Multilevel community assembly of the tadpole gut microbiome

Decio T. Correa[1*], David Rodriguez[2], Carine Emer[3], Daniel Saenz[4], Cory K. Adams[4], Luis Schiesari[5], Mikhail Matz[1], Mathew A. Leibold[6]

1. Department of Integrative Biology, The University of Texas at Austin, Austin, TX, USA
2. Department of Biology, Texas State University, San Marcos, TX, USA
3. Chico Mendes Institute for Biodiversity Conservation, National Centre for Wild Birds Research and Conservation, João Pessoa, PB, Brazil
4. Southern Research Station, US Forest Service, Nacogdoches, TX, USA
5. School of Arts, Sciences and Humanities, University of São Paulo, São Paulo, SP, Brazil
6. Department of Biology, University of Florida, Gainesville, FL, USA

* Corresponding author, decio.tadeu@gmail.com

Abstract

The assembly of local communities is likely to reflect the effects of local environmental factors associated with filters that act at larger spatial scales. Dissecting these multiscale effects remains a timely challenge that is particularly important for host-associated microbiomes. We investigated the relative roles of local selection (due to host species identity) and regional effects (due to water body identity) on the community structure of bacteria in the gut of tadpoles from three biogeographic areas and used graph theory and metanetwork approaches to explore and illustrate the distribution of bacteria across different ponds. The pond of origin, which represents a regional species pool of bacteria, was in general more important in shaping the gut microbiome of tadpoles than host species identity. The resulting metanetworks are modular and indicate relatively few species of bacteria occurring in more than one pond. Thus, each pond represents a relatively distinct species pool of bacteria available for community assembly of the tadpole microbiomes. Our findings indicate that microbiome community assembly in amphibian larvae, as in many other communities, is a multiscale process with important regional effects that constrain how local (i.e. host-dependent) filters act to influence microbiome community composition.

Introduction

The origin and assembly of host-associated microbial communities (microbiomes) is a complex process of fundamental importance in helping us understand the symbiotic relationship between hosts and their microbes [1, 2]. Various factors can play a role in the community assembly of microbiomes at multiple levels. Perhaps the most obvious is the effect of host diet on the gut microbiome, which has been shown in many organisms from insects to humans [3–6]. Other local factors include host sex, health state, and genetics [7–11]. These effects can be thought of as local 'filters' to microbial community assembly. Here we evaluate these effects as they are determined by host species identity only, recognizing that there are likely to be other more subtle effects involving inter-individual factors. However, there can also be processes that constrain which microbes are available to colonize any given host. Here, we focus on factors that determine the microbial species pool available to colonize the host and evaluate these by the identity of the site of the host community.

We view community assembly of microbiomes as being driven by metacommunity processes [12]. This recognizes that the assembly of a local community depends strongly on the species available to colonize that community, known as the regional species pool [13, 14]. From that pool, if a species is able to disperse to a given community, its persistence in that community will depend on local demographic processes and on the ability of the species to tolerate the local biotic and abiotic conditions, which is known as the local selection component of community assembly

54 [15]. Therefore, communities with identical environmental conditions could have a
55 completely different species composition if colonized by different species pools [16].
56 While most work in metacommunity ecology has focused on macroorganisms, these
57 ideas should apply equally well to microbes [1, 2, 12, 17, 18].

58 Host-associated microbial communities appear to be influenced by effects
59 related to the individual host or individual species such as diet, physiology, and
60 genotype [10, 19, 20], as well as regional and historical processes, as in the case of
61 exposure to potentially different species pools of microorganisms [21–26], or vertical
62 and horizontal transmission of the microbiome [27, 28]. Considering the importance
63 that the microbiome can have to host health and well-being, understanding the relative
64 importance of effects acting at different levels is fundamental, but studies explicitly
65 tackling that question are still lacking [29].

66 We analyzed variation in tadpole microbiomes to evaluate the effects of local
67 environmental filters by focusing on host species identity, and regional effects by
68 focusing on the locality of the water body. We assume that different microbes would be
69 differentially favored by selection associated with the physiology and diet of different
70 tadpole species, and different microbes may be present at different sites due to either
71 site-specific habitat variables or to isolation and other differences among water bodies.
72 Our studies were conducted across three different geographic locations to test if they
73 show consistent patterns. We also used graph theory and metanetwork approaches to

74 explore the degree to which each pond consists of a unique species pool, and how they
75 are connected to each other in terms of shared bacterial species.
76 The gut microbiome of tadpoles is critically important to their digestive function
77 [30], and some evidence shows gut microbiomes from the same host species are more
78 similar to each other than from different species [31]. Further, there are often multiple
79 species of tadpole per water body, which represent different hosts with different
80 characteristics that are exposed to the same pool of bacteria. Many of these tadpole
81 species can occupy a variety of different water bodies, each with potentially distinct
82 bacterial species pools, which makes them a suitable system to test questions regarding
83 the effects of environment and host identity on microbiome assembly [32]. We evaluate
84 two hypotheses: i) if the bacterial community of each pond is different across the
85 landscape, then the pond of origin would be more important in determining the gut
86 microbiome of tadpoles, i.e., the microbiomes of different species within each pond
87 would be more similar to each other than the microbiomes of the same species across
88 distinct ponds, and the metanetworks would be structured into modules according to
89 each pond; ii) if ponds have very similar bacterial communities across the landscape,
90 then the species of tadpole would be more important in determining their gut
91 microbiome, i.e., tadpoles of the same species across ponds would have more similar
92 microbiomes than tadpoles of different species from the same pond, and the
93 metanetworks would be structured into modules according to each species.
94

Methods

Study area

We sampled tadpoles from 22 water bodies located in three different localities: i) six water bodies in the Boracéia Biological Station, located in the Atlantic Forest east of the state of São Paulo, Brazil, hereafter BR; ii) nine water bodies in the David Crockett and Angelina National Forests, located in the pine forest ecosystem in eastern Texas, United States, hereafter ET; and iii) seven water bodies in the Edwards Plateau, a mix of grassland and juniper/oak woodlands in the central part of Texas, United States, hereafter CT. We classified water bodies as either lentic or lotic. For ET ponds, we did a more thorough environmental characterization quantifying in each pond the dissolved oxygen (mg/L), pH, conductivity (uS/cm), chlorophyll a (ug/L), and water temperature (^{o}C) using a Eureka multiparameter sonde, and the absence or presence of piscivorous (Bass) and/or insectivorous fish (Green Sunfish). Even though three water bodies in BR were lotic, we will refer to all of them as *ponds* for convenience.

Tadpole and microbiome sampling

In each pond we collected tadpoles with a dipnet that was swept through the pond. We then placed the larvae in a sterile plastic bag (Whirl Pak®) filled with water from their place of origin. Environmental microorganisms from water from CT and ET were sampled by filtering approximately 50 ml of water from three different places through a sterile syringe filter with a 0.45 μm cellulose acetate membrane (VWR). From

116 BR, approximately 2L of water was collected in a sterile plastic bag (Whirl Pak®), mixed,
117 and 100 mL was vacuum-filtered through a 0.45 μm cellulose acetate membrane. The
118 sediment in all ponds was sampled in three locations in each pond by removing the first
119 three centimeters with a sterile plastic straw of 1.3 cm of diameter and placed in a
120 sterile tube.

