1   **Title:** Geography, not lifestyle, explains the population structure of free-living and host-

2   associated deep-sea hydrothermal vent snail symbionts

3

4   **Authors:** Michelle A. Hauer[1], Corinna Breusing[1], Elizabeth Trembath-Reichert[2], Julie A.

5   Huber[3], Roxanne A. Beinart[1*]

6

7   [1]University of Rhode Island, Graduate School of Oceanography, Narragansett, RI, USA

8   [2]Arizona State University, School of Earth and Space Exploration, Tempe, AZ, USA

9   [3]Woods Hole Oceanographic Institution, Department of Marine Chemistry and

10 Geochemistry, Woods Hole, MA, USA

11

12 ***Corresponding author**: Roxanne A. Beinart, rbeinart@uri.edu

13

14 **Keywords**: symbiosis; hydrothermal vents; population genomics; *Alviniconcha*; Mariana

15 Back-Arc; microbial biogeography

16

17

18

19

20

21

22

23

Abstract

24

25 Marine symbioses are predominantly established through horizontal acquisition of

26 microbial symbionts from the environment. However, genetic and functional

27 comparisons of free-living populations of symbionts to their host-associated

28 counterparts are sparse. Here, we assembled the first genomes of the

29 chemoautotrophic gammaproteobacterial symbionts affiliated with the deep-sea snail

30 *Alviniconcha hessleri* from two separate hydrothermal vent fields of the Mariana Back-

31 Arc Basin. We used phylogenomic and population genomic methods to assess

32 sequence and gene content variation between free-living and host-associated

33 symbionts. Our phylogenomic analyses show that the free-living and host-associated

34 symbionts of *A. hessleri* from both vent fields are populations of monophyletic strains

35 from a single species. Furthermore, genetic structure and gene content analyses

36 indicate that these symbiont populations are differentiated by vent field rather than by

37 lifestyle. Together, this work suggests that, despite the potential influence of host-

38 mediated acquisition and release processes on horizontally transmitted symbionts,

39 geographic isolation and/or adaptation to local habitat conditions are important

40 determinants of symbiont population structure and intra-host composition.

41

Background

43 Mutualistic animal-microbe associations are globally significant phenomena, shaping

44 the ecology and evolution of both host animals and microbial symbionts (1). These

45 symbiotic associations are maintained by transmission of symbionts from host parent to

46 progeny either 1) directly, for example via the germline (vertical transmission), 2)

47   indirectly, for example through an environmental population of symbionts (hereafter

48   referred to as "free-living" symbionts) (horizontal transmission); or, 3) via a combination

49   of both vertical and horizontal transmission (mixed mode transmission) (2).

50   Horizontal transmission is more commonly found in aquatic than terrestrial

51   habitats, likely due to the ease with which microbes can be transported in water

52   compared to air or soil (3). However, even for marine symbioses where horizontally

53   transmitted microbial symbionts are observed in the environment (4), it is not yet clear

54   whether free-living, environmental populations of symbionts represent host-associated

55   populations at the strain level, or whether their diversity and composition differs. Free-

56   living symbiont populations are likely shaped by the dynamic interactions with their

57   host—for example, host animals may "seed" the environment by the release of their

58   symbionts into the water column only upon host death (5) or via continuous release

59   from live adults (6,7). In addition, ecological and evolutionary processes, such as

60   dispersal barriers, natural selection, and genetic drift, can contribute to the diversity and

61   biogeography of environmental symbionts (8,9).

62   Deep-sea hydrothermal vents are discontinuous, island-like habitats dominated

63   by vent-endemic invertebrates that host primarily horizontally transmitted

64   chemoautotrophic bacterial symbionts, making them opportune natural systems for

65   understanding the biogeography of free-living microbial symbionts. In these mutualisms,

66   the symbiotic bacteria are either obtained during a narrow competence window in early

67   developmental stages or throughout the lifetime of the host (10,11) and are, in most

68   cases, housed intracellularly within the host's gill tissue. The symbionts oxidize

69   chemical reductants (e.g., H_2S, H_2, CH_4) in venting fluids to generate energy for the

70   production of organic matter, thereby providing the primary food source for the host in

71   an otherwise oligotrophic deep ocean (12) and accounting for the high ecosystem

