Lizard from Sonora, Mexico, Is the "Missing Link" in the Evolutionary Transition to Polyploidy. *Am Nat*. 2021;198(2):295-309. doi:10.5061/dryad.2z34tmpm2

55. Tucker DB, Colli GR, Giugliano LG, et al. Methodological congruence in phylogenomic analyses with morphological support for teiid lizards (Sauria: Teiidae). *Mol Phylogenet Evol*. 2016;103:75-84.
56. Walker JM, Sullivan BK, Cordes JE. Sanctioned Nomenclature for Triploid Parthenogenetic Colorado Checkered Whiptail and for Other Taxa of the Lizard Genus Aspidoscelis (Family Teiidae). *Herpetol Rev*. 2021;52(3):550-558.
57. R Core Team. R: A language and environment for statistical computing. Published online 2021.

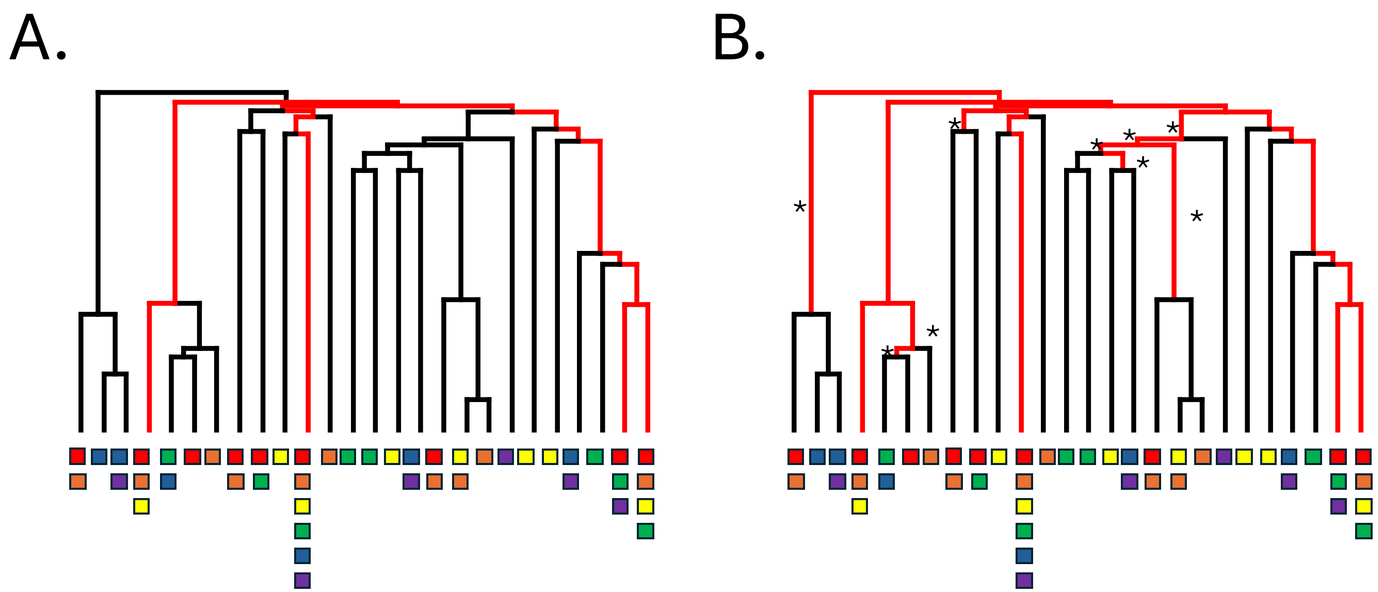
Electronic Supplementary Material

core_community_phylogeny_example.docx

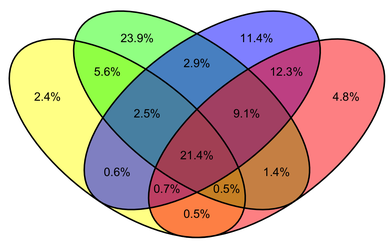
Additional file 1: A simplified system demonstrating the construction of core community phylogenies and calculations of associated metrics.