121 In the lab, tadpoles were euthanized by overdose in a solution of Milli-Q® filtered
122 water and tricaine methanesulfonate (MS 222). After the death was confirmed, each
123 individual was dissected using instruments that were cleaned in 100% ethanol and
124 flame-sterilized. The entire tadpole gut was removed and placed directly in a MoBio
125 PowerSoil® bead tube. The sediment from each pond was homogenized, centrifuged to
126 remove the excess water, and an aliquot of approximately 0.2 g was used for DNA
127 extraction. DNA from the water sample was extracted from the entire filter membrane.
128 We performed DNA extraction using MoBio PowerSoil® DNA isolation kit following the
129 manufacturer's recommendation. The same procedures without a real sample were
130 repeated to count as negative control samples, which did not have DNA when quantified
131 using Qubit™ dsDNA high sensitivity assay.

132 We amplified the V4 hypervariable region of the 16s rRNA gene (515F, 806F). CT
133 samples were amplified using the version F: GTGYCAGCMGCCGCGGTA / R:
134 GGACTACHVGGGTWTCTAAT and ET and BR the version F: GTGYCAGCMGCCGCGGTAA /
135 R: GGACTACNVGGGTWTCTAAT. The library preparation and sequencing (Illumina MiSeq
136 2x250bp) of the CT samples was realized at the Genomic Sequencing and Analysis

137 Facility (wikis.utexas.edu/display/GSAF) at The University of Texas at Austin following
138 their protocols. The library preparation and sequencing (Illumina MiSeq 2x250bp) of the
139 ET samples was realized at the Argonne National Laboratory (www.anl.gov) following
140 their protocols. For BR samples, we prepared the libraries following the Earth
141 Microbiome Project protocol (www.earthmicrobiome.org), with the exception of using
142 20uL PCR reactions and dual-index barcodes instead. Samples were sequenced in an
143 Illumina MiSeq platform (2x300 bp paired-end sequences).
144 The sequences were processed using the dada2 pipeline [33] in the software
145 Microsoft R Open v 3.5.1 [34]. Sequences from each site were processed separately.
146 Briefly, primers, adapters, and barcodes were removed and the quality profile of the
147 reads were visually inspected using the function plotQualityProfile aggregating over all
148 the samples from a site. BR, CT, and ET reads were truncated at 200 and 120, 200 and
149 180, and 200 and 160 forward and reverse reads, respectively. Sequencing errors were
150 calculated, sequences were clustered as Amplicon Sequence Variants based on the
151 DADA2 algorithm [33] and paired ends were merged. Chimeras were removed using the
152 consensus method in the function removeBimeraDenovo. Taxonomy was assigned
153 based on the Silva database v128 [35]. ASV sequences were exported to Qiime2 2018.4
154 [36], aligned using the MAFFT algorithm [37], and a phylogenetic tree was constructed
155 using FastTree [38]. We discarded all ASVs that were classified as Archaea, chloroplast,
156 mitochondria, or that were not at least assigned to bacteria. Hereafter, we will refer to
157 ASVs as species of bacteria.

158

159 **Data analysis**

160 **Diversity**

161 For each sample we constructed a rarefaction curve and compared richness and
162 Faith's phylogenetic diversity [39]. To avoid sequencing depth effects, each sample was
163 rarefied 100 times to the number of reads of the sample with the lowest coverage using
164 the R package BAT [40]. For comparison, we considered the median richness from all
165 rarefactions from a sample as our richness measure.

166

167 **Species and pond effects**

168 We analyzed our data as compositional data (termed CODA) [41]. It requires a
169 centered log-ratio transformation of the read counts and therefore cannot have zeros,
170 so we added a pseudocount of one to all ASVs in all samples. We transformed the data
171 using the codaSeq.clr function from the R package CoDaSeq [42, 43]. We also analyzed
172 data that takes into account phylogenetic relatedness of microbes by using the PHILR
173 transformation from the R package philr [44]. This metric takes into account the
174 phylogenetic relationship between the ASVs, which is equivalent to the Unifrac metric
175 [45], but it considers the compositional nature of the data [44]. We then ran a Principal
176 Component Analysis (PCA) on the Euclidean distance matrix of the CODA- and PHILR-
177 transformed data to visualize the relationship between samples and tested if samples
178 from tadpoles, water, and sediment are different from each other using a Permutational

179 Multivariate Analysis of Variance (perMANOVA). If a difference was detected we then
180 used additional perMANOVAs for pairwise comparisons with Bonferroni correction.
181 To first test if either the identity of the species of tadpole or the pond of origin
182 are more related to the gut microbiome composition we used a perMANOVA to fit a
183 model with both pond and species identity to the CODA- and PHILR-transformed data
184 using the function adonis2 from the R package vegan [46]. The significance of the
185 marginal effects was tested based on 999 permutations. In addition, to test for the
186 unique effects of the species of tadpole or the pond of origin as well their shared effects
187 on the gut microbiome we performed variance partitioning and Redundancy Analysis
188 (RDA) [47] using the function varpart in the R package vegan [46]. The significance of
189 each unique component in the variation partitioning was tested using the vegan
190 function anova.cca with 1000 permutations. We repeated perMANOVA and variation
191 partitioning analyses to the data grouping the species of bacteria at Genus and Family
192 level.
193
194 **Inter-pond variation**
195 We also investigated if there was evidence for dispersal limitation in the bacteria
196 from water and sediment using multiple regression on distance matrices using the
197 function MRM from the R package ecodist [48]. We used the Euclidean distance matrix
198 based on the PHILR and CODA transformations as the dependent variable and a
199 geographical distance matrix between sites as the independent variable. In addition, to

200 account for environmental effects, we also include in the model a distance matrix based
201 on the type of the pond (lentic or lotic) for BR, and a distance matrix based on the
202 environmental variables measured for ET. The latter was a Gower distance matrix [49]
203 created by the function dist.ktab from the R package ade4 [50]. Quantitative variables
204 were Gower-standardized (divided by maximum minus minimum). All CT ponds were
205 lentic and there were no other variables besides spatial distance to include in the
206 model. For models where distance was significant in the MRMs, we applied a Partial
207 Mantel Correlogram approach using the function mpmcorrelogram from the R package
208 with the same name [51]. That tests, at several distance classes, in our case estimated
209 by Sturge's rule [52], the relationship between water and sediment microbiome with
210 geographical distance while controlling for environmental effects.
211
212 **Metanetworks**
213 We further explored the structure of three different metanetworks within each
214 locality (BR, CT, and ET) to understand the distribution of bacteria across ponds. Shared
215 species (here bacteria) connect different ponds in a metacommunity perspective [53,
216 54]. We built three bacteria metanetworks for each locality: (i) the tadpole microbiome
217 metanetwork, composed by the bacteria living in the gut of all tadpoles within a pond,
218 independently of the tadpole species, (ii) the water metanetwork, with bacteria found in
219 the water, and (iii) the sediment metanetwork, composed by bacteria found in the
220 sediment of each pond. Then, we built an a_{mn} adjacency matrix for each metanetwork in