72   productivity characteristic of hydrothermal vents (13–15). Despite reliance on horizontal

73   transmission, the majority of host species from hydrothermal vents affiliate with only one

74   or two specific endosymbiont phylotypes (i.e., species or genera based on 16S rRNA

75   gene sequence similarity) (12), possibly as a means to reduce the acquisition of

76   cheaters (16). While a significant number of studies have focused on the diversity,

77   composition and structure of the host-associated symbiont populations (e.g., 10,17–22),

78   their free-living, environmental stages remain poorly investigated, partly due to the

79   difficulty of detecting low abundance free-living symbionts in environmental samples. As

80   a consequence, few free-living symbiont studies exist. Most of these studies have so far

81   relied on investigations of particular marker genes (4,23); only one used an -omics

82   approach but was limited to a single metagenome (24).

83   A recent shotgun metagenomic study found putative free-living symbiont

84   populations of the provannid snail *Alviniconcha hessleri* in low-temperature diffuse

85   venting fluids at two distinct vent fields of the Mariana Back-Arc, Northwest Pacific

86   (15.5–18°N), (25), providing a unique opportunity to compare free-living and host-

87   associated stages of chemosynthetic symbionts at hydrothermal vents. *Alviniconcha*

88   *hessleri* belongs to the dominant fauna at hydrothermal vents in the Mariana Back-Arc

89   Basin, where it lives in nutritional endosymbiosis with one species of sulfur-oxidizing,

90   environmentally acquired Gammaproteobacteria (26,27). As an endemic species to the

91   Mariana region*, A. hessleri* is currently listed as "Vulnerable" on the International Union

92   for Conservation of Nature Red List of Threatened Species (https://www.iucnredlist.org),

93   highlighting the need to identify the factors that contribute to its limited biogeographic

94   range, including the population structure of its obligate microbial symbiont.

95   In this study, we applied phylogenomic and population genomic methods to

96   evaluate the evolutionary relationships as well as the genetic and functional variation of

97   *Alviniconcha hessleri* symbionts based on lifestyle by comparing free-living and host-

98   associated symbiont populations collected from the same habitats. In addition, we

99   addressed the effect of geography by comparing populations of both host-associated

100 and free-living symbionts at two vent fields, Illium and Hafa Adai, that are approximately

101 140 km apart and differ notably in their geochemistry: Hafa Adai is known to support

102 black smokers that emit high temperature fluids, with high amounts of hydrogen sulfide

103 (H_2S), whereas Illium is characterized by diffuse flow habitats that are lower in

104 temperature and concentration of H_2S (25).

105

106 **Methods**

107 <u>Host-associated symbiont collection, sequencing, and genome assemblies</u>

108 Three *A. hessleri* specimens each were collected from snail beds at the Illium vent field

109 (3582 m) and the Voodoo Crater-2 (VC2) location within the Hafa Adai vent field (3277

110 m) in the Mariana Back-Arc Basin (Figure 1) using the remotely operated vehicle

111 *SuBastian* on board the R/V *Falkor* in 2016. Symbiont-bearing snail gill tissues were

112 dissected and stored in RNALater™ (Thermo Fisher Scientific, Inc.) at -80ºC until DNA

113 extraction with the Zymo Quick DNA 96 Plus and ZR-96 Clean-up kits (Zymo Research,

114 Inc.). High-throughput Illumina 150-bp paired-end libraries for all six samples were

115 prepared and sequenced by Novogene Corporation, Inc. (Beijing, China) with an

116 average yield of 58 million total Illumina reads (Supplementary Table 1). In addition, one

117 gill sample from each vent field—Hafa Adai 172 (VC2) and Illium 13—was selected for

118 long-read Nanopore sequencing on 2–3 MinION flow cells (Oxford Nanopore

119 Technologies, Oxford, UK) using the SQK-LSK109 ligation kit (Supplementary Table 1).

120　　　　Raw Illumina reads were trimmed with Trimmomatic v0.36 (28) and common

121 sequence contaminants were removed by mapping against the PhiX and human

122 (GRCh38) reference genomes. Nanopore reads were base-called with Albacore (Oxford

123 Nanopore Technologies, Oxford, UK) and adapter-clipped with Porechop

124 (https://github.com/rrwick/Porechop). For Illium 13, symbiont Illumina and Nanopore