biom_files.zip

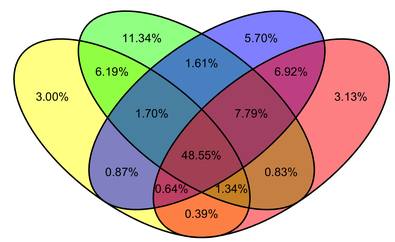
Additional file 2: BIOM files, read counts, taxonomy assignments, microbial phylogenies, and lizard metadata for gut and skin samples used in this study.


A.
B.

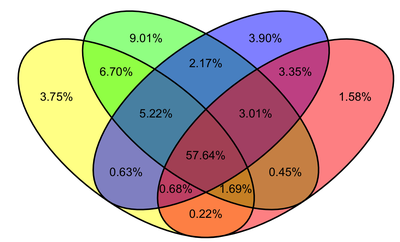
A. ASV Counts (Skin)



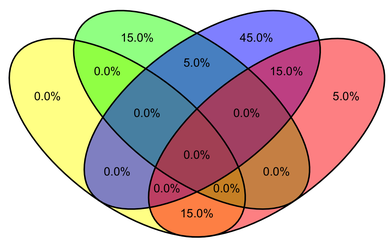
B. Tip-based (Skin)



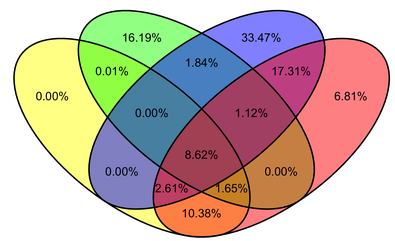
C. Branch-based (Skin)



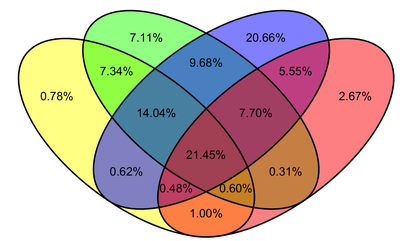
D. ASV Counts (Gut)