221 which *m* corresponds to a single pond (i.e., considering bacteria from water, sediment
222 or gut of tadpoles), and *n* to each bacterium. The *mn* element represents the
223 presence/absence of species of bacteria *n* in a pond *m,* and is represented by a link in
224 the graphical representation. Therefore, each node in our metanetwork is either a
225 bacterium or a pond; when the same bacterium is found in more than one pond, there
226 is a link connecting them, forming the edges of the metanetwork.
227 To test whether ponds consist of different bacterial communities, we calculated
228 the modularity of the metanetworks using the FastGreedy algorithm [55, 56] in the
229 software Modular v 0.1 [57]. We tested the significance of the observed modularity
230 against two different null models with 100 permutations each: the Erdős-Rényi [58], in
231 which pond-bacteria interactions are connected randomly, and the 'null model 2' [59],
232 in which the probability of a pond-bacteria interaction is proportional to their number of
233 links in the observed matrix. To identify the bacteria with higher potential to connect
234 the metanetwork through shared occurrence in a higher number of ponds, we
235 calculated the degree (k) for each metanetwork using the software Pajek v. 4.10 [60].
236 Degree measures how many links each node establishes in its correspondent network.
237 For the bacteria, it represents the number of ponds in which it occurs; for the pond, it
238 represents the number of bacteria it has (from tadpoles, water or sediment). For each
239 metanetwork we also calculated the proportion of bacteria found uniquely in p_1, p_2, …,
240 p_n, where n is the total number of ponds in one locality.
241

242 **Results**

243 **Diversity**

244 We registered 19 species of tadpole in all ponds: three in CT, six in ET, and
245 eleven in BR (Table S1). After the filtering steps, there were a total of 533 777, 1 289
246 662, and 2 064 298 16s rRNA gene fragment reads for CT, ET, and BR, respectively. The
247 rarefaction curve shows that all samples reached an asymptote, even the ones with the
248 lowest number of reads (Figure S1). For richness and phylogenetic diversity measures of
249 bacteria, CT, ET, and BR samples were rarefied to 3 950, 1 791, and 3 916 reads,
250 respectively. There was no significant difference among CT samples in terms of richness,
251 but there was in terms of phylogenetic diversity, with sediment samples having the
252 highest diversity and *Rana berlandieri* tadpoles the lowest (Figure S2). Sediment
253 samples had the highest richness and phylogenetic diversity in ET and *Hylodes phyllodes*
254 the highest in BR (Figures S3 and S4). A great part of the tadpole microbiome was
255 composed by Fusobacteria, Firmicutes, Proteobacteria, and Bacteroidetes, while
256 sediment and water samples were mostly composed by Proteobacteria (Figures S5-S7).
257

258 **Species and pond effects**

259 The composition of the tadpole gut microbiome is different from both water and
260 sediment samples in all localities (Figure 1, Figure S8, Table S2). Also, in most cases,
261 tadpole gut microbiome samples cluster more with other samples from the same pond
262 than with samples from the same species from a different pond (Figure 1, Figures S8-

263 S10). However, within a pond, there can be separation by species and, in fact, the
264 perMANOVA shows that both species identity of tadpole and pond of origin are
265 significant predictors of the tadpole gut microbiome in all localities when using either
266 PHILR- or CODA-transformed data, except for CT CODA, in which species identity was
267 only marginally significant (Table 1). In CT and ET, pond of origin was a stronger
268 predictor of the tadpole gut microbiome than tadpole species identity. The overall
269 results are consistent over all the taxonomic levels considered (Tables S3-S6).
270 For CT and ET, the pond of origin explained considerably more of the variance in
271 tadpole gut microbiome than the species of tadpole (Figure 2). For BR, the tadpole
272 species explained a slightly higher proportion of the variance in their gut microbiome.
273 However, most of the variation in that environment is explained by the joint effects of
274 tadpole species and pond of origin, i.e., the variation that could not be separated in
275 unique components (Figure 2). Again, the overall results are comparable across all the
276 taxonomic levels considered (Tables S7-S8).
277
278 **Inter-pond variation**
279 Only the bacterial communities of water from CT were significantly related to the
280 spatial distance between ponds (Table S9). However, that was the only dataset where
281 environmental data is not available and could not be accounted for, allowing for the
282 possibility that such distance effects are due to spatially structured environmental
283 factors. Partial Mantel Correlograms showed that, for both CODA- (p_{adj} = 0.044) and

284 PHILR-transformed data (p_{adj} = 0.022), there is a spatial signal only for the first distance
285 range, from 174 to 6 223 m. ET water and sediment communities were significantly
286 related to environmental variables for PHILR-transformed data and BR water
287 communities were significantly related to both PHILR- and CODA-transformed data
288 (Table S9).
289
290 **Metanetworks**
291 All metanetworks are significantly modular according to the two null models
292 used ($p < 0.001$ for all null models; Figure 3, Figure S11). In most cases, each module
293 corresponds to a single pond (Figure 3, Figure S11 and Table S10). Metanetworks of
294 water and sediment are more modular than those of tadpoles for all localities (Figure 3,
295 Figure S11). In all cases, the vast majority of bacteria occur in a single pond (k = 1),
296 varying from 63.2% in East Texas to 91.3% in Brazil (Figure 4, Figures S12-S16). Only 0.1
297 to 1.1% of bacteria were present in all ponds (Figure 4, Figure S12-S16). From those, for
298 BR and ET they were mostly from the phyla Bacteroidetes and Proteobacteria and
299 Orders Bacteroidales and Desulfovibrionales. CT had only three bacteria present across
300 all ponds. From the ponds' perspective, their degree, which corresponds to the number
301 of unique links they have in the metanetwork, varied from 118 to 615 for water, 90 to
302 478 for sediment, and 241 to 2 751 for bacteria from tadpoles (Figures S17-S19).
303

304 **Discussion**

305 We found that microbial community assembly of tadpole microbiomes was
306 structured by processes happening at different levels and broadly resemble findings
307 from macroorganisms [1, 15, 61–63]. At higher levels, there are inter-pond differences
308 in microbial communities due to site-specific variables, isolation, and/or stochasticity. It
309 was not our goal to understand the drivers of inter-pond differences, which is ultimately
310 responsible for the different bacterial species pool tadpoles are exposed to. However, in
311 some cases, for the ponds where we had environmental data there is a relationship
312 between environmental variables and the bacterial community of water and/or
313 substrate. In these cases, each pond could act as a regional scale environmental filter for
314 bacteria. For CT, we found some evidence for a spatial correlation in bacterial
315 communities among nearby ponds, which could indicate dispersal limitation structuring
316 the pond communities, but that was the only dataset for which we did not have
317 environmental variables. Nevertheless, independently of the causes of the differences
318 among ponds, each one of them constitutes a unique species pool of colonizers for the
319 microbiomes of tadpoles. As tadpoles develop, the community assembly of their
320 microbiome is restricted by that species pool of microbes. Each tadpole comprises a
321 smaller-scale filter to a subset of the microbes within the species pool. Therefore, even
322 if similar habitats (same species of tadpole in our case) are exposed to different species
323 pools, they will not end up with the same microbial community.

The composition and diversity of the gut microbiome of tadpoles can be variable across species. The microbiome of tadpoles from this study is slightly different than previous studies. Proteobacteria and Firmicutes dominated in other studies [31, 64–66] but here we found that Fusobacteria were also important. Alpha and phylogenetic diversity were also mostly higher in our study than in previous ones [31, 64–66]. The knowledge about the gut microbiome of tadpoles is still limited to a few species and quantitative comparisons of diversity across studies is problematic because of differences in sampling procedures, sample processing, data cleaning, and the algorithms and databases. However, overall differences in microbiome composition could be due to the exposure of tadpoles to different pools of colonist microbes.