125 reads were extracted through mapping against a draft co-assembly constructed with

126 MEGAHIT (29) and reassembled with SPAdes v3.13.1 (30) using kmers between 21-91

127 in increments of 10. For Hafa Adai 172, all trimmed Illumina and Nanopore reads were

128 assembled with metaSPAdes v3.13.1 (31) using the same parameters. Manual binning

129 of the assemblies was conducted with GBTools (32) and contigs <200 bp and <500 bp

130 were excluded from the Illium and Hafa Adai-VC2 assemblies, respectively. Assemblies

131 were then scaffolded with SSPACE-Standard v3.0 (33) and SLR (34) and gapfilled with

132 GapFiller v1-10 (35) and LR_GAPCLOSER (36). Final assemblies were polished with

133 Pilon (37). Trimmed Illumina reads for the remaining samples—Illium 11, Illium 17, Hafa

134 Adai 60, and Hafa Adai 64—were assembled with metaSPades v3.13.1 as described

135 above and binned with MaxBin (38). All metagenome-assembled genomes (MAGs)

136 were quality checked with Quast v5.0.2 (39) and CheckM v1.0.18 (40) and

137 taxonomically assigned through the Genome Taxonomy Database toolkit (41).

138

139   Free-living symbiont collection, sequencing, and genome assembly

140   High-quality MAGs of environmental *A. hessleri* symbionts from the Illium

141   (GCA_003972985.1) and Hafa Adai-VC2 (GCA_003973075.1) vent sites were retrieved

142   from a previous study (25). The fluid samples from which these MAGs were assembled

143   were collected in direct vicinity of snail beds where the *A. hessleri* specimens were

144   obtained (25). We further included a third MAG from diffuse-venting fluids at the Voodoo

145   Crater-1 (VC1) (GCA_003973045.1) location within the Hafa Adai vent field, ~5 meters

146   from Hafa Adai-VC2. All details of the hydrothermal fluid collection, sample storage,

147   sample processing, sequencing, assembly, and binning of metagenome-assembled

148   genomes can be found in ref. (25). Information about raw sequencing reads is provided

149   in Supplementary Table 1.

150   Raw Illumina metagenomics reads from the other vent sites sampled in ref. (25)

151   —Burke, Alice Springs, Perseverance and Hafa Adai-Alba—were assembled and

152   binned with the methods described above, but did not produce usable symbiont MAGs.

153

154   Genome similarity and phylogenomic analyses

155   To confirm that all symbiont MAGs belong to the same bacterial species, we calculated

156   average nucleotide identities (ANIs) via FastANI (42). A phylogenomic tree that included

157   the six host-associated and the three free-living symbiont MAGs as well as reference

158   genomes of other chemosynthetic Gammaproteobacteria (Supplementary Table 2) was

159   then constructed with IQ-TREE2 (43) based on 70 single copy core genes in the

160   Bacteria_71 collection (44). Parameter choice for phylogenomic reconstructions

161   followed ref. (18).

162

Population structure and gene content analysis

164 To determine symbiont population structure according to geography and lifestyle, we

165 inferred DNA sequence polymorphisms in the free-living and host-associated samples

166 by mapping metagenomic reads to a pangenome created with Panaroo (45) from all

167 nine symbiont MAGs. Variants were called and filtered following the pipeline in ref. (18).

168 All samples from Illium and Hafa Adai met our minimum 10x coverage threshold. Free-

169 living samples from Burke and Alice Springs mapped at 5.9x and 3.9x coverage,

170 respectively. The population structure and gene content analyses (see below) were

171 repeated for these lower-coverage samples. All other free-living metagenomic samples

172 from the remaining vent sites collected in ref. (25) had an insufficient number of reads

173 mapped for further analyses. Principal coordinate analysis (PCoA) plots were created

174 based on nucleotide counts converted to Bray-Curtis dissimilarities with the ggplot2 (46)

175 and vegan (47) packages in Rstudio (48). To quantify the qualitative variant calling

176 results depicted in the PCoAs, fixation indices (F_{ST}) between individual samples were

177 calculated following ref. (49) and plotted with pheatmap (50). The method from ref. (49)

178 as well as scikit-allel (https://github.com/cggh/scikit-allel) were further used to calculate

179 pairwise F_{ST} values between samples pooled by lifestyle or vent field.

