Integrating population genetic structure, microbiome, and pathogens presence data in *Dermacentor variabilis*

Paula Lado Corresp., 1 , **Bo Luan** 2 , **Michelle EJ Allerdice** 3 , **Christopher D Paddock** 3 , **Sandor E Karpathy** 3 , **Hans Klompen** 1

1 Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, Ohio, United States

2 Statistics, Ohio State University, Columbus, Ohio, United States

3 Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, United States

Corresponding Author: Paula Lado
Email address: ladohenaise.1@osu.edu

Tick-borne diseases (TBDs) continue to emerge and re-emerge in several regions of the world, highlighting the need for novel and effective control strategies. The development of effective strategies requires a better understanding of TBDs ecology, and given the complexity of these systems, interdisciplinary approaches are required. In recent years, the microbiome of vectors has received much attention, mainly because associations between native microbes and pathogens may provide a new promising path towards the disruption of pathogen transmission. However, we still do not fully understand how host genetics and environmental factors interact to shape the microbiome of organisms, or how pathogenic microorganisms affect the microbiome and vice versa. The integration of different lines of evidence may be the key to improve our understanding of TBDs ecology. In that context, we generated microbiome and pathogen presence data for *Dermacentor variabilis*, and integrated those data sets with population genetic data, and metadata for the same individual tick specimens. Clustering and multivariate statistical methods were used to combine, analyze, and visualize data sets. Interpretation of the results is challenging, likely due to the low levels of genetic diversity and the high abundance of a few taxa in the microbiome. *Francisella* was dominant in almost all ticks, regardless of geography or sex. Nevertheless, our results showed that, overall, ticks from different geographic regions differ in their microbiome composition. Additionally, DNA of *Rickettsia rhipicephali, R. montanensis, R. bellii,* and *Anaplasma* spp., was detected in *D. variabilis* specimens. This is the first study that successfully generated microbiome, population genetics, and pathogen presence data from the same individual ticks, and that attempted to combine the different lines of evidence. The approaches and pre-processing steps used can be applied to a variety of taxa, and help better understand ecological processes in biological systems.

1 **Integrating population genetic structure, microbiome, and pathogens presence data in**

2 ***Dermacentor variabilis***

4 Paula Lado[1], Bo Luan[2], Michelle E. J. Allerdice[3], Christopher D. Paddock[3], Sandor E. Karpathy[3],

5 Hans Klompen[1]

7 [1] Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH, USA

8 [2] Statistics, The Ohio State University, Columbus, OH, USA

9 [3] Rickettsial Zoonoses Branch, National Center for Emerging and Zoonotic Infectious Diseases,

10 Centers for Disease Control and Prevention, Atlanta, GA, USA.

12 Corresponding Author:

13 Paula Lado[1]

14 1315 Kinnear Rd, Columbus, OH, 43212, U.S.A

15 Email address: pau.parasito@gmail.com

17 **ABSTRACT**

18 Tick-borne diseases (TBDs) continue to emerge and re-emerge in several regions of the world,

19 highlighting the need for novel and effective control strategies. The development of effective

20 strategies requires a better understanding of TBDs ecology, and given the complexity of these

21 systems, interdisciplinary approaches are required. In recent years, the microbiome of vectors

22 has received much attention, mainly because associations between native microbes and

23 pathogens may provide a new promising path towards the disruption of pathogen transmission.

24 However, we still do not fully understand how host genetics and environmental factors interact

25 to shape the microbiome of organisms, or how pathogenic microorganisms affect the

26 microbiome and vice versa. The integration of different lines of evidence may be the key to

27 improve our understanding of TBDs ecology. In that context, we generated microbiome and

28 pathogen presence data for *Dermacentor variabilis*, and integrated those data sets with

29 population genetic data, and metadata for the same individual tick specimens. Clustering and

30   multivariate statistical methods were used to combine, analyze, and visualize data sets.

31   Interpretation of the results is challenging, likely due to the low levels of genetic diversity and

32   the high abundance of a few taxa in the microbiome. *Francisella* was dominant in almost all

33   ticks, regardless of geography or sex. Nevertheless, our results showed that, overall, ticks from

34   different geographic regions differ in their microbiome composition. Additionally, DNA of

35   *Rickettsia rhipicephali, R. montanensis, R. bellii,* and *Anaplasma* spp., was detected in *D.*

36   *variabilis* specimens. This is the first study that successfully generated microbiome, population

37   genetics, and pathogen presence data from the same individual ticks, and that attempted to

38   combine the different lines of evidence. The approaches and pre-processing steps used can be

39   applied to a variety of taxa, and help better understand ecological processes in biological

40   systems.

41   **INTRODUCTION**

42   As vector-borne diseases continue to emerge and re-emerge in several regions of the

43   world, there is an urgent need for a better understanding of their ecology, including the tripartite

44   pathogen-vector-host relationship, to ultimately develop effective control strategies. Ticks

45   represent an increasing threat to One Health given the range expansion of some species, and the

46   frequent identification of new tick-borne pathogens (Paddock et al. 2016; Eisen et al. 2017). Due

47   to the complex nature of diseases transmitted by ticks, their control requires interdisciplinary

48   studies and collaboration (Murgia et al. 2019).

49   In recent years, the microbiome (referring to bacteria only throughout the manuscript) of

50   vectors (mostly mosquitoes and ticks) has received much attention. The main reason is that

51   associations between native microbes and pathogens may provide a new promising path towards

52   the disruption of pathogen transmission (Narasimhan & Fikrig 2015; Bonnet et al. 2017). In the

53   case of ticks, microbes can be acquired from the environment (e.g through the spiracles),

54   horizontally (from the host, during blood feeding), or vertically (maternally inherited).

55   Relationships between the tick and a particular microbe or set of microbes can range from

56   mutualistic to parasitic (Casadevall et al. 2011; Bonnet et al. 2017). Microbes can affect tick

57   fitness, vector competence, and pathogen transmission (Bonnet et al. 2017; Budachetri et al.

58   2018). For example, a few maternally inherited microbes have been proposed as primary

59   symbionts necessary for tick success by providing vitamins that are lacking in blood (Smith et al.

60   2015b; Duron et al. 2017; Guizzo et al. 2017). Without these microbes, the ticks' fitness is

61   negatively impacted (Zhong et al. 2007). Other symbionts interact with pathogenic

62   microorganisms, either positively or negatively. High proportions of the endosymbiotic

63   *Rickettsia bellii* inhibit the transmission of the pathogenic *Anaplasma marginale* (Gall et al.

64   2016), and *R. parkeri* is likely excluded from *A. maculatum* by Candidatus *Rickettsia andeanae*

65   (Paddock et al. 2015), presumably through a phenomenon known as interference. On the other

66   hand, cases of facilitation have also been reported; such as a positive relationship between the

67   proportion of the microbiome occupied by *Francisella* endosymbiont (FLE) and the infection

68   level of *F. novicida* (Gall et al., 2016).

69   Even though a considerable amount of information and knowledge has been generated

70   during the last decade or so, microbiome research on ticks is still in its infancy. In fact, many

71   results of microbiome research in ticks have led to contradictory findings, highlighting the

72   complexity of these systems, and the caution that needs to be taken when attempting to

73   generalize results (Couper et al. 2019). However, some findings have been consistent. For

74   example, it is now pretty much accepted that tick microbiomes differ among genera, and, quite

75   frequently, among species within a genus (Van Treuren et al. 2015; Rynkiewicz et al. 2015;

76   Chicana et al. 2019). Other factors have also been suggested as shaping the microbiome, such as

77   tick stage, sex, environment, and geography (Williams-Newkirk et al. 2014; Van Treuren et al.

78   2015; Trout Fryxell & DeBruyn 2016), although these results have not been consistent across

79   studies.

80   But we still do not fully understand how host genetics and environmental factors interact

81   to shape the microbiome of organisms (Spor et al. 2011; Goodrich et al. 2014; Steury et al.

82   2019), or how pathogenic microorganisms affect the microbiome and vice versa. Integration of

83   different lines of evidence may be the key to improve our understanding of TBDs ecology. As

84   highlighted by Griffiths et al. (2018), research exploring the links between, for example, vector

85   genetics, microbiome composition and structure, and pathogen susceptibility may enable a better

86   understanding of the factors governing disease in vulnerable populations. Unfortunately, to the

87   best of our knowledge, there are no such studies on ticks. There are a few studies based on

88   vertebrate taxa that have attempted to incorporate genomic and microbial data with data on

89   environmental variation (Fietz et al. 2018; Griffiths et al. 2018; Steury et al. 2019). Overall, these

90   studies have shown that genetically divergent host populations, exhibited more divergent

91   microbiomes (Smith et al. 2015a; Griffiths et al. 2018; Steury et al. 2019). In amphibians,

92   Griffiths et al. (2018) showed that at the landscape level, the microbiome was not structured

93   according to host population structure even though the genetic distance among hosts was still

94   significantly correlated to microbial community dissimilarity when controlling for geographic

95   distance. Steury et al. (2019) investigated populations of Threespine Stickleback and concluded

96   that the microbiome composition was better predicted by fish population genetic divergence than

97   by geographic distance and environment. It is worth noting that the global trends appear to be

98   driven by a subset of the microbiome. In other words, the influence of host genetic factors on the

99   microbiome composition depends on the bacterial taxa in question; host genetics may affect the

100 presence of some microbes, but presence of others may be better explained by, for example, the

101 environment (Fietz et al. 2018). In the same way, differences between populations could be the

102 result of differences in the relative abundance of a small subset of microorganisms (Steury et al.

103 2019).