The pond of origin of the tadpoles was more important in determining their gut microbiome than the species of the tadpole in both of the Texas locations. We assume the same species of tadpole has similar requirements in terms of microbes and/or constitutes a similar patch for colonization by microbes. There were only a few bacteria in common across all ponds in all three localities, therefore each pond constitutes a unique bacterial species pool from which the tadpoles can be colonized. In Brazil, the pond of origin and species of tadpole showed similar relationships with the gut microbiome of tadpoles. However, most of variation in Brazil was due to variation that could not be separately attributed to either because of the uniquely distinct tadpole faunas of ponds.

344 The limited and unique pool of bacteria tadpoles are exposed to and the higher
345 similarity within a pond can be visualized on the metanetworks of bacteria from water,
346 sediment, and tadpole gut. The metanetworks are modular and each pond constitutes a
347 unique community of bacteria. The vast majority of bacteria occur only in a single pond,
348 and just a few are present in all of them. In fact, for water and sediment, the maximum
349 number of bacteria found in all ponds was only eight. That number would probably
350 increase if more sediment or water was sampled. On the other hand, we likely sampled
351 all species of tadpoles from each water body and the proportion of bacteria from their
352 gut found in all water bodies is still very low (0.1-1.1%) despite similar phyla found in all
353 of them. The major bacteria in tadpole guts belong to orders Bacteroidales and
354 Desulfovibrionales, which do not overlap with the most widespread taxa found in water
355 or sediment. It is common to not find much overlap between bacteria from environment
356 and gut of aquatic organisms [67, 68]. Likewise, we show that the gut microbiome of
357 tadpoles is distinct from both water and sediment from the ponds suggesting that the
358 gut constitutes a strong environmental filter that favors certain taxa. Desulfovibrionales,
359 for example, is a group of sulfur-reducing bacteria that is mostly anaerobic [69].
360 Therefore, at least certain parts of the digestive tract of tadpoles might be anaerobic
361 [31, 70]. Pryor and Bjorndal [30] detected significant levels of fermentation in the
362 hindgut of bullfrog tadpoles. It is unknown to which extent other species of tadpoles are
363 hindgut fermenters. For those that are, even though they might not have a core

364 microbiome, at least not at fine-scale taxonomy levels, there may be a functional core
365 gut microbiome.
366 In terrestrial and aquatic environments, the importance of the regional species
367 pool in shaping the microbial communities can be noticed even when not tested
368 directly. In many species, from insects to humans, individuals likely exposed to the same
369 regional species pool, such as being from the same environment, have more similar
370 microbiomes [21–26]. There are many difficulties in quantifying and testing how
371 regional effects affect community assembly [29, 62], but comparing results with
372 predictions from theory can lead to important insights. If hosts from different
373 environments have similar microbiomes it could indicate that (i) there is not much
374 difference in the regional species pools between environments, (ii) the hosts move
375 between species pools, or (iii) there are other methods for microbiome acquisition such
376 as transmission from conspecifics. For example, Kueneman *et al*. [71] and McKenzie *et*
377 *al*. [32] studying the skin microbiome of tadpoles found that the identity of the species
378 was the strongest predictor of the skin microbiome, with the water body of origin
379 explaining additional [71] or no variance at all [32]. Vertical transmission of the
380 microbiome is unlikely in amphibians [72], and they probably acquire their microbiome
381 from the environment at every generation [73]. Therefore, the results from Kueneman
382 *et al*. [71] and McKenzie *et al*. [32] indicate there is at least a partially shared species
383 pool of microbes across their sampling sites.

384 The enormous intra and inter-specific variation in the microbiomes of different
385 species of animals and plants can be at least partially explained by differences in
386 regional species pool during the community assembly of the microbiome. The gut
387 microbiome of animals can be stable over time [63, 74] and can also be resistant to
388 invasions and colonization by other bacteria [75–77]. Therefore, if the community
389 assembly of individuals of the same species occurred while they were exposed to
390 different species pools, historical effects can cause lasting effects in the development of
391 the microbiome [78]. Moreover, recent efforts to manipulate the microbiome of plants
392 and animals to improve health and/or increase productivity might not be as effective if
393 the manipulation is done only at the level of the individual host, i.e., at the local scale.
394 As stated by Chase [16] "if local and regional processes determine the community
395 composition, then both processes need to be restored to achieve the desired
396 community". Our results indicate that the same logic is valid for microbes.
397 In summary, we showed that local assembly of host-associated bacterial
398 communities are affected by regional scale processes, more specifically changes in the
399 regional species pool of colonizers. The inter-pond variation in bacterial community can
400 be due to stochastic, historical, spatial, and/or environmental processes or even the
401 result of local evolution [79]. Nevertheless, microbiome assembly, as any other
402 community, is a multiscale process. Microbial and microbiome ecology could be better
403 linked to ecological theory by considering the multiscale dynamics of community
404 assembly [12, 17, 80, 81]. Because it is much easier to quantify effects at the scale of the

405 individual host or host species, i.e., local effects [29, 62], the role of regional effects is
406 probably underestimated. Regional effects, however, appear to be a fundamental piece
407 of the community assembly puzzle that can be used to understand variation in
408 microbiomes.
409

Acknowledgements

We thank Ananda Brito de Assis and Carlos Arturo Navas for laboratory access during field work in Brazil and Christopher Schalk and James Childress for help in the field in East Texas. We thank the Zoology Museum of University of São Paulo (MZUSP), The National Forests and Grasslands of Texas, and several private landowners for granting permission to access sampling sites and ICMBio (47263-1), Texas Parks and Wildlife (SPR-0414-067), and IACUC (AUP-2014-00286) for collecting and animal care permits. Funding was provided by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasil, grant number 458796/2014-0) and Texas Ecolab. DTC received a Science Without Borders scholarship during the research (CAPES 1176/13-7). Finally, we thank Daniel Bolnick, Michel Ryan, Ulrich Mueller, and members of Leibold and Matz labs for comments throughout the project.