180 Gene content variation among symbiont populations was determined via

181 Pangenome-based Phylogenomic Analysis (PanPhlAn) (51) and visualized through a

182 PCoA plot based on the Jaccard Similarity Coefficient. Genes that were uniquely

183 associated with lifestyle and vent field, respectively, were extracted from the PanPhlAn

184 gene presence/absence matrix. Functional predictions for these genes were either

185   obtained from the Prokka (52) annotations created during pangenome construction or

186   inferred by blasting the respective protein sequences against the NR database.

187   Hypothetical and unknown proteins were further annotated via KEGG (53) and

188   Alphafold (54). Differences in gene content between symbiont populations were

189   visualized through Likert plots with the HH package (55) in RStudio.

190

191   <u>Validation of free-living symbiont populations</u>

192   To gain confidence that the symbionts detected in our environmental samples

193   represented truly "free-living" symbiont stages as opposed to symbionts associated with

194   host larvae or shed gill cells, we calculated the ratio of symbiont 16S rRNA genes to

195   host mitochondrial CO1 genes in all nine samples by mapping metagenomic reads to

196   *Alviniconcha* symbiont 16S rRNA and host mtCO1 gene databases. To account for false

197   positive mappings, we created additional background databases consisting of select

198   bacterial (SUP05 clade bacteria, *Thiomicrospira*, and *Marinomonas*) and mollusk gene

199   sequences. Bacterial 16S rRNA genes were downloaded from SILVA (56), while all

200   *Alviniconcha* and mollusk mtCO1 genes were downloaded from BOLD (57). BBSplit

201   (https://sourceforge.net/projects/bbmap/) was then used to separate *Alviniconcha*

202   symbiont and host reads based on the taxon-specific and background 16S rRNA and

203   CO1 gene databases.

204

205   **Results**

206   <u>Free-living and host-associated symbionts belong to the same bacterial species</u>

207 Our analysis included nine *A. hessleri* symbiont MAGs from the Illium and Hafa Adai

208 vent fields: six host-associated symbiont genomes assembled in this study, and three

209 previously published, free-living symbiont genomes from the diffuse venting fluids

210 around *A. hessleri* beds (25) (Supplementary Table 3). All host-associated and two of

211 the three free-living MAGs were of very high quality, with >90% completeness and <3%

212 contamination. The third free-living MAG, Hafa Adai-VC1, had a medium quality (~67%

213 completeness). ANI values between all MAGs were >97.7% (Supplementary Table 4),

214 suggesting that the nine *A. hessleri* symbiont genomes belong to the same bacterial

215 species (42), *Thiolapillus* sp. based on the Genome Taxonomy Database, and

216 confirming that the previously assembled free-living symbiont genomes were indeed *A.*

217 *hessleri* symbionts (Supplementary Table 3). Corroborating the ANI results, the nine *A.*

218 *hessleri* MAGs were monophyletic in our phylogenomic analysis relative to the

219 gammaproteobacterial symbionts of other vent invertebrates (Figure 2). In agreement

220 with phylogenetic analyses of the 16S rRNA gene (27), the nearest neighbors of the *A.*

221 *hessleri* symbionts were the *Ifremeria nautilei* SOX symbiont, as well as *Thiolapillus*

222 *brandeum*, a microbe not known to be symbiotic (58).

223

224 Environmental samples contain free-living symbiont populations

225 To investigate whether the symbionts observed in the diffuse flow samples were true

226 free-living symbionts rather than symbionts associated with *A. hessleri* larvae or shed

227 gill tissue, we calculated the ratio of symbiont 16S rRNA gene to host mitochondrial

228 CO1 gene reads in all nine environmental and host-associated samples (Table 1). If the

229 environmental symbiont samples were associated with larvae or host tissue debris/cells,

230   we expect the ratio in the environmental and host-associated samples to be similar to

231   one another. However, the 16S rRNA : mtCO1 ratio was consistently orders of

232   magnitude higher in environmental samples than in host-associated samples, indicating

233   a population of symbiont cells independent from host tissue. This finding provides

234   evidence that our environmental samples include truly free-living *A. hessleri* symbiont

235   populations.