104   The vertebrate results suggest that integrating microbiome and population genetic data

105 may lead to better understanding of the ecology of tick-borne diseases. In this context, it is

106 paramount to first generate high quality data, and then to develop an appropriated framework to

107 integrate the different lines of evidence. To do so, we generated high quality microbiome data,

108 determined presence/absence data for several pathogens, and combined this with an existing

109 population genetics data set (see Lado et al. 2019). The three pieces of data (microbiome,

110 pathogens, and population genetics) were derived from the same individual ticks. The goal of

111 this study is to integrate the different lines of evidence using clustering methods and multivariate

112 statistics.

113 **METHODS**

114 **Tick samples**

115   The tick specimens employed in this study are the same individual ticks from Lado et al.

116 (2019), with the exception of two specimens for which we did not have enough genetic material.

117 The sample includes 64 adult *D. variabilis* collected from California (n=3), Georgia (n=2),

118 Indiana (n=3), Maryland (n=4), Maine (n=3), Minnesota (n=5), North Dakota (n=1), Ohio

119 (n=31), Oklahoma (n=5), Tennessee (n=5), Virginia (n=1), and Washington (n=1) (Fig. 1). All

120 specimens were wild caught by dragging a 1m x 0.8m cloth thorugh the vegetation, and

121 preserved in 95% ethanol. We used flat (unfed) ticks in an effort to limit the amount of genetic

122   material from the host. All collection data are accessible online through the OSAL (Ohio State

123   Acarology Collection) database (https://acarology.osu.edu/database).

DNA extraction

125   Before DNA extraction, the ticks were surface sterilized following Lado et al. (2018).

126   Briefly, the washes consist on following this procedure twice: 1 minute in commercial bleach

127   (3%), washed in distilled water, and 1 minute in 95% ethanol. Individual DNA extractions were

128   performed using the QIAgen Blood and Tissue kit following the manufacturers' instructions,

129   with one exception: while in ATL buffer a portion of the tick idiosoma was cut with a scalpel

130   and the entire specimen cuticle was recovered to serve as voucher. Cuticles were individually

131   barcoded and data based in the OSAL database (https://acarology.osu.edu/database). OSAL

132   accession numbers are listed in Table 1. Once the DNA was extracted and the cuticles preserved,

133   the DNA was quantified using a Qubit 3.0 fluorometer. DNA was aliquoted and kept in a freezer

134   until needed.

Population genetic data

136   For the integrative portion of the analysis, we employed all genetic data generated

137   previously, and followed the genetic clusters nomenclature from Lado et al. (2019). Details on

138   how the population genetics data set was generated and analyzed can be found elsewhere (Lado

139   et al. 2019). In brief, ticks were assigned to three different genetic clusters according to their

140   population genetic structure: a generally "western cluster", an "eastern cluster", and a "northern

141   cluster". This clusters usually correspond to the locations where ticks were collected.

142

Detection of Rickettsiales through PCR

144   PCR screening

145   Tick DNAs were tested by real-time PCR to detect three genera that include pathogenic

146   microorganisms: *Rickettsia*, *Anaplasma*, and *Ehrlichia.* All ticks were initially screened using

147   two real-time assays: 1) a TaqMan Pan-*Rickettsia* assay, which amplifies a portion of the 23S

148   gene for all *Rickettsia* species using primers PanR8-F and PanR8-R (Kato et al. 2013), and 2) an

149   EvaGreen Anaplasmataceae assay that targets a portion of the 16S gene using primers

150   ECHSYBR-F and ECHSYBR-R, amplifying all *Anaplasma* and *Ehrlichia* species (Li et al.

151   2002).

152   Positive tick samples for Pan-*Rickettsia* (23S real-time assay) where further screened to

153   identify the *Rickettsia* species present. Samples were subjected to conventional semi-nested PCR

154   targeting the *omp*A gene of all spotted fever group *Rickettsia* (Regnery et al. 1991; Eremeeva et

155   al. 2006). PCRs were performed using 1 μM of each primer (Rr190.70, Rr190.602, Rr190.701),

156   10 μL of Taq PCR Master Mix (QIAGEN), 2 μL of sample DNA in the primary reaction or 2 μl

157   of the primary reaction product in the secondary reaction, and water to bring the final reaction

158   volume to 20 μl. Positive samples were processed as described below to identify species present.

159   We also performed a *R. bellii* specific TaqMan assay targeting *glt*A, the citrate synthase gene

160   (Hecht et al. 2016) for all positive samples for Pan-*Rickettsia*.

161   All real-time PCRs were performed in duplicate on a BioRad CFX 96 thermal cycler

162   using 4 μL of template DNA in a final reaction volume of 25 μL for the Pan-*Rickettsia* and 20

163   μL for both the Anaplasmataceae and *R. bellii*-specific assays. We considered samples positive if

164   one of the duplicates had a cycle threshold (Ct) <40. Two sets of negative controls and one set of

165   positive controls were included on each plate, where water was used as the negative non-

166   template control and DNA from cultured *R. rickettsii*, *E. canis*, or an *R. bellii* plasmid were used

167   as positive controls, depending on the assay (Hecht et al. 2019).

168   Amplicon purification and sequencing

169   Amplicons from *omp*A semi-nested PCR were visualized on 1.5% agarose gels using

170   ethidium bromide. Amplicons were extracted and purified using the Promega Wizard SV Gel

171   and PCR Clean-up System (Promega, Madison, WI). Products were bidirectionally sequenced

172   using a BigDye Terminator v3.1 kit on an ABI 3500 genetic analyzer (Applied BioSystems,

173   Carlsbad, CA) and assembled using Geneious version 7.0.4. (http:// geneious.com, Kearse et al.

174   2012). Resulting sequences were compared to available sequences in the NCBI GenBank

175   database using the BLASTn option. Positive amplicons from the Anaplasmataceae assay were

176   also sequenced following the above-mentioned procedure.

177   **Microbiome**

178   Library preparation and sequencing

179   Extracted DNA samples were normalized to a final concentration of 10 ng/µl, and sent to

180   Argonne National Laboratory for library preparation and sequencing. The V4-V5 variable

181   regions of the 16S rDNA were amplified using modified 515F and 926R primers (Walters et al.

182   2016). The amplicons were sequenced on a MiSeq illumina platform, 2x251bp. The resulting

183   reads were initially demultiplexed in MiSeq Reporter.

184   Quality filtering, OTU picking, taxonomic assignments, and diversity calculations

185   After demultiplexing, a primary quality filter of the reads was performed in Cutadapt

186   (Martin 2011) using a cutoff of Q10. If adaptors were still present in the reads after the quality

187   filtering step, they were trimmed in Cutadapt (Martin 2011). After trimming and quality filtering,

188   the reads were assembled in QIIME 1.9.1 (Caporaso et al. 2010). An additional quality filtering

189   step using the *multiple_split_libraries.py* was performed. QIIME 1.9.1 was also used for

190   clustering (97% threshold), open reference OTUs (operational taxonomic units) picking using

191   uclust, and taxonomic assignment using the Greengenes (DeSantis et al. 2006)(DeSantis et al.

192   2006) and Silva132 (Quast et al. 2013) data bases. The aligned representative sequences were

193   used to build a phylogenetic tree using FastTree (Price et al. 2009), and to build OTU tables for

194   different taxonomic levels (i.e genus, family).

195   **Statistical analysis: integrating individual data sets**

196   OTU tables were imported into R, and finer filtering was performed. First, we eliminated

197   all OTUs that appeared in no more than two samples. Then, with the reduced data, we further

198   eliminated OTUs with less than 0.5% relative abundance in all ticks. The dimensions of the data

199   were therefore substantially reduced. Metadata was then added to the tables, including

200   coordinates, sex, genetic cluster they belong to (following Lado et al. 2019), and the presence of:

201   *Rickettsia*, *Anaplasma* or *Ehrlichia* species for each tick specimen. Additionally, ticks were

202   assigned to three geographic regions following Lado et al. 2019: eastern, western, and northern.

203   Clustering approach

204   To determine how samples are related to one another when integrating all variables, a

205   hierarchical clustering approach was taken. In this analysis the variables included were the

206   filtered OTU table, tick ID, sex, genetic cluster, location (as coordinates), and presence of

207   *Ricketts*ia or Anaplasmataceae agents. Hierarchical clustering was performed in R and the

208   number of clusters was set to three (k=3), which is the number of genetic clusters. We employed

209   Gower's distance (Gower 1971) as the measure of dissimilarity between samples because it is

210   compatible with mixed data types (quantitative, nominal, and binary variables). We visualized

211   the clustering results through the dendrogram generated by the hierarchy and demonstrated

212   sample organization in space by using the first two principal components of all the filtered OTUs

213   and location coordinates.

214   As a further exploration, we evaluated how well our clustering matched with the genetic

215   clusters by the concordance index (C-index), which is an internal validation measure of goodness

216   of matching. We also employed methods to determine the optimal number of clusters from the

217   data. This allowed the comparison between the predicted number of clusters (k=3), and the

218   number of clusters determined inherently from the data set. For this purpose, two methods were

219   used: the average silhouette and the elbow method. The former method computes the average

220   silhouette of observations for different values of k. The optimal number of clusters is the one that

221   maximizes the average silhouette over a range of possible values for k (Kaufman & Rousseeuw

222   1990). On the other hand, the elbow method looks at the total within-cluster sum of squares

223   (wss) as a function of the number of clusters. The optimal k will be the smallest one such that

224   adding another cluster doesn't improve much better the total wss.

225   Ordination methods

226   Once the data was integrated, ordination methods were used to visualize the samples in

227   space. We used both non-metric multidimensional scaling (NMDS) and Principal coordinates

228   analyses (PCoA) to determine if there was any patterning within the data. Both techniques

229   reduce the high dimensional data into a two-dimensional representation. All analysis and graphs

230   were done in R, employing the *phyloseq* package (McMurdie & Holmes 2013).

231   Test for significance for different variables

232   We used Welch's t-test to detect if the relative abundance of a specific species of

233   microorganism differed across host sex or between genetic clusters. Such tests were shown to be

234   flexible and robust even when samples sizes were unbalanced and group variances were unequal

235   (Delacre et al. 2017).