423 **References**

424 1. Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA. The
425 application of ecological theory toward an understanding of the human
426 microbiome. *Science* 2012; **336**: 1255–62.
427 2. Christian N, Whitaker BK, Clay K. Microbiomes: unifying animal and plant systems
428 through the lens of community ecology theory. *Front Microbiol* 2015; **6**: 1–15.
429 3. Broderick NA, Raffa KF, Goodman RM, Handelsman J. Census of the Bacterial
430 Community of the Gypsy Moth Larval Midgut by Using Culturing and Culture-
431 Independent Methods. *Appl Environ Microbiol* 2004; **70**: 293–300.
432 4. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The Effect of Diet
433 on the Human Gut Microbiome: A Metagenomic Analysis in Humanized
434 Gnotobiotic Mice. *Sci Transl Med* 2009; **1**: 6ra14-6ra14.
435 5. Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Knight R, Caporaso JG, et al.
436 Individuals' diet diversity influences gut microbial diversity in two freshwater fish
437 (threespine stickleback and Eurasian perch). *Ecol Lett* 2014; **17**: 979–87.
438 6. Gomez A, Rothman JM, Petrzelkova K, Yeoman CJ, Vlckova K, Umaña JD, et al.
439 Temporal variation selects for diet–microbe co-metabolic traits in the gut of
440 Gorilla spp. *ISME J* 2016; **10**: 514–526.
441 7. Fierer N, Hamady M, Lauber CL, Knight R. The influence of sex, handedness, and
442 washing on the diversity of hand surface bacteria. *Proc Natl Acad Sci U S A* 2008;
443 **105**: 17994–9.
444 8. Pflughoeft KJ, Versalovic J. Human Microbiome in Health and Disease. *Annu Rev*
445 *Pathol Mech Dis* 2012; **7**: 99–122.
446 9. Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability
447 and resilience of the human gut microbiota. *Nature* 2012; **489**: 220–30.
448 10. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human
449 genetics shape the gut microbiome. *Cell* 2014; **159**: 789–799.
450 11. Bolnick DI, Snowberg LK, Hirsch PE, Lauber CL, Org E, Parks B, et al. Individual diet
451 has sex-dependent effects on vertebrate gut microbiota. *Nat Commun* 2014; **5**:
452 4500.
453 12. Miller ET, Svanbäck R, Bohannan BJM. Microbiomes as Metacommunities:
454 Understanding Host-Associated Microbes through Metacommunity Ecology.
455 *Trends Ecol Evol* 2018; **33**: 926–935.
456 13. Ricklefs RE. Community Diversity: Relative Roles of Local and Regional Processes.
457 *Science (80-)* 1987; **235**: 167–171.
458 14. Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, et al.
459 The metacommunity concept: a framework for multi-scale community ecology.
460 *Ecol Lett* 2004; **7**: 601–613.
461 15. Vellend M. Conceptual synthesis in community ecology. *Q Rev Biol* 2010; **85**: 183–
462 206.

463 16. Chase JM. Community assembly: when should history matter? *Oecologia* 2003;
464 **136**: 489–98.
465 17. Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, et al.
466 The role of ecological theory in microbial ecology. *Nat Rev Microbiol* 2007; **5**:
467 384–392.
468 18. Fierer N, Ferrenberg S, Flores GE, González A, Kueneman J, Legg T, et al. From
469 Animalcules to an Ecosystem: Application of Ecological Concepts to the Human
470 Microbiome. *Annu Rev Ecol Evol Syst* 2012; **43**: 137–155.
471 19. Phillips CD, Phelan G, Dowd SE, McDonough MM, Ferguson AW, Delton Hanson J,
472 et al. Microbiome analysis among bats describes influences of host phylogeny, life
473 history, physiology and geography. *Mol Ecol* 2012; **21**: 2617–2627.
474 20. Fitzpatrick BM, Allison AL. Similarity and differentiation between bacteria
475 associated with skin of salamanders (*Plethodon jordani*) and free-living
476 assemblages. *FEMS Microbiol Ecol* 2014; **88**: 482–94.
477 21. Staubach F, Baines JF, Künzel S, Bik EM, Petrov DA. Host Species and
478 Environmental Effects on Bacterial Communities Associated with Drosophila in
479 the Laboratory and in the Natural Environment. *PLoS One* 2013; **8**: e70749.
480 22. Schloss PD, Iverson KD, Petrosino JF, Schloss SJ. The dynamics of a family's gut
481 microbiota reveal variations on a theme. *Microbiome* 2014; **2**: 25.
482 23. McCord AI, Chapman CA, Weny G, Tumukunde A, Hyeroba D, Klotz K, et al. Fecal
483 microbiomes of non-human primates in Western Uganda reveal species-specific
484 communities largely resistant to habitat perturbation. *Am J Primatol* 2014; **76**:
485 347–354.
486 24. Eichmiller JJ, Hamilton MJ, Staley C, Sadowsky MJ, Sorensen PW. Environment
487 shapes the fecal microbiome of invasive carp species. *Microbiome* 2016; **4**: 44.
488 25. Hernández-Gómez O, Hoverman JT, Williams RN. Cutaneous Microbial
489 Community Variation across Populations of Eastern Hellbenders (*Cryptobranchus*
490 *alleganiensis alleganiensis*). *Front Microbiol* 2017; **8**: 1–16.
491 26. Muletz Wolz CR, Yarwood SA, Campbell Grant EH, Fleischer RC, Lips KR. Effects of
492 host species and environment on the skin microbiome of Plethodontid
493 salamanders. *J Anim Ecol* 2018; **87**: 341–353.
494 27. Sanders JG, Powell S, Kronauer DJC, Vasconcelos HL, Frederickson ME, Pierce NE.
495 Stability and phylogenetic correlation in gut microbiota: Lessons from ants and
496 apes. *Mol Ecol* 2014; **23**: 1268–1283.
497 28. Moeller AH, Suzuki TA, Phifer-Rixey M, Nachman MW. Transmission modes of the
498 mammalian gut microbiota. *Science (80-)* 2018; **362**: 453–457.
499 29. Lindström ES, Langenheder S. Local and regional factors influencing bacterial
500 community assembly. *Environ Microbiol Rep* 2012; **4**: 1–9.
501 30. Pryor GS, Bjorndal KA. Symbiotic fermentation, digesta passage, and
502 gastrointestinal morphology in bullfrog tadpoles (*Rana catesbeiana*). *Physiol*
503 *Biochem Zool* 2005; **78**: 201–215.

31. Vences M, Lyra ML, Kueneman JG, Bletz MC, Archer HM, Canitz J, et al. Gut bacterial communities across tadpole ecomorphs in two diverse tropical anuran faunas. *Sci Nat* 2016; **103**: 25.
32. McKenzie VJ, Bowers RM, Fierer N, Knight R, Lauber CL. Co-habiting amphibian species harbor unique skin bacterial communities in wild populations. *ISME J* 2012; **6**: 588–96.
33. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016; **13**: 581–583.
34. Microsoft and R Core Team. Microsoft R Open. 2018. Redmond, Washington.
35. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* 2014; **42**: D643–D648.
36. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ Prepr* 2018; **6**: e27295v2.
37. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013; **30**: 772–780.
38. Price MN, Dehal PS, Arkin AP. FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS One* 2010; **5**.
39. Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv* 1992; **61**: 1–10.
40. Cardoso P, Rigal F, Carvalho JC. BAT - Biodiversity Assessment Tools, an R package for the measurement and estimation of alpha and beta taxon, phylogenetic and functional diversity. *Methods Ecol Evol* 2015; **6**: 232–236.
41. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* 2017; **8**: 2224.
42. Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can J Microbiol* 2016; **62**: 692–703.
43. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. *Ann Epidemiol* 2016; **26**: 322–329.
44. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* 2017; **6**: 1–20.
45. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 2005; **71**: 8228–35.
46. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2018.
47. Legendre P, Legendre L. Numerical ecology, 3rd ed. 2012. Elsevier.
48. Goslee SC, Urban DL. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *J Stat Softw* 2007; **22**.
49. Pavoine S, Vallet J, Dufour A-B, Gachet S, Daniel H. On the challenge of treating