236

237   *A. hessleri* symbiont populations are structured primarily by vent field, not lifestyle

238   Our genome assemblies from both host tissue and diffuse vent fluids likely represent

239   the dominant symbiont strain in each sample, but do not reveal the full extent of strain-

240   level population variation between samples. To determine whether *A. hessleri*

241   symbionts form subpopulations consistent with geography or lifestyle, we created a

242   pangenome out of the individual symbiont MAGs that we used as reference for variant

243   calling (Supplementary Tables 5, 6). Our variant detection method resulted in 2177

244   sequence polymorphisms for investigation of population genomic structure based on F_{ST}

245   and ordination analyses (Figure 3).

246   F_{ST} values were calculated pairwise between all nine populations (Figure 3a), as

247   well as between samples pooled by lifestyle and vent field. F_{ST} values range from 0 to 1,

248   where an F_{ST} value of 1 indicates that samples form genetically isolated subpopulations,

249   while an F_{ST} value of 0 indicates that the samples form a single, well-mixed population.

250   Between individual samples, pairwise F_{ST}s showed a moderate (0.2–0.5) to strong

251   (>0.5) differentiation, indicating that all samples represent distinct subpopulations with

252   limited genetic exchange among each other (Figure 3a, Supplementary Table 7).

253   Genetic isolation among individual samples was typically stronger between (0.5395–

254   0.7560) than within (0.2123–0.4594) vent localities (i.e., Illium vs Hafa Adai-VC2).

255   Within vent sites, the degree of differentiation was comparable among samples

256   independent of lifestyle at Illium, while host-associated samples were more similar to

257   one another than to free-living samples at Hafa Adai-VC2. When samples were pooled,

258   overall pairwise F_{ST} values were markedly higher by vent field (0.4679 ± 0.0300 s.d.)

259   than by lifestyle (0.0544 ± 0.0139 s.d.). The dominant effect of geography on symbiont

260   population structure was supported by PCoAs where both free-living and host-

261   associated samples from Illium clustered distinctly from Hafa Adai (VC1 and VC2)

262   (Figure 3b). Despite the fact that Hafa Adai-VC1 and -VC2 differ spatially by only ~5

263   meters, the free-living VC1 sample formed its own distinct subpopulation from both

264   host-associated and free-living populations at VC2 (F_{ST}: 0.5771–0.6731), suggesting

265   very fine-scale geographic or environmental structuring.

266   These patterns were consistent in analyses based on 1271 and 793 variant sites

267   that included the free-living, low coverage symbiont samples from Burke and Alice

268   Springs, respectively (Supplementary Figure 1, 4; Supplementary Tables 8, 9). Burke

269   represented the most divergent population, reaching F_{ST} values > 0.8 in all pairwise

270   comparisons. Although Alice Springs clustered closely with free-living and host-

271   associated symbionts from Illium in the PCoAs, F_{ST} values indicated a high degree of

272   genetic isolation for this population (F_{ST} > 0.7). Analyses with samples pooled by vent

273   field confirmed patterns of strong genetic differentiation between geographic locations

274   without evidence for isolation-by-distance (Supplementary Table 10).

275

276   *A. hessleri* symbiont gene content differs by vent field, not lifestyle

277   Gene content variation between symbiont populations was assessed based on lifestyle

278   and geography. Similar to the population structure analyses, PCoA plots based on gene

279   content variation across all nine host-associated and free-living populations revealed

280   clustering by vent field but not by lifestyle: symbiont populations from Hafa Adai-VC1

281   and Hafa Adai-VC2 were more similar to one another than to Illium (Figure 4), although

282   Hafa Adai-VC1 clustered as an independent population from all other samples.

283   Gene content differed more substantially by geography than by lifestyle: the

284   Illium symbionts had 44 unique gene clusters, while the Hafa Adai (VC1 and VC2)

285   symbionts had 26 (Figure 5a, Supplementary Table 11). Hypothetical and unknown

286   proteins based on Prokka and the NR database were also assessed via KEGG and

287   Alphafold, but yielded low confidence results. Of the successfully annotated genes

288   unique to the Illium symbionts, most were predicted to be involved in the mobilome and

289   DNA metabolism, followed by membrane transport, virulence, disease, defense; RNA

290   metabolism; sulfur metabolism; cell signaling and regulation; conjugation; iron

291   metabolism; glycolysis and gluconeogenesis; and detoxification and stress response.