236   To compare microbial beta diversity across treatments (sex, genetic cluster, and region)

237   we calculated the distance between microbial communities by using two metrics. The first one,

238   Bray-Curtis distance, considers the relative abundance of bacteria, while the second, Jaccard's

239   distance (or dissimilarity index), measures differences in presence/absence of Bacteria. All

240   distances were calculated using the *vegan* package (Oksanen et al. 2019) in R. We then

241   partitioned the matrices by treatment and performed permutational multivariate analysis of

242   variance (PERMANOVA) (Anderson 2001) to determine if beta diversity differed between

243   treatments. PERMANOVA calculations were performed using the *adonis* function in the vegan

244   package with 999 permutations. For those treatments that included more than two groups, a post

245   hoc test was performed to identify the pairs of groups between which the bulk of differences

246   occurred. For that purpose, we employed the *paiwise.adonis* function from the same package.

247

248   **RESULTS**

249   **Detection of Rickettsiales through PCR**

250   DNA belonging to the genus *Rickettsia* was detected in 15.6% (10/64) of the screened

251   ticks including *R. montanensis* (n=6, 9.3%); *R. rhipicephali* (n=2, 3.1%); and *R. bellii* (n=2,

252   3.1%). *Rickettsia bellii* was detected only in specimens collected from the western U.S.

253   *Anaplasma* DNA was found in a single tick (1.6%), collected in Oklahoma. The DNA sequence

254   obtained was identified as "Uncultured *Anaplasma* sp. clone 15-3642 16S ribosomal RNA gene,

255   GenBank number MG429812". The sequence generated in this study was 100% identical to

256   MG429812, and the query coverage was also 100%, corresponding to bases 23 to 168.

257   None of the ticks were infected with *Ehrlichia* species. A summary of PCR results and metadata

258   for each tick specimen is presented in Table 2.

259   **Microbiome**

260   Microbiome data was successfully generated for all ticks. The number of Illumina

261   forward (R1) reads after quality filtering was 2,129,331 with and average length of 251 bp. After

262   OTU picking and taxonomic assignment many of the reads were categorized as "unassigned"

263   and therefore, the average number of sequences per tick specimen included in downstream

264   analyses was fairly low, ranging from a few hundred to ~15,500.

265   A total of 41 orders of bacteria were identified (13 remaining after filtering), and three of them

266   were dominant: Rickettsiales, Legionellales, and Enterobacteriales (Fig. 2A). At the genus level,

267   three taxa were highly abundant at the genus level: *Rickettsia* spp., *Francisella* spp., and

268   Enterobacteriaceae "other". (Fig. 2B). *Francisella spp.* were present in all ticks, with a relative

269   abundance ranging from 0.5 to 100% (mean = 80.4%, median = 99.1%). *Rickettsia* was detected

270   in 10 (15.6%) of the ticks. In all cases where *Rickettsia* spp. had a relative abundance of at least

271   0.5% by microbiome analysis, tick specimens were also PCR positive, thus microbiome and PCR

272   results are overall, congruent (Table 2). *Rickettsiella*, an intracellular bacterial pathogen of

273   arthropods (Leclerque 2008), was found in a tick from Ohio (1/64; tick ID 119572D) with a

274   relative abundance of 11%.

275   **Statistical analysis: integrating individual data sets**

276   Clustering approach

277   The dendrogram resulting from hierarchical clustering showed that the two most distinct

278   samples (clustering together) were from the same location and collection event: Fernald

279   Preserve, Hamilton Co., Ohio (110559-1 and 110559-2) (Fig. 3). The second clade grouped two

280   samples from CA, the two samples collected in Lake Co (115102A,B). The next pair of samples

281   corresponded to the two remaining samples from the west coast, a sample from Napa Co, CA

282   (115101C); and one from Whitman Co, WA (115105B). These two samples were closely

283 associated with a clade that included all Northern samples. Lastly, all Eastern samples clustered

284 together, without any internal grouping according to geography (Fig. 3).

285 No differences were evident upon visualization of community membership or structure. PCA did

286 not show any specific clustering pattern between samples (Supplementary file 1). In fact, most of

287 the ticks were clustered in a single aggregate. The ticks that clustered somewhat separated were

288 firstly the two that also appeared separated in the clustering analysis (110559-1 and 110559-2);

289 and a tick from Ohio, 119241-3. It is important to note that the first two principal components

290 explained only approximately 40% of the variance present in our sample.

291 As for k, the optimal number of clusters, silhouette method and elbow method suggested

292 $k = 2$ and $k = 4$ respectively (Supplementary file 2). Both values were close to the number of

293 genetic clusters, which was also the number of clusters we used in the analysis. The C-index

294 between the predicted clusters and genetic clusters was 0.61, indicating some weak but

295 nonnegligible consistency between these two pairs.

296 Ordination methods

297 When visualizing the samples in the space through NMDS, the samples were spread out

298 without any specific pattern (Fig. 4A). In PCoA with Gower's distance as metric, most samples

299 appeared in a single tight aggregate, while a few others dispersed randomly along the axes (Fig.

300 4B).

301 Test for significance for different variables

302 The number of bacterial genera found in ticks belonging to the Eastern genetic cluster

303 (24) was higher than that of the Western (16), and Northern (10) clusters. The genus *Francisella*

304 was dominant in almost all the ticks, regardless of geography. Its relative abundance did not

305 differ significantly between males and females (p-val 0.8); nor between ticks belonging to

306   different genetic clusters (Eastern vs Western p-val 0.57, Eastern vs Northern p-val 0.96,

307   Northern vs Western p-val 0.60) (Fig. 5). The relative abundance of this genus was also not

308   significantly different between infected and uninfected ticks (p-val 0.14), although it was

309   generally lower in infected ticks. Infected ticks were defined as those that were PCR positive for

310   either *Anaplasma* or *Rickettsia* species.

311   *Rickettsia* relative abundance did not significantly differ between males and females (p-

312   val 0.44), nor between ticks belonging to different genetic clusters (Northern vs. Eastern p-val =

313   0.26; Northern vs. Western p-val = 0.72; and Eastern vs. Western p-val = 0.45) (Fig. 5).

314   *Rickettsia rhipicephali* and *Anaplasma* spp. were present only in ticks belonging to the Eastern

315   genetic cluster, and *R. montanensis* was present in ticks belonging to the Northern and Eastern

316   clusters. *Rickettsia bellii* was found infecting ticks from the Western and Northern genetic

317   clusters, although of these ticks were collected in the Western region (Fig 6). One of the ticks

318   positive for this species was 115101C, a tick collected in California but with the genetic make-up

319   corresponding to Northern ticks.

320   Overall, microbial beta diversity differed between regions (p-val <0.05), but not between

321   sexes (p = 0.63) when using Bray-Curtis distance. Post hoc tests indicated that ticks collected in

322   the Western region harbored a different microbiome from those collected in the Eastern or

323   Northern regions (p < 0.05). In the case of presence/absence of Bacteria (Jaccard distance), the

324   results were the same. A significant difference between ticks collected from different regions

325   was detected (p< 0.05), and the post hoc tests showed that the differences were significant

326   between the Western and both the Eastern and Northern regions.

327   The Adonis test for genetic clusters showed that there were differences (p < 0.05) only

328   when considering relative abundance of bacterial taxa (Bray-Curtis distance). When the pairwise

329 Adonis test was performed to determine what genetic clusters differed in microbiome structure

330 (p <0.05) between Eastern and Western genetic clusters, but its correction was not (plFDR

331 0.096). Eastern and Western clades are genetically, the two more distant (Lado et al. 2019).

332

333 **DISCUSSION**

334   This work corresponds, to the best of our knowledge, to the first attempt in tick research

335 to integrate population genetic, microbiome, and pathogens presence data to better understand

336 the ecology of TBDs. To optimize comparisons, it was our goal to generate those three pieces of

337 data for each of the individual ticks included in the data set. We show that this is possible, even

338 when techniques based on next generation sequencing require high quantities of high quality

339 DNA.

340   We developed an effective pre-processing and processing procedure for researchers

341 interested in microbiomes of small organisms, or parts of organisms with limited amounts of

342 genetic material. This was motivated by a need for consensus regarding the pre-processing of

343 ticks for generating microbiome data. For example, it is likely that many of the incongruencies

344 between microbiome studies arise from differences in approaches to decontamination of the

345 ticks' surface. Several studies in the literature performed surface sterilization of the ticks before

346 DNA extraction (e.g. Lado et al. 2018; Trout Fryxell and DeBruyn 2016), whereas others did not

347 (e.g. Clow et al. 2018); and the differences in bacterial communities may be a reflection of

348 extensive environmental "contamination" in the latter. Even between studies that "washed" the

349 ticks, comparisons need to be done cautiously, since not all sterilization methods are equally

350 effective. The incorporation of bleach appears necessary for effective decontamination (Binetruy

351 et al. 2019). Increased consistency in results of microbiome studies of ticks requires some

352   community consensus related to the generation and analysis of microbiome data. This is

353   especially important for non-model organisms such as ticks, given the "noise" (Alpha and Beta

354   diversity inflation) generated by the presence of environmental microorganisms.

355   Consistent with most other studies of tick microbiomes (Van Treuren et al. 2015; Gall et al.

356   2016; Chicana et al. 2019), our analysis showed that the microbiome of many species is heavily

357   dominated by a few genera, while the microbiomes of others, such as *Ixodes angustus* (Chicana

358   et al. 2019) and *Haemaphysalis lemuris* (Lado et al. 2018), are more diverse. The microbiome of

359   *D. variabilis* is dominated by *Francisella* (Chicana et al. 2019; Travanty et al. 2019; current

360   study). Clow et al. (2018) presented, to some extent, opposing results (far higher microbial

361   diversity), but they analyzed only 9 specimens, and the ticks were not surface sterilized. These

362   results casts some doubt on the concept of a "core microbiome". For example, Chicana et al.