545 various types of variables: application for improving the measurement of
546 functional diversity. *Oikos* 2009; **118**: 391–402.
547 50. Dray S, Dufour A. The ade4 package: implementing the duality diagram for
548 ecologists. *J Stat Softw* 2007; **22**: 1–20.
549 51. Matesanz S, Gimeno TE, de la Cruz M, Escudero A, Valladares F. Competition may
550 explain the fine-scale spatial patterns and genetic structure of two co-occurring
551 plant congeners. *J Ecol* 2011; **99**: 838–848.
552 52. Borcard D, Gillet F, Legendre P. Numerical Ecology with R. 2011. Springer New
553 York, New York, NY.
554 53. Hagen M, Kissling WD, Rasmussen C, De Aguiar MAM, Brown LE, Carstensen DW,
555 et al. Biodiversity, Species Interactions and Ecological Networks in a Fragmented
556 World. *Advances in Ecological Research*, 1st ed. 2012. Elsevier Ltd., pp 89–210.
557 54. Emer C, Galetti M, Pizo MA, Guimarães PR, Moraes S, Piratelli A, et al. Seed-
558 dispersal interactions in fragmented landscapes – a metanetwork approach. *Ecol*
559 *Lett* 2018; **21**: 484–493.
560 55. Clauset A, Newman MEJ, Moore C. Finding community structure in very large
561 networks. *Phys Rev E* 2004; **70**: 066111.
562 56. Wakita K, Tsurumi T. Finding Community Structure in Mega-scale Social Networks.
563 *WWW*. 2007. pp 1275–1276.
564 57. Marquitti FMD, Guimarães PR, Pires MM, Bittencourt LF. MODULAR: Software for
565 the autonomous computation of modularity in large network sets. *Ecography*
566 *(Cop)* 2014; **37**: 221–224.
567 58. Erdős P, Rényi A. On random graphs I. *Publ Math* 1959; **6**: 290–297.
568 59. Bascompte J, Jordano P, Melian CJ, Olesen JM. The nested assembly of plant-
569 animal mutualistic networks. *Proc Natl Acad Sci* 2003; **100**: 9383–9387.
570 60. Batagelj V, Mrvar A. Pajek - Analysis and Visualization of Large Network. In:
571 Juenger M, Mutzel P (eds). *Graph Drawing Software*. 2003. Springer, Berlin, pp
572 77–103.
573 61. Emerson BC, Gillespie RG. Phylogenetic analysis of community assembly and
574 structure over space and time. *Trends Ecol Evol* 2008; **23**: 619–630.
575 62. Cornell H V., Harrison SP. What Are Species Pools and When Are They Important?
576 *Annu Rev Ecol Evol Syst* 2014; **45**: 45–67.
577 63. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: Networks,
578 competition, and stability. *Science (80-)* 2015; **350**: 663–666.
579 64. Kohl KD, Cary TL, Karasov WH, Dearing MD. Restructuring of the amphibian gut
580 microbiota through metamorphosis. *Environ Microbiol Rep* 2013; **5**: 899–903.
581 65. Chai L, Dong Z, Chen A, Wang H. Changes in intestinal microbiota of *Bufo*
582 *gargarizans* and its association with body weight during metamorphosis. *Arch*
583 *Microbiol* 2018; **0**: 1–13.
584 66. Lyra ML, Bletz MC, Haddad CFB, Vences M. The Intestinal Microbiota of Tadpoles
585 Differs from Those of Syntopic Aquatic Invertebrates. *Microb Ecol* 2018; **76**: 121–

586 124.
587 67. Sullam KE, Essinger SD, Lozupone CA, O'Connor MP, Rose GL, Knight R, et al.
588 Environmental and ecological factors that shape the gut bacterial communities of
589 fish: a meta-analysis. *Mol Ecol* 2012; **21**: 3363–3378.
590 68. Parris DJ, Brooker RM, Morgan MA, Dixson DL, Stewart FJ. Whole gut microbiome
591 composition of damselfish and cardinalfish before and after reef settlement.
592 *PeerJ* 2016; **4**: e2412.
593 69. Madigan MT, Martinko JM, Bender KS, Buckley DH, Stahl DA. Brock Biology of
594 Microorganisms. *International Microbiology* . 2015. Pearson.
595 70. Pryor G. Anaerobic bacteria isolated from the gastrointestinal tracts of bullfrog
596 tadpoles (*Rana catesbeiana*). *Herpetol Conserv Biol* 2008; **3**: 176–181.
597 71. Kueneman JG, Parfrey LW, Woodhams DC, Archer HM, Knight R, McKenzie VJ. The
598 amphibian skin-associated microbiome across species, space and life history
599 stages. *Mol Ecol* 2014; **23**: 1238–1250.
600 72. Hughey MC, Delia J, Belden LK. Diversity and stability of egg-bacterial
601 assemblages: The role of paternal care in the glassfrog *Hyalinobatrachium*
602 *colymbiphyllum*. *Biotropica* 2017; **49**: 792–802.
603 73. Walke JB, Becker MH, Loftus SC, House LL, Cormier G, Jensen R V, et al.
604 Amphibian skin may select for rare environmental microbes. *ISME J* 2014; **8**:
605 2207–2217.
606 74. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, et
607 al. The Long-Term Stability of the Human Gut Microbiota. *Science (80-)* 2013;
608 **341**: 1237439–1237439.
609 75. Britton RA, Young VB. Interaction between the intestinal microbiota and host in
610 *Clostridium difficile* colonization resistance. *Trends Microbiol* 2012; **20**: 313–319.
611 76. Risely A, Waite D, Ujvari B, Klaassen M, Hoye B. Gut microbiota of a long-distance
612 migrant demonstrates resistance against environmental microbe incursions. *Mol*
613 *Ecol* 2017; **26**: 5842–5854.
614 77. Zmora N, Zilberman-Schapira G, Suez J, Mor U, Dori-Bachash M, Bashiardes S, et
615 al. Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is
616 Associated with Unique Host and Microbiome Features. *Cell* 2018; **174**: 1388-
617 1405.e21.
618 78. Fukami T. Historical Contingency in Community Assembly: Integrating Niches,
619 Species Pools, and Priority Effects. *Annu Rev Ecol Evol Syst* 2015; **46**: 1–23.
620 79. Urban MC, De Meester L. Community monopolization: local adaptation enhances
621 priority effects in an evolving metacommunity. *Proc Biol Sci* 2009; **276**: 4129–38.
622 80. Mihaljevic JR. Linking metacommunity theory and symbiont evolutionary ecology.
623 *Trends Ecol Evol* 2012; **27**: 323–9.
624 81. Shade A, Dunn RR, Blowes SA, Keil P, Bohannan BJM, Herrmann M, et al.
625 Macroecology to Unite All Life, Large and Small. *Trends Ecol Evol* 2018; **33**: 731–
626 744.