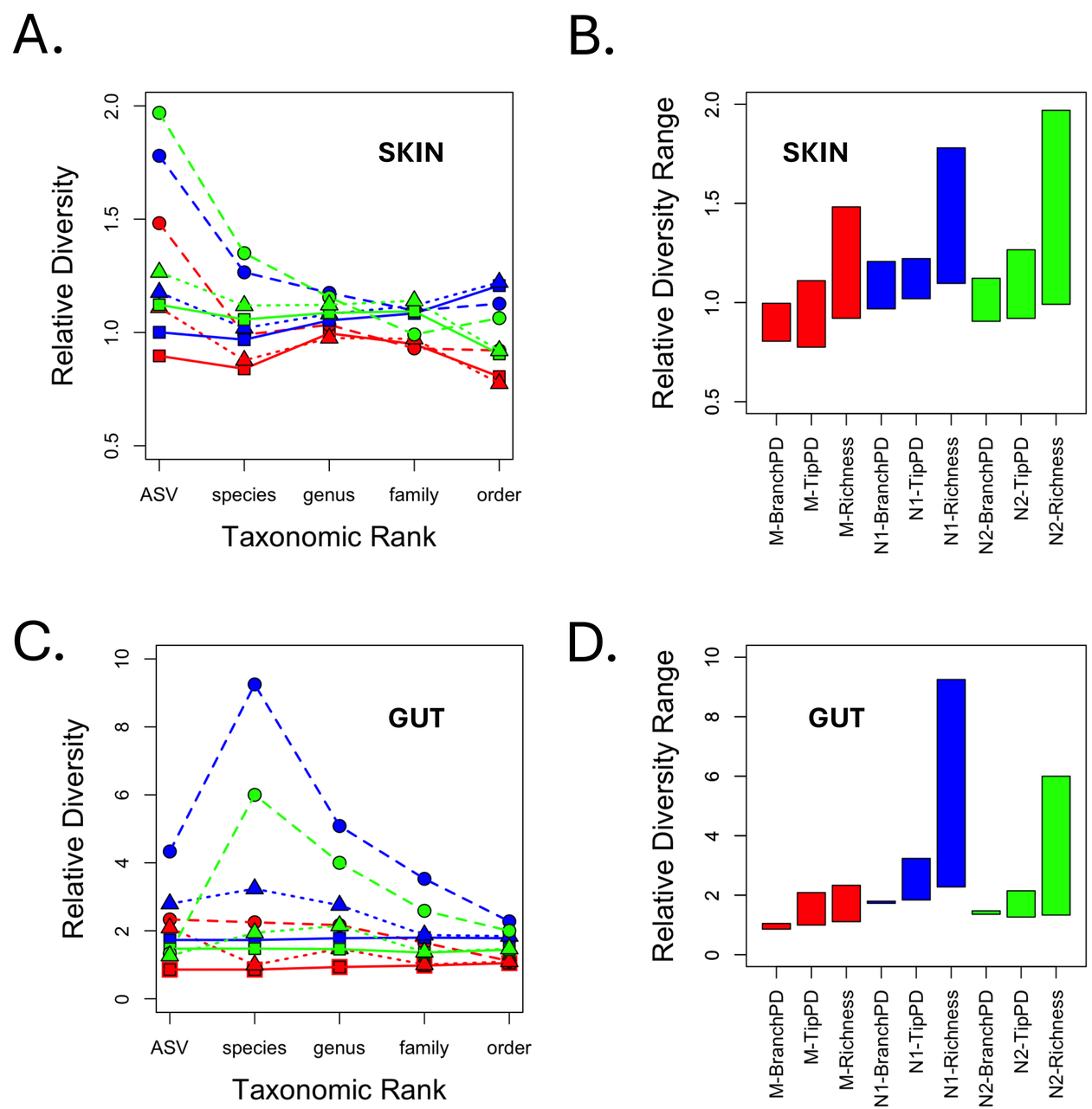


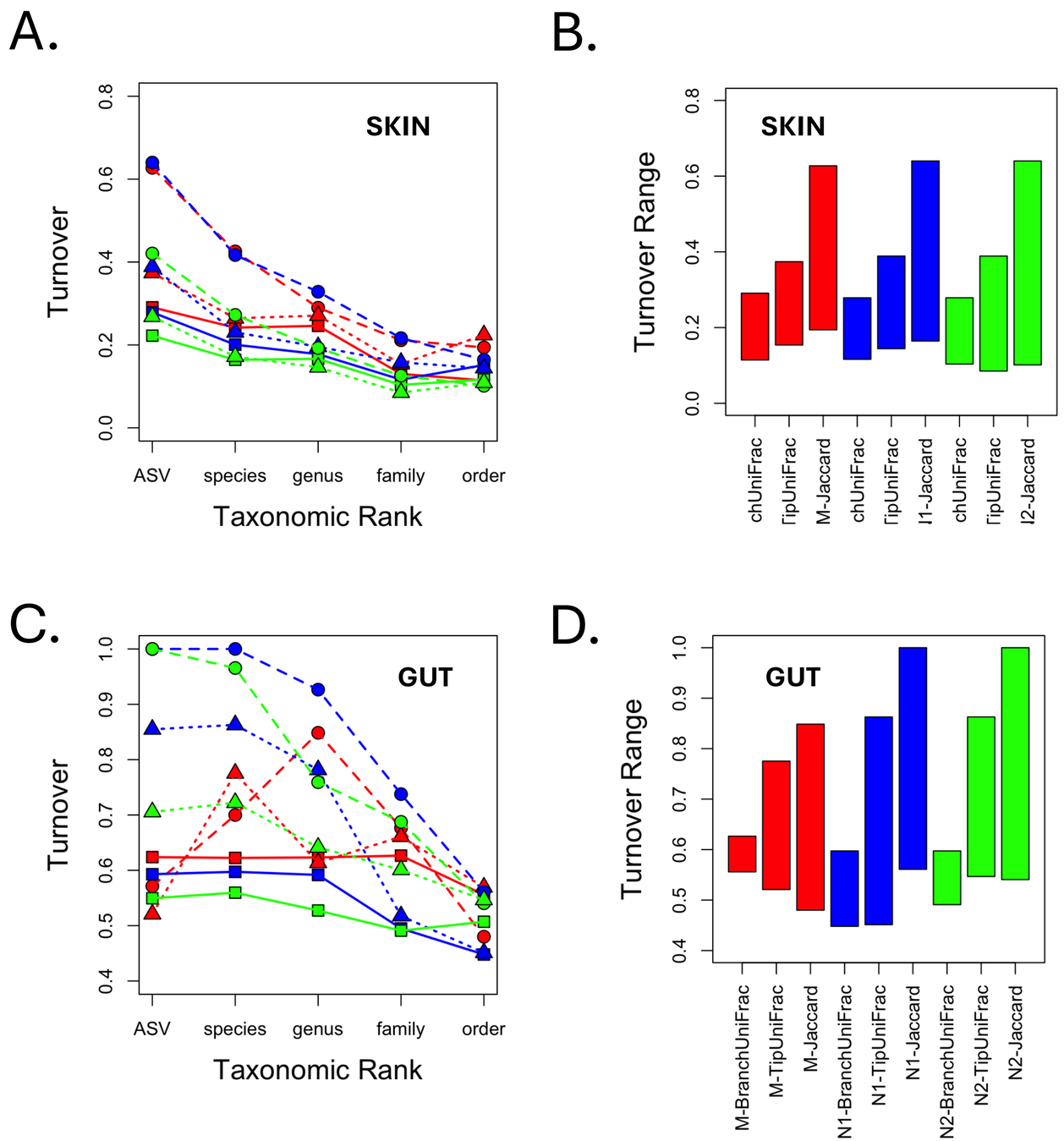
E. Tip-based (Gut)



F. Branch-based (Gut)




A.
SKIN
Relative Diversity
0.5
1.0
1.5
2.0
ASV
species
genus
family
order
Taxonomic Rank
B.
SKIN
Relative Diversity Range
0.5
1.0
1.5
2.0
M-BranchPD
M-TipPD
M-Richness
N1-BranchPD
N1-TipPD
N1-Richness
N2-BranchPD
N2-TipPD
N2-Richness
C.
GUT
Relative Diversity
0
2
4
6
8
10
ASV
species
genus
family
order
Taxonomic Rank
D.
GUT
Relative Diversity Range
0
2
4
6
8
10
M-BranchPD
M-TipPD
M-Richness
N1-BranchPD
N1-TipPD
N1-Richness
N2-BranchPD
N2-TipPD
N2-Richness


A.
SKIN
Turnover
0.0
0.2
0.4
0.6
0.8
ASV
species
genus
family
order
Taxonomic Rank
B.
SKIN
Turnover Range
0.0
0.2
0.4
0.6
0.8
chUniFrac
TipUniFrac
M-Jaccard
chUniFrac
TipUniFrac
I1-Jaccard
chUniFrac
TipUniFrac
I2-Jaccard
C.
GUT
Turnover
0.4
0.5
0.6
0.7
0.8
0.9
1.0
ASV
species
genus
family
order
Taxonomic Rank
D.
GUT
Turnover Range
0.4
0.5
0.6
0.7
0.8
0.9
1.0
M-BranchUniFrac
M-TipUniFrac
M-Jaccard
N1-BranchUniFrac
N1-TipUniFrac
N1-Jaccard
N2-BranchUniFrac
N2-TipUniFrac
N2-Jaccard