363   (2019) report a *D. variabilis* core microbiome as composed by *Francisella*, *Sphingomonas*, and

364   *Methylobacterium*. On the other hand, Travanty et al. (2019) reported a *D. variabilis* core

365   microbiome that includes *Francisella* spp., *Sphingomonas* spp., *Delftia* spp.,

366   and *Hymenobacter* spp. Our study recovered all of these taxa but only *Fransicella* spp. Is nearly

367   universally present. *Francisella* has been well established as endosymbiont and dominant in *D.*

368   *variabilis*, as well as other *Dermacentor* species. All other taxa reported as "core microbiome"

369   are not consistent across studies (e.g. Chicana et al. 2019; Clow et al. 2018; Rynkiewicz et al.

370   2015; Travanty et al. 2019). Thus, the true "core" of taxa overlapping across studies, can be

371   reduced to *Francisella* only. The concept of "core microbiome" is generally used in microbiome

372   research to refer to a suite of microbes, and not to refer to only one taxon. The utility of the

373   concept of "core microbiome" for *D. variabilis* is therefore unclear.

374   Recent studies focusing on the pathogen transmission by *D. variabis* and on its role as

375   vector of human diseases have consistently reported low prevalence of infections with known

376   pathogenic microorganisms. For example, *R. rickettsii*, is usually present in ~1% of the ticks

377   analyzed e.g. (Hecht et al. 2019). The results obtained in the present study are consistent with

378   these literature results, and show an overall *Rickettsia* spp. prevalence of 17.2%, including *R.*

379   *montanensis, R. bellii,* and *R. amblyommatis* infections. *Rickettsia rickettsii* DNA was not found

380   in any of the ticks analyzed. We detected DNA of *Anaplasma* spp. in one tick sample collected

381   in Oklahoma. Its DNA sequence matched 100% with GenBank sequences submitted by

382   researchers at the CDC Fort Collins, who isolated the agent from a human blood sample. No

383   additional information is available in the literature about this case. This study comprises the first

384   report of that specific bacterial agent in *D. variabilis*. However, an *Anaplasma* spp. *bovis*-like

385   agent has been previously reported in *D. andersoni* ticks from Canada (Dergousoff & Chilton

386   2011). Unfortunately, the region of 16S amplified differs between our study and that of *D.*

387   *andersoni* (Dergousoff and Chilton, 2011), making it difficult to determine if the sequences

388   correspond to the same specific agent. Regardless, it appears that uncharacterized *Anaplasma*

389   agents are circulating in nature (ticks and humans), and their characterization, together with the

390   determination of whether or not they are capable of causing disease in humans should be further

391   explored. Furthermore, due to the finding of this agent's DNA in a *D. variabilis* sample in this

392   study, the role of *D. variabilis* as potential vector should be considered and further explored.

393   With respect to the primary goal of this study, the integration of the different types of

394   data was done successfully, although the interpretation of the results is somewhat challenging.

395   The latter problem is most likely caused by to the low levels of genetic diversity and moderate

396   levels of population structure (see details in Lado et al. 2019), and a microbiome highly

397   dominated by a few taxa. As a result, ordination methods failed to show patterns of variation

398   across different groups. Nonetheless, the dendrogram resulting from the clustering analysis, in

399   which all lines of evidence and geographic location were considered, was largely congruent with

400   the observations at the population genetic level. Thus, the addition of microbiome, pathogen

401   presence, host sex, and geographic location led to conclusions that are, in general, consistent

402   with host genetics. All Eastern samples clustered together (with the exception of two samples

403   from Hamilton Co, OH), and separated from both Western and Northern samples. The

404   distinctiveness and separation of those two samples from Hamilton Co, OH reflects their

405   distinctiveness at the microbiome level (Fig. 2). The fact that both samples appeared very

406   similar to all other Eastern samples at the population genetic level (Lado et al. 2019), supports

407   that hypothesis. It is possible that these two specimens were not well surface sterilized, or that

408   the environment at that collection location is different.

409   The higher number of bacterial genera found in ticks belonging to the Eastern genetic

410   cluster when compared to either Northern or Western ticks, is likely a product of the higher

411   number of ticks analyzed, and the greater geographic area covered. Even though in the broader

412   sense the microbiome is not diverse (dominated by three genera), it differed between the three

413   main geographic regions: Eastern, Western, and Northern. Statistical results looking at the beta

414   diversity between genetic clusters are less clear: initial test demonstrated a difference in the

415   microbial communities between genetic clusters, but pairwise tests with their corresponding

416   corrections failed to reach the same result. These findings should be further explored using more

417   samples, and more samples per genetic cluster. From our analyses, geography explains the

418   differences in the microbial communities better than host genetics. This could be the result of

419   certain microorganisms occurring only in certain geographic areas.

420   Despite challenges arising during the interpretation of the results, the integration of lines

421 of evidence and metadata, revealed that Eastern ticks canbe separated from Norther and Western

422 ticks. It can also be noted that some *Rickettsia* species were associated with a certain geographic

423 area. For example, *R. bellii* was found only in samples collected in CA, a finding consistent with

424 that reported in Hecht et. al. (2019). It seems possible that *R. bellii* is more common along the

425 west coast. And while the two *R. bellii* positive tick samples were collected in CA, one of them

426 (ID 115101C) did not belong to the Western genetic cluster; it belonged to the Northern cluster.

427 From the host genetics perspective, this tick is more similar to ticks from Northern locations;

428 whereas from the rickettsial agents perspective, it is more similar to Western samples. Looking at

429 the microbiome composition, this tick could belong to any geographic region. This underscores

430 the value of integrating different types of data when thinking about disease ecology.

431   The approaches taken during this study, both to generate and analyze data, can be applied

432 to a wide variety of taxa, and groups of organisms. It is likely that in organisms with a higher

433 level of variation, the clustering methods will be more informative, and their interpretation more

434 straightforward. In the particular case of *D. variabilis,* given the low levels of variation (both in

435 the microbiome and population genetics), together with the low prevalence of pathogenic

436 microorganisms, it may be valuable to attempt different approaches to assess risk of infection in

437 different geographic areas.

438

439 **CONCLUSIONS**

440   In conclusion, this is the first study that successfully generated microbiome, population

441 genetics, and pathogens presence data for the same individual ticks. General methodologies and

442 pre-processing steps are replicable, and applicable to different groups of organisms across the

443   tree of life. This work also comprises, at present, one of the few studies aiming at integrating

444   population genetics and microbiome data to better understand ecological processes and disease,

445   and it is the first one to do so for ticks. The integration of different lines of evidence allows a

446   more holistic approach; and clustering and ordination methods are very helpful to summarize and

447   visualize the results. Finally, this study comprises the first report of "Uncultured *Anaplasma* sp.

448   clone 15-3642" in *D. variabilis*. This agent was previously isolated from human blood, and it is

449   important to determine if it is capable of causing disease in humans, and if so, the role of

450   *Dermacentor* ticks as potential vectors.

451

452   **ACKNOWLEDGEMENTS**

453   We are grateful to M. Yoshimizu, A. Donohue, B. Ryan, G. Keeney, C. Nelson, L. Beati,

454   L. Durden, C. Lubekzyk, D. Neitzel, G. Hickling, and B. Pagac who provided some of the tick

455   specimens used in this study. We also thank The National Ecological Observatory Network, a

456   program sponsored by the National Science Foundation and operated under cooperative

457   agreement by Battelle Memorial Institute, for providing some specimens used in this work. We

458   also thank the Ohio Supercomputer Center for computing resources (allocation grant PAS1306).

459

460   **REFERENCES**

461   Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral

462   Ecology **26**:32–46.

463   Binetruy F, Dupraz M, Buysse M, Duron O. 2019. Surface sterilization methods impact

464   measures of internal microbial diversity in ticks. Parasites and Vectors **12**(268).

465   Bonnet SI, Binetruy F, Hernández-Jarguín AM, Duron O. 2017. The Tick Microbiome: why non-

466   pathogenic microorganisms matter in tick biology and pathogen transmission. Frontiers in

467   Cellular and Infection Microbiology **7**:1–14. Available from

468   http://journal.frontiersin.org/article/10.3389/fcimb.2017.00236/full.

469 Budachetri K, Kumar D, Crispell G, Beck C, Dasch G, Karim S. 2018. The tick endosymbiont

470   Candidatus *Midichloria mitochondrii* and selenoproteins are essential for the growth of

471   Rickettsia parkeri in the Gulf Coast tick vector. Microbiome **6**:141.

472 Caporaso JG et al. 2010. QIIME allows analysis of high-throughput community sequencing data.

473   Nature Methods **7**:335–336. Nature Publishing Group. Available from

474   http://www.nature.com/doifinder/10.1038/nmeth.f.303 (accessed November 6, 2016).

475 Casadevall A, Fang FC, Pirofski L anne. 2011. Microbial virulence as an emergent property:

476   Consequences and opportunities. PLoS Pathogens **7**:e1002136.

477 Chicana B, Couper LI, Kwan JY, Tahiraj E, Swei A. 2019. Comparative microbiome profiles of

478   sympatric tick species from the far-western United States. Insects **10**:1–12.

479 Clow KM, Weese JS, Rousseau J, Jardine CM. 2018. Microbiota of field-collected *Ixodes*

480   *scapularis* and *Dermacentor variabilis* from eastern and southern Ontario, Canada. Ticks

481   and Tick-borne Diseases **9**:235–244.

482 Couper LI, Kwan JY, Ma J, Swei A. 2019. Drivers and patterns of microbial community

483   assembly in a Lyme disease vector. Ecology and Evolution **9**:7768–7779.