627

628 **Tables**

629 Table 1. Results of a Permutational Analysis of Variance (perMANOVA) testing if the
630 species of tadpole or the pond of origin are related to the gut microbiome composition
631 of tadpoles using the bacteria as Amplicon Sequence Variants (ASVs). Models were
632 fitted to CODA and PHILR transformed data. The significance of the marginal effects was
633 tested based on 999 permutations. BR = Brazil, CT = Central Texas, ET = East Texas.
634

		PHILR			CODA		
		R^2	F	p	R^2	F	p
BR	Water body$_{(4,80)}$	0.08	3	0.005	0.1	2.7	0.001
	Species$_{(9,80)}$	0.15	2.5	0.002	0.16	2.1	0.001
CT	Water body$_{(5,25)}$	0.48	6	0.001	0.41	3.6	0.001
	Species$_{(1,25)}$	0.05	2.8	0.010	0.04	1.8	0.056
ET	Water body$_{(8,76)}$	0.27	4.7	0.001	0.24	3.2	0.001
	Species$_{(5,76)}$	0.18	5.1	0.001	0.12	2.6	0.001

635

636

Figures

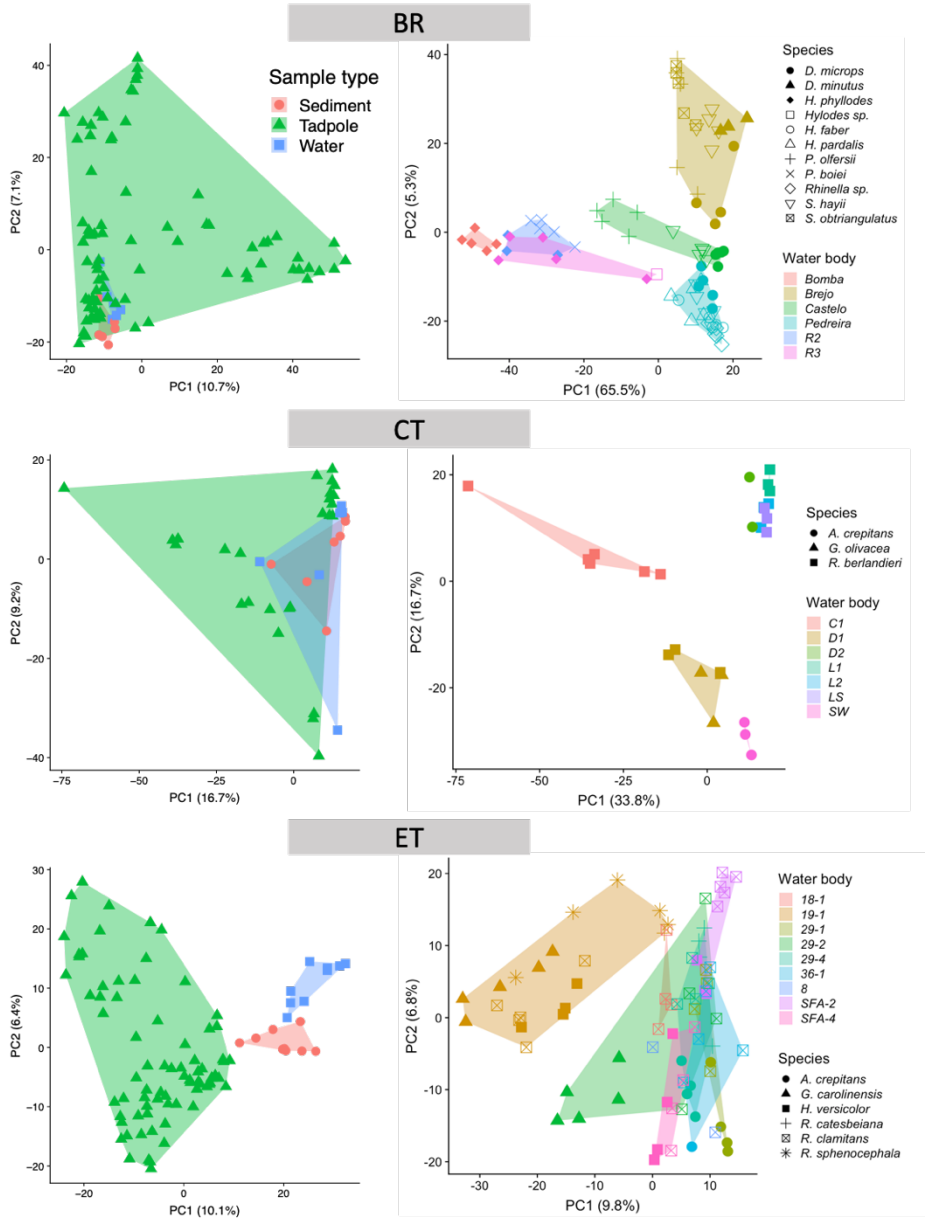

Figure 1. Plot of the first two axes of a Principal Component Analysis on CODA transformed microbiome data (amplicon sequence variants) from tadpole samples from Brazil (BR), Central Texas (CT), and Eastern Texas (ET). Panels on the left show sediment, water, and tadpole samples. Panels on the right show samples from tadpoles only. Left and right panels are results from separate analyses. Shapes represent different species of tadpoles and colors represent water body of origin. See Table S1 for abbreviations.