484 Delacre M, Lakens D, Leys C. 2017. Why psychologists should by default use welch's t-Test

485   instead of student's t-Test. International Review of Social Psychology **30**:92–101.

486 Dergousoff SJ, Chilton NB. 2011. Novel genotypes of *Anaplasma bovis*, "Candidatus

487   Midichloria" sp. and Ignatzschineria sp. in the Rocky Mountain wood tick, *Dermacentor*

488   *andersoni*. Veterinary Microbiology **12**;150(1-2):100-106.

489 DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P,

490   Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and

491   workbench compatible with ARB. Applied and environmental microbiology **72**:5069–72.

492   American Society for Microbiology (ASM). Available from

493   http://www.ncbi.nlm.nih.gov/pubmed/16820507 (accessed November 6, 2016).

494 Duron O et al. 2017. Evolutionary changes in symbiont community structure in ticks. Molecular

495   Ecology **26**:2905–2921.

496 Eisen RJ, Kugeler KJ, Eisen L, Beard CB, Paddock CD. 2017. Tick-borne zoonoses in the

497   United States: Persistent and emerging threats to human health. ILAR Journal **58**:319–335.

498 Eremeeva ME, Bosserman EA, Demma LJ, Zambrano ML, Blau DM, Dasch GA. 2006. Isolation

499   and identification of *Rickettsia massiliae* from *Rhipicephalus sanguineus* ticks collected in

500   Arizona. Applied and Environmental Microbiology **72**:5569–5577.

501 Fietz K, Rye Hintze CO, Skovrind M, Kjærgaard Nielsen T, Limborg MT, Krag MA, Palsbøll

502   PJ, Hestbjerg Hansen L, Rask Møller P, Gilbert MTP. 2018. Mind the gut: Genomic

503   insights to population divergence and gut microbial composition of two marine keystone

504   species. Microbiome **6**:82.

505 Gall CA, Reif KE, Scoles GA, Mason KL, Mousel M, Noh SM, Brayton KA. 2016. The bacterial

506   microbiome of *Dermacentor andersoni* ticks influences pathogen susceptibility. The ISME

507   Journal **10**:1846–1855. Nature Publishing Group. Available from

508   http://www.nature.com/doifinder/10.1038/ismej.2015.266 (accessed September 20, 2016).

509 Goodrich JK et al. 2014. Human genetics shape the gut microbiome. Cell **159**(4):789-799.

510 Gower JC. 1971. A General Coefficient of Similarity and Some of Its Properties. Biometrics

511   **27**:857–871.

512   Griffiths SM, Harrison XA, Weldon C, Wood MD, Pretorius A, Hopkins K, Fox G, Preziosi RF,

513   Antwis RE. 2018. Genetic variability and ontogeny predict microbiome structure in a

514   disease-challenged montane amphibian. ISME Journal **12**:2506–2517.

515   Guizzo MG et al. 2017. A *Coxiella* mutualist symbiont is essential to the development of

516   *Rhipicephalus microplus*. Scientific Reports **7**:17554.

517   Hecht JA et al. 2019. Multistate Survey of American Dog Ticks (*Dermacentor variabilis*) for

518   *Rickettsia* Species. Vector-Borne and Zoonotic Diseases **19**:652–657.

519   Hecht JA, Allerdice MEJ, Krawczak FS, Labruna MB, Paddock CD, Karpathy SE. 2016.

520   Development of a *Rickettsia bellii*-specific taq man assay targeting the citrate synthase

521   gene. Journal of Medical Entomology **53**(6):1492-1495.

522   Kato CY, Chung IH, Robinson LK, Austin AL, Dasch GA, Massunga RF. 2013. Assessment of

523   real-time PCR assay for detection of *Rickettsia* spp. and *Rickettsia rickettsii* in banked

524   clinical samples. Journal of Clinical Microbiology **51**:314–417.

525   Kaufman L, Rousseeuw P. 1990. Finding Groups in Data: An Introduction to Cluster Analysis.

526   John Wiley & Sons, Inc.

527   Kearse M et al. 2012. Geneious Basic: An integrated and extendable desktop software platform

528   for the organization and analysis of sequence data. Bioinformatics **28**:1647–1649.

529   Lado P, Cox C, Wideman K, Hernandez A, Klompen H. 2019. Population genetics of

530   *Dermacentor variabilis* say 1821 (Ixodida: Ixodidae) in the United States Inferred from

531   ddRAD-seq SNP Markers. Annals of the Entomological Society of America **112**:433–442.

532   Lado P, Qurollo B, Williams C, Junge R, Klompen H. 2018. The microbiome of *Haemaphysalis*

533   *lemuris* (Acari: Ixodidae), a possible vector of pathogens of endangered lemur species in

534   Madagascar. Ticks and Tick-borne Diseases **9**:1252–1260. Available from

535   https://doi.org/10.1016/j.ttbdis.2018.05.003.

536   Leclerque A. 2008. Whole genome-based assessment of the taxonomic position of the arthropod

537   pathogenic bacterium *Rickettsiella grylli*. FEMS Microbiology Letters **283**(1):117-127.

538   Li JS, Chu F, Reilly A, Winslow GM. 2002. Antibodies Highly Effective in SCID Mice During

539   Infection by the Intracellular Bacterium *Ehrlichia chaffeensis* Are of Picomolar Affinity and

540   Exhibit Preferential Epitope and Isotype Utilization. The Journal of Immunology **169**:1419–

541   1425.

542   Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.

543   EMBnet.journal **17**:10. Available from

544   http://journal.embnet.org/index.php/embnetjournal/article/view/200.

545   McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis

546   and Graphics of Microbiome Census Data. PLoS ONE **8**.

547   Murgia M V. et al. 2019. Meeting the challenge of tick-borne disease control: A proposal for

548   1000 *Ixodes* genomes. Ticks and Tick-borne Diseases **10**:213–218.

549   Narasimhan S, Fikrig E. 2015. Tick microbiome: the force within. Trends in parasitology

550   **31**:315–23. Available from http://www.ncbi.nlm.nih.gov/pubmed/25936226 (accessed

551   September 20, 2016).

552   Oksanen J et al. 2019. vegan: Community Ecology Package. R package version 2.5-2.

553   Paddock CD et al. 2015. High prevalence of "Candidatus *Rickettsia andeanae*"; and apparent

554   exclusion of *Rickettsia parkeri* in adult *Amblyomma maculatum* (Acari: Ixodidae) from

555   Kansas and Oklahoma. Ticks and tick-borne diseases **6**:297–302.

556   Paddock CD, Lane RS, Staples JE, Labruna MB. 2016. Changing paradigms for tick-borne

557   diseases in the Americas. Page Global Health Impacts of Vector-Borne Diseases: Workshop

558   Summary.

559   Price MN, Dehal PS, Arkin AP. 2009. Fasttree: Computing large minimum evolution trees with

560   profiles instead of a distance matrix. Molecular Biology and Evolution **26**:1641–1650.

561   Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The

562   SILVA ribosomal RNA gene database project: Improved data processing and web-based

563   tools. Nucleic Acids Research.

564   Regnery RL, Spruill CL, Plikaytis BD. 1991. Genotypic identification of rickettsiae and

565   estimation of intraspecies sequence divergence for portions of two rickettsial genes. Journal

566   of Bacteriology **173**:1576–1589.

567   Rynkiewicz EC, Hemmerich C, Rusch DB, Fuqua C, Clay K. 2015. Concordance of bacterial

568   communities of two tick species and blood of their shared rodent host. Molecular ecology

569   **24**:2566–79. Available from http://www.ncbi.nlm.nih.gov/pubmed/25847197 (accessed

570   September 20, 2016).

571   Smith CCR, Snowberg LK, Gregory Caporaso J, Knight R, Bolnick DI. 2015a. Dietary input of

572   microbes and host genetic variation shape among-population differences in stickleback gut

573   microbiota. ISME Journal **9**:2515–2526.

574   Smith TA, Driscoll T, Gillespie JJ, Raghavan R. 2015b. A Coxiella-like endosymbiontis a

575   potential vitamin source for the lone star tick. Genome Biology and Evolution **7**.

576   Spor A, Koren O, Ley R. 2011. Unravelling the effects of the environment and host genotype on

577   the gut microbiome. Nature Reviews Microbiology **9**(4):279-290.

578   Steury RA, Currey MC, Cresko WA, Bohannan BJM. 2019. Population genetic divergence and

579   environment influence the gut microbiome in oregon threespine stickleback. Genes

580   **10**(7):484.

581   Travanty N V., Ponnusamy L, Kakumanu ML, Nicholson WL, Apperson CS. 2019. Diversity

582   and structure of the bacterial microbiome of the American dog tick, *Dermacentor variabilis*,

583   is dominated by the endosymbiont *Francisella*. Symbiosis **79**:239-250.

584   Trout Fryxell RT, DeBruyn JM. 2016. The microbiome of *Ehrlichia*-infected and uninfected

585   lone star ticks (*Amblyomma americanum*). PLoS ONE **11**:1–19.

586   Van Treuren W et al. 2015. Variation in the Microbiota of *Ixodes* Ticks with Regard to

587   Geography, Species, and Sex **81**:6200–6209.

588   Walters W et al. 2016. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal

589   Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. mSystems

590   **1**(1):e0009-15.

591   Williams-Newkirk AJ, Rowe LA, Mixson-Hayden TR, Dasch GA. 2014. Characterization of the

592   bacterial communities of life stages of free living lone star ticks (*Amblyomma americanum*).

593   PLoS ONE **9**.

594   Zhong J, Jasinskas A, Barbour AG. 2007. Antibiotic treatment of the tick vector *Amblyomma*

595   *americanum* reduced reproductive fitness. PLoS ONE 2(5):e405.

General information of *Dermacentor variabilis* samples used in this study. All individuals are from the USA. Each row corresponds to an individual tick specimen and the columns to the collection information.

1 **Table 1.** General information of *Dermacentor variabilis* samples used in this study. All individuals are from the USA. Each row
2 corresponds to an individual tick specimen and the columns to the collection information.

RAW DATA ID	VOUCHER ID	COLLECTION EVENT ID	US STATE	LOCALITY	COORDINATES
119592-6	OSAL119943	OSAL119592	IN	Hoosier National Forest	38.52, -86.44
119592-7	OSAL119944	OSAL119592	IN	Hoosier National Forest	38.52, -86.44
119273-2	OSAL119553	OSAL119273	IN	Splinter Ridge Wildlife area	38.75, -85.20
119241-2	OSAL119260	OSAL119241	OH	Battelle Darby Creek MP	39.9, -83.21
119241-3	OSAL119261	OSAL119241	OH	Battelle Darby Creek MP	39.9, -83.21
119241-4	OSAL119262	OSAL119241	OH	Battelle Darby Creek MP	39.9, -83.21
119243-1	OSAL119267	OSAL119243	OH	Glacier Ridge MP	40.13, -83.18
119243-2	OSAL119268	OSAL119243	OH	Glacier Ridge MP	40.13, -83.18
119243-3	OSAL119269	OSAL119243	OH	Glacier Ridge MP	40.13, -83.18
119243-4	OSAL119270	OSAL119243	OH	Glacier Ridge MP	40.13, -83.18
119244-1	OSAL119400	OSAL119244	OH	High banks MP	40.15, -83.03
119244-4	OSAL119403	OSAL119244	OH	High banks MP	40.15, -83.03
119247-1	OSAL119404	OSAL119247	OH	Pickerington Ponds MP	39.88, -82.79
119247-4	OSAL119407	OSAL119247	OH	Pickerington Ponds MP	39.88, -82.79
119248-1	OSAL119392	OSAL119248	OH	Pickerington Ponds MP	39.88, -82.80
119248-4	OSAL119395	OSAL119248	OH	Pickerington Ponds MP	39.88, -82.80
119250-2	OSAL119397	OSAL119250	OH	Sharon Woods MP	40.11, -82.95
119250-4	OSAL119399	OSAL119250	OH	Sharon Woods MP	40.11, -82.95
110559-1	OSAL110503	OSAL110559	OH	Fernald Preserve	39.29, -84.69
110559-2	OSAL110504	OSAL110559	OH	Fernald Preserve	39.29, -84.69
119600B	OSAL129714	OSAL119600	OH	Roads intersection	39.13, -84.79
115093A	OSAL129721	OSAL115093	OH	Cuyahoga Valley	41.289, -81.573
115093B	OSAL129722	OSAL115093	OH	Cuyahoga Valley	41.289, -81.573
115093C	OSAL129723	OSAL115093	OH	Cuyahoga Valley	41.289, -81.573
115093D	OSAL129724	OSAL115093	OH	Cuyahoga Valley	41.289, -81.573
115093E	OSAL129725	OSAL115093	OH	Cuyahoga Valley	41.289, -81.573
119928A	OSAL129691	OSAL119928	OH	Strouds Run SP	39.369, -82.042
119928B	OSAL129692	OSAL119928	OH	Strouds Run SP	39.369, -82.042
119928C	OSAL129693	OSAL119928	OH	Strouds Run SP	39.369, -82.042
119928D	OSAL129694	OSAL119928	OH	Strouds Run SP	39.369, -82.042

119572A	OSAL129695	OSAL119572	OH	Oak Openings MP	41.549, -83.854
119572B	OSAL129696	OSAL119572	OH	Oak Openings MP	41.549, -83.854
119572C	OSAL129697	OSAL119572	OH	Oak Openings MP	41.549, -83.854
119572D	OSAL129698	OSAL119572	OH	Oak Openings MP	41.549, -83.854
119567A	OSAL129707	OSAL119567	TN	Knoxville	35.390, -84.226
119567B	OSAL129708	OSAL119567	TN	Knoxville	35.390, -84.226
119567C	OSAL129709	OSAL119567	TN	Knoxville	35.390, -84.226
119567D	OSAL129710	OSAL119567	TN	Knoxville	35.390, -84.226
119567E	OSAL129712	OSAL119567	TN	Knoxville	35.390, -84.226
N8805A	OSAL129702	USNMENT01358805	OK	Washita Co.	35.411, -99.059
N8805B	OSAL129703	USNMENT01358805	OK	Washita Co.	35.411, -99.059
N8805C	OSAL129704	USNMENT01358805	OK	Washita Co.	35.411, -99.059
N8806A	OSAL129705	USNMENT01358806	OK	Washita Co.	35.411, -99.059
N8806B	OSAL129706	USNMENT01358806	OK	Washita Co.	35.411, -99.059
N8464B	OSAL129701	USNMENT01358464	VA	Warren Co.	38.893, -78.14
119951	OSAL129711	OSAL119951	GA	Statesboro	32.42, -81.77
119952A	OSAL129715	OSAL119952	GA	Statesboro	32.42, -81.77
115086	OSAL129836	OSAL115086	MN	Carlos Avery	45.287, -93.122
115087A	OSAL129717	OSAL115087	MN	Camp Ripley	46.076, -94.349
115087B	OSAL129718	OSAL115087	MN	Camp Ripley	46.076, -94.349
119918A	OSAL129829	OSAL119918	MN	Columbus	45.31, -93.02
119918B	OSAL129830	OSAL119918	MN	Stutsman Co.	47.23, -98.87
N128168B	OSAL129979	USNMENT01358520	ND	Stutsman Co.	47.23, -98.87
115139A	OSAL129834	OSAL115139	ME	Crescent Beach	43.56, -70.23
115139B	OSAL129835	OSAL115139	ME	Crescent Beach	43.56, -70.23
115140	OSAL129833	OSAL115140	ME	Unknown	Unknown
119276-2	OSAL129559	OSAL119276	MD	Aberdeen Providing Ground	39.46, -76.12
119276-3	OSAL129560	OSAL119276	MD	Aberdeen Providing Ground	39.46, -76.12
119276-5	OSAL129562	OSAL119276	MD	Aberdeen Providing Ground	39.46, -76.12
119276-6	OSAL129563	OSAL119276	MD	Aberdeen Providing Ground	39.46, -76.12
115101	OSAL129845	OSAL115101	CA	Napa Co	38.215, -122.33
115102A	OSAL129846	OSAL115102	CA	Lake Co.	39.139, -122.886
115102C	OSAL129848	OSAL115102	CA	Lake Co.	39.139, -122.886
115105B	OSAL129852	OSAL115105	WA	Whitman Co.	46.623, -117.228

Rickettsia spp. and *Anaplasma* spp. screening results divided by state.

1   **Table 2.** *Rickettsia* spp. and *Anaplasma* spp. screening results divided by state.

State	Year of collection	Number of ticks positive for rickettsiales	Total number of ticks positive			
			R. rhipicephali	*R. montanensis*	*R. bellii*	*Anaplasma* spp.
California	2017	2/3 (66.7%)			+ (2)	
Georgia	2013, 2016	1/2 (50%)		+ (1)		
Indiana	2017	0/3				
Maryland	2017	0/4				
Maine	2017	0/3				
Minnesota	2017	0/5				
North Dakota	2017	1/1 (100%)		+ (1)		
Ohio	2016, 2017	1/31 (3.2%)		+ (1)		
Oklahoma	2017	3/5 (60%)	+ (2)			+ (1)
Tennessee	2017	3/5 (60%)		+ (3)		
Virginia	2017	0/1				
Washington	2017	0/1				
Total		11/64 (17.2%)	2	6	2	1

Figure 1

Map of the USA showing the sampling locations.

The size of the circles reflects the number of specimens analyzed from each locality, from 1 to 5.
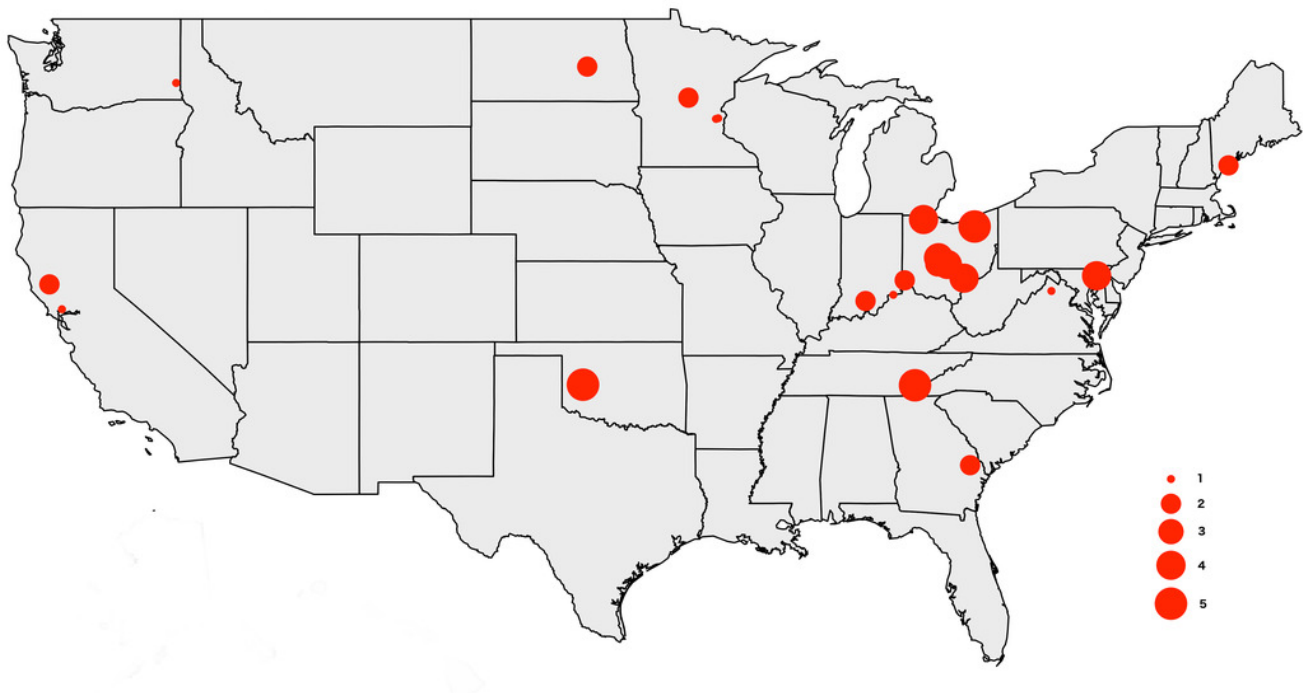
Figure 2

Barplots of the relative abundance of bacteria.