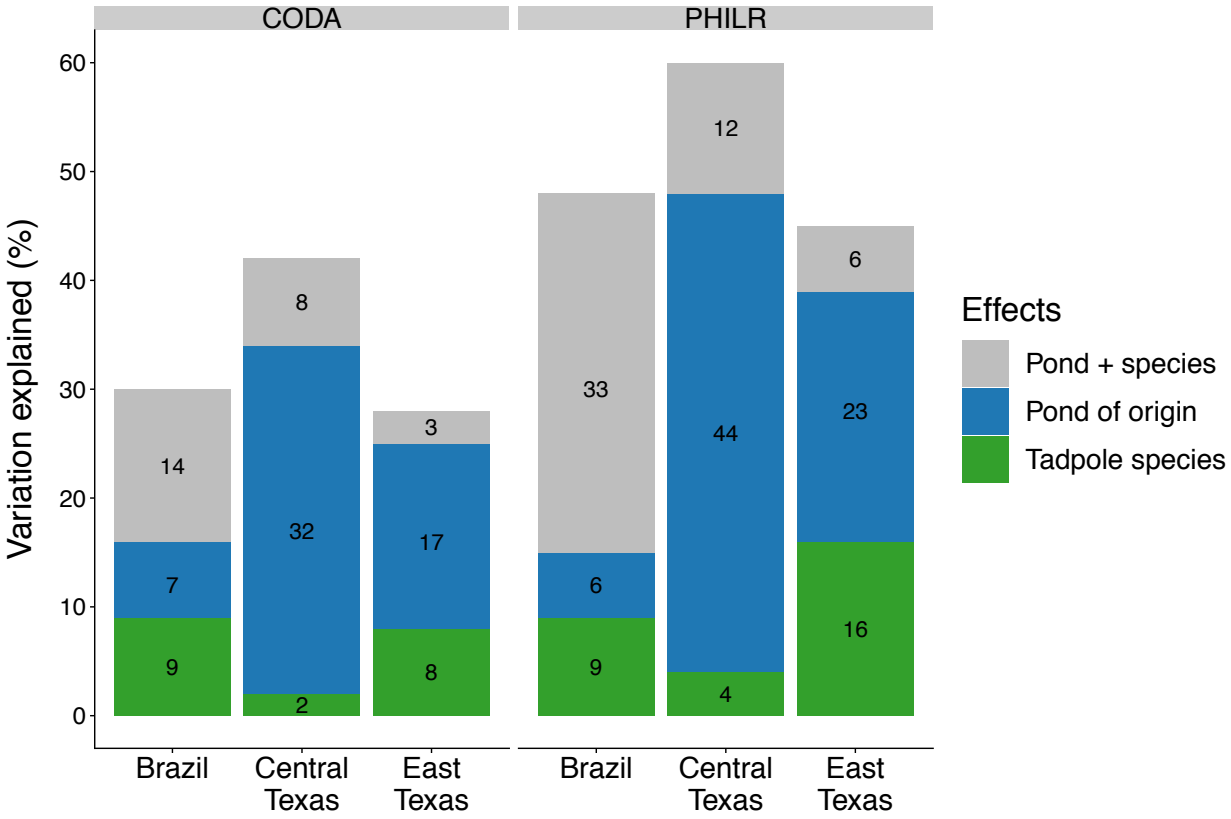

Figure 2. Results of a variation partitioning analysis testing the unique effects of the species of tadpole or pond of origin as well their shared effects on the tadpole gut microbiome considering the bacteria as Amplicon Sequence Variants (ASVs) for CODA- and PHILR-transformed data. The effects of pond of origin and species of tadpole were all significant ($p < 0.05$) based on 1000 permutations, except for the shared component, which cannot be tested.

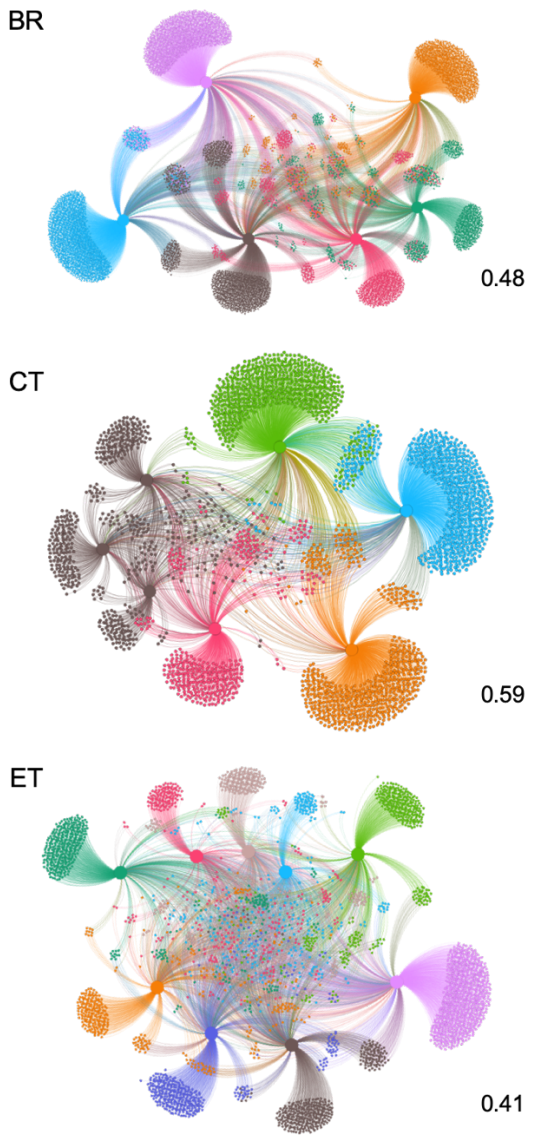

Figure 3. Microbiome metanetworks of the shared bacteria (amplicon sequence variant - ASV) found in the gut of tadpoles. Larger nodes represent each water body for each locality, *i.e.*, Brazil (BR), Central Texas (CT), and East Texas (ET). Smaller nodes represent a unique bacterium that can be found in a single water body (link between smaller and larger nodes) or shared among ponds thus connecting them. Numbers on the bottom-right are the modularity values for each metanetwork. Distinct colors represent different modules which, with a few exceptions, correspond to a single water body (Table S10).

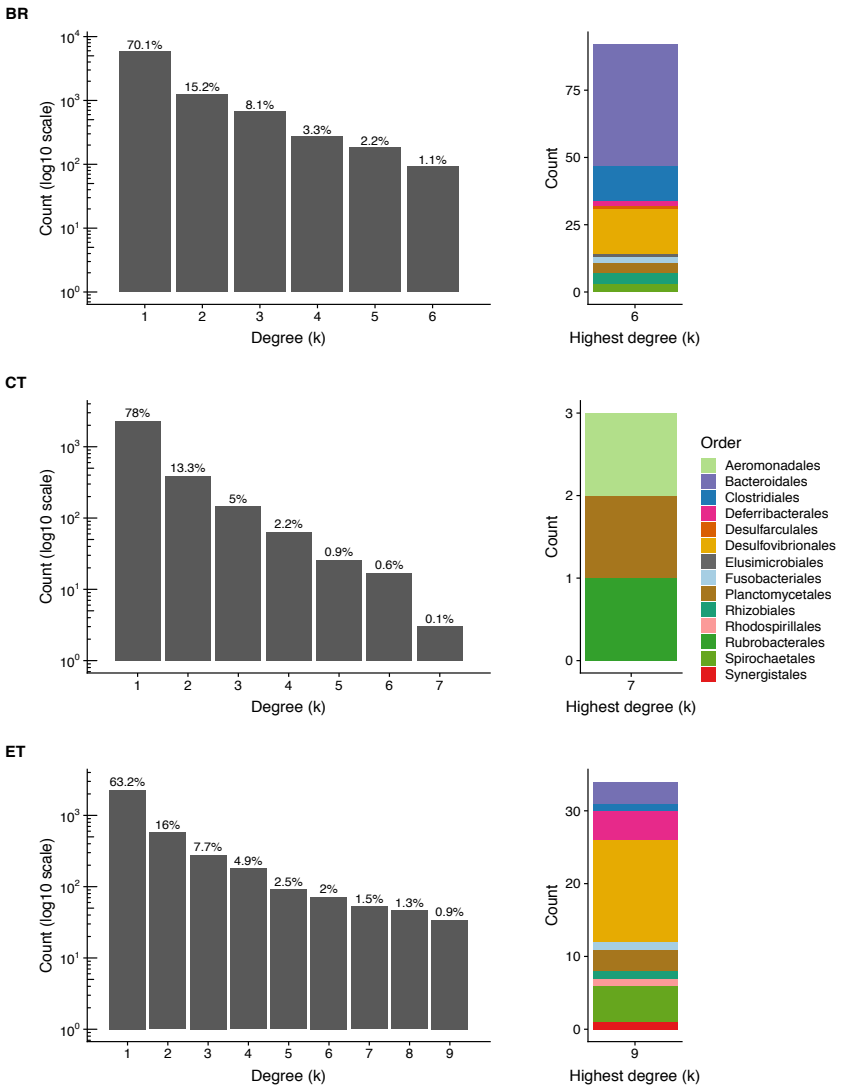

Figure 4. Number of bacteria from the gut of tadpoles for each degree (k) on the left and Order of bacteria present in the highest degree on the right. Each degree represents one pond. Bacteria found in only one pond are represented in the degree category 1. Likewise, bacteria present in tadpoles across all ponds are represented in the highest category of each histogram. The Order of bacteria present in tadpoles across all ponds from each locality is shown on the barplot on the right. Same color across barplots represent the same Order. See Figures S11-S16 for water and sediment metanetworks. BR = Brazil, CT = Central Texas, ET = East Texas.