'(A) Barplot of the relative abundance of bacteria at the Order level. The x axis corresponds to individual tick specimens, and the y axis shows the relative abundance of the different bacteria Orders. Bacteria orders are color-coded as shown below the plot; (B) Similar to A), barplot of the relative abundance of the bacteria at the genus level.'
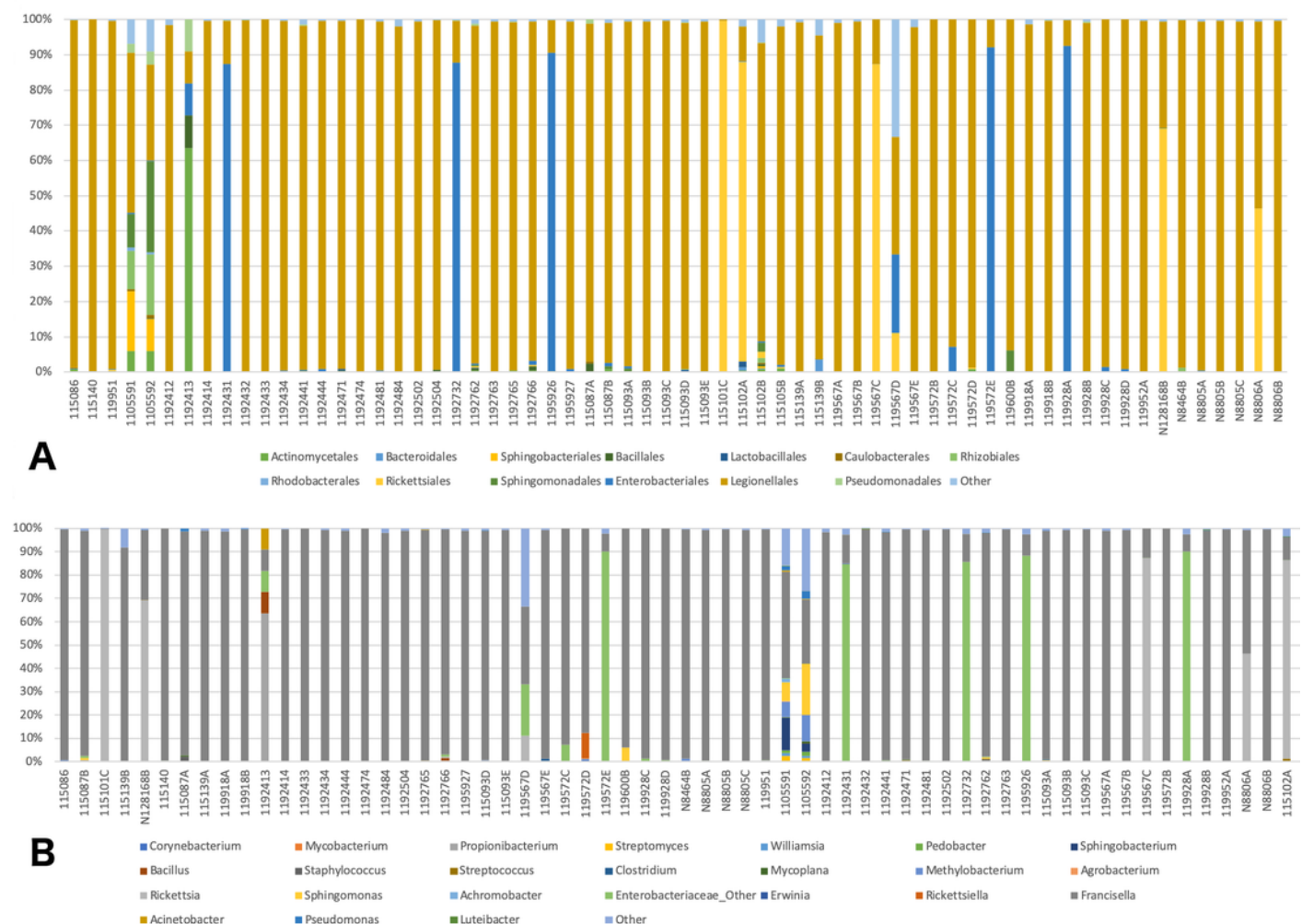
Figure 3

Dendrogram showing the relationship between all samples, and the different groupings.

Groupings names: Eastern, Western, CA (samples from California), WA (samples from Washington), and OH Hamilton Co (samples collected in Hamilton County, Ohio).
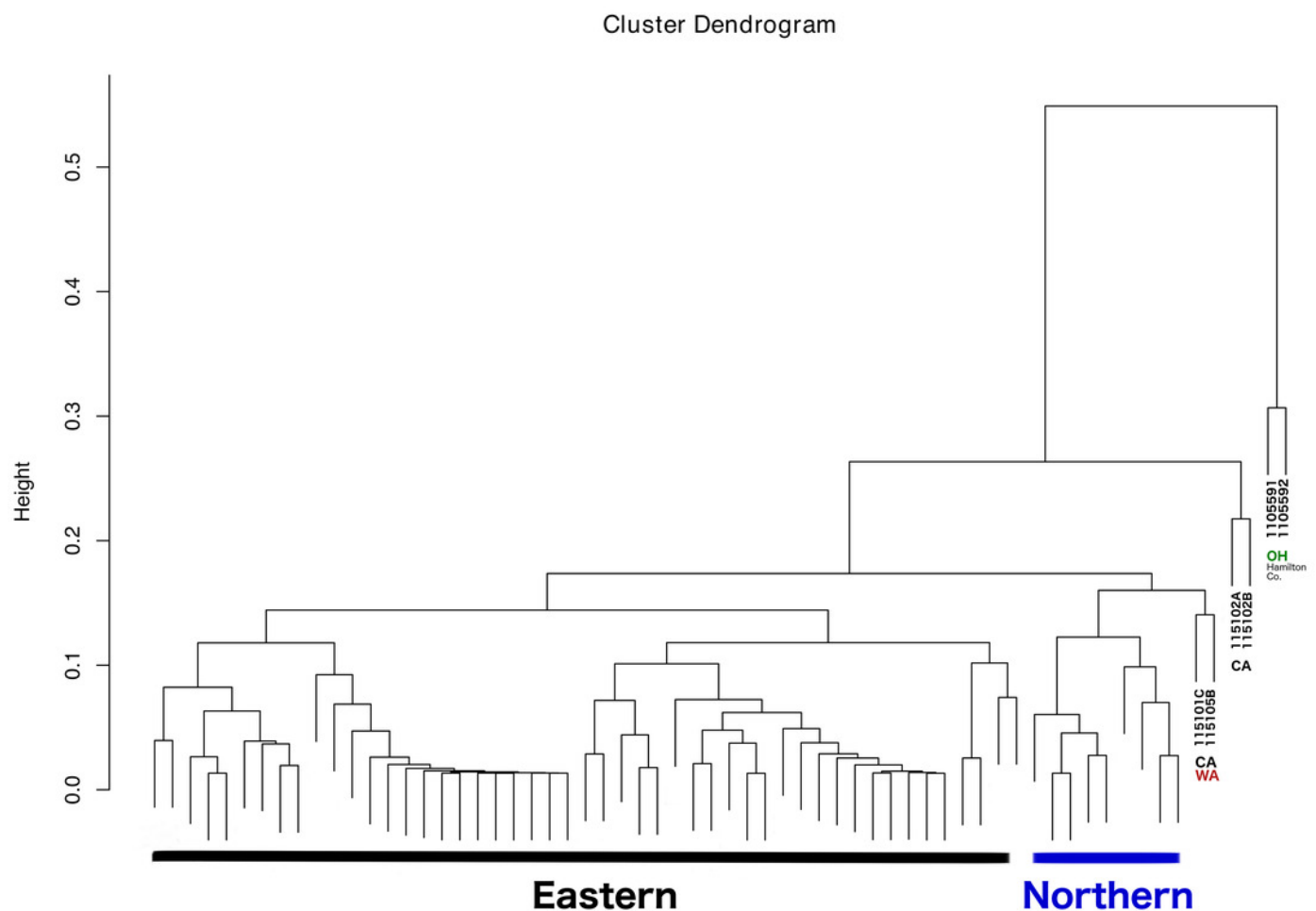
Figure 4

Non-metric multidimensional scaling and Principal coordinates analyses.

'(A) Non metric multidimensional scaling analysis of the samples. Different colors represent the different genetic clusters: Eastern (red), Northern (green), and Western (blue). (B) Principal coordinates analysis of the samples. Different colors represent the different genetic clusters: Eastern (red), Northern (green), and Western (blue). Axis 1 explains 74.9% of the variance, and Axis 2 explains 10% of it.'
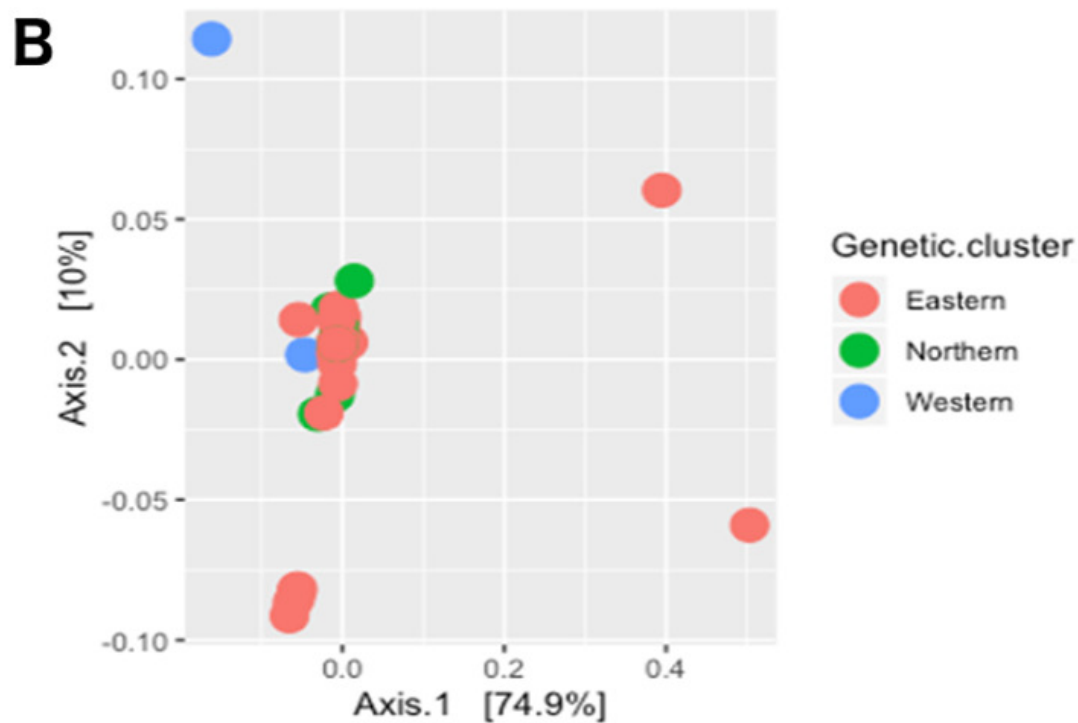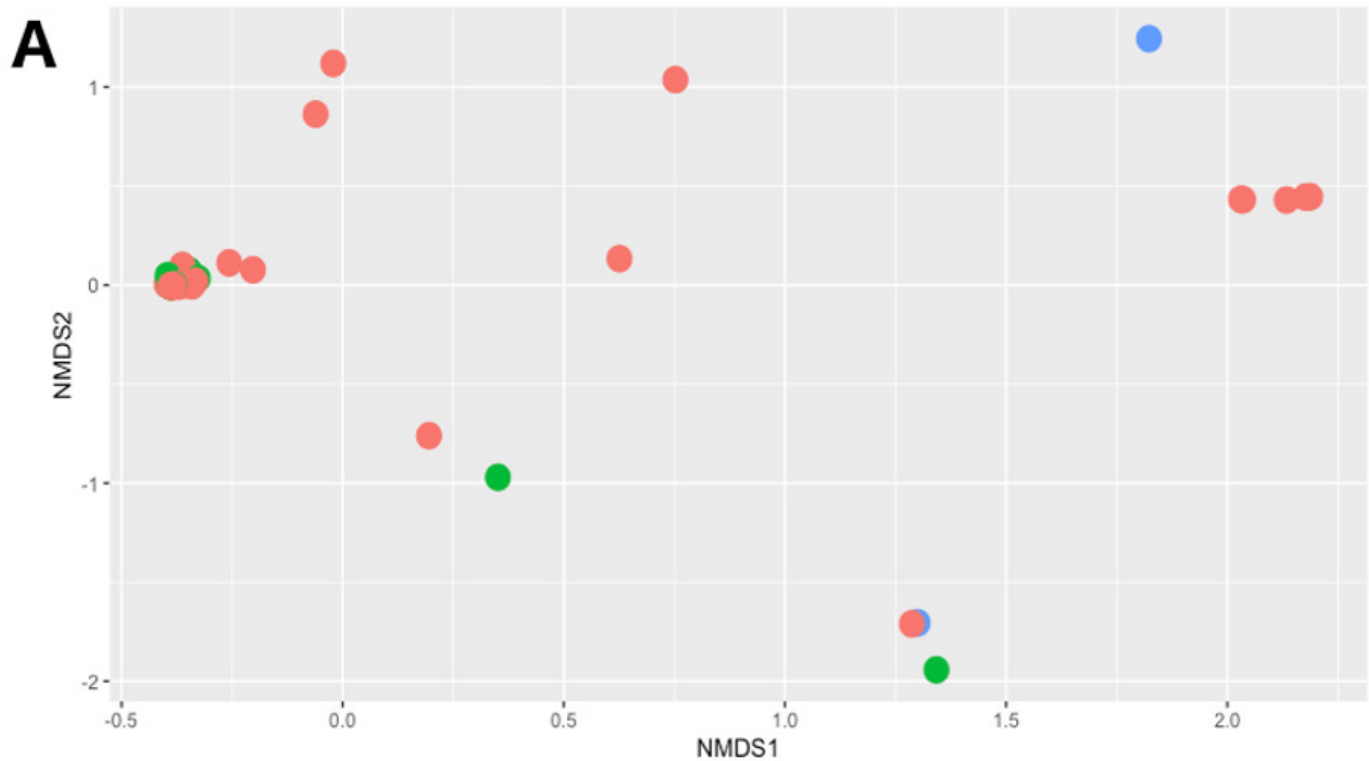
Figure 5

Average abundance of bacteria genera by genetic cluster.

Average abundance (y axis) of bacteria genera (x axis) by genetic clusters (Northern, Eastern, Western), and sex (male in red, female in blue).
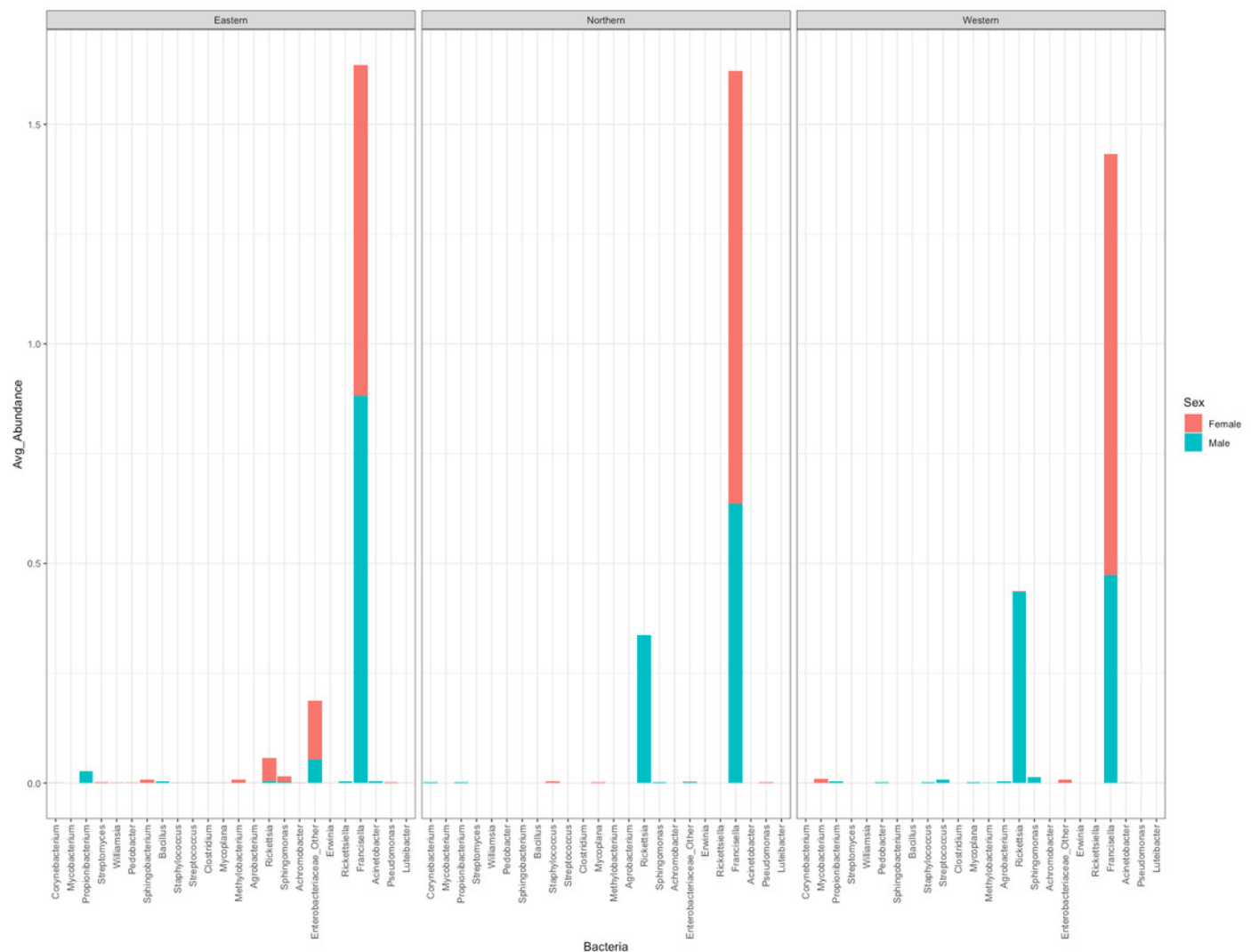
Figure 6

Pie chart plots of proportions of infection by genetic clusters.

Pie chart plots of proportions of infection by genetic clusters. Each plot corresponds to an agent (i.e *R. rhipicephali*, *R. montanensis*), and the colors represent the genetic clusters (N: Northern in red; E: Eastern in green; and W: Western in blue).