292   Genes unique to the Hafa Adai (VC1 & VC2) symbionts were predominantly associated

293   with the mobilome, followed by membrane transport; RNA metabolism; motility and

294   chemotaxis; DNA metabolism; virulence, disease and defense; and glycolysis and

295   gluconeogenesis. Only three unique genes differentiated the host-associated and free-

296   living symbiont populations across both vent fields (group_681, group_2104,

297   group_2131, Supplementary Table 11). However, these genes could not be

298   characterized by any database we used for functional annotations.

13

299 Given the small-scale geographic structuring found between VC1 and VC2 at

300 Hafa Adai, we also compared the unique genes between Illium and Hafa-Adai VC2

301 symbionts alone (i.e., without VC1) (Supplementary Table 12, Figure 5b). In this case,

302 there were an additional 18 unique genes for symbionts from Illium and 2 unique genes

303 for symbionts from Hafa Adai-VC2 (Figure 5b). Only one of the genes unique to the

304 Hafa Adai symbionts could be annotated and fell under the larger subcategory of

305 "Virulence, Disease and Defense", whereas unique genes of the Illium symbionts

306 spanned a variety of metabolic functions.

307 Analyses that included symbiont reads from Alice Springs and Burke

308 (Supplementary Figures 2, 3, 5; Supplementary Tables 13, 14) further supported the

309 effect of geography over lifestyle on gene content variation in the *A. hessleri* symbionts.

310 The population at Burke harbored a single unique, uncharacterized gene

311 (Supplementary Table 13). When pooled with Illium as a "northern site", additional

312 genes unique to DNA metabolism and membrane transport were found, followed by

313 genes involved in the mobilome, RNA metabolism, virulence, glycolysis and

314 gluconeogenesis, cell signaling, conjugation and stress response (Supplementary

315 Figure 3; Supplementary Table 13).

316 Alice Springs harbored three uncharacterized or hypothetical genes. When all

317 three northern sites (Alice Springs, Illium, and Burke) were pooled together, seven

318 unique genes were found. Four of these were related to DNA metabolism, virulence,

319 conjugation and transposition (Supplementary Table 14). Since Alice Springs and Illium

320 are more geochemically similar to one another than either vent is to Burke, we also

321 investigated the unique genes shared by these two vent fields alone: four unique genes

322 were found, one of which fell under the functional category of Virulence, Disease and

323 Defense.

324

Discussion

326 Here, we compared free-living and host-associated symbiont populations of

327 *Alviniconcha hessleri* from two vent fields in the Mariana Back-Arc. Based on ANI and

328 taxonomic assignments, our nine representative, medium- to high-quality MAGs can be

329 considered to represent a single species within the genus *Thiolapillus* (58). Our results

330 provide strong evidence that diffuse fluid flow microbial communities include populations

331 of free-living symbionts, further supporting an expected model of horizontal transmission

332 in *Alviniconcha* species (18,59).

333   Both population structure and gene content analyses suggest that *A. hessleri*

334 symbionts form subpopulations that segregate by geography more strongly than by

335 lifestyle. These patterns agree with previous studies of host-associated *A. hessleri*

336 symbiont biogeography at the 16S rRNA gene level (27) and with observations in other

337 horizontally-transmitted associations from hydrothermal vents, such as bathymodiolin

338 mussels (49,60) and provannid snails (18), that have been shown to partner with

339 habitat-specific symbiont strains. These results are also consistent with studies of non-

340 symbiotic hydrothermal vent microbial communities, which show that microbes are

341 shaped by their local environment (61). Uptake of environmental symbiont strains bears

342 a risk of infection to the host by cheaters (16), but also enhances an animal's ability to

343 flexibly associate with locally available symbiont strains and, therefore, to maximize the

344 habitat range in which they can settle (2,18). Hydrothermal vents are ephemeral and

345 geochemically dynamic habitats, harboring microbial communities that are shaped by

346 environmental conditions both within and between vent fields (61); therefore, it may be

347 ecologically and evolutionarily advantageous for vent animals to acquire symbiont

348 strains that are likely locally adapted (62).

349   The dynamics of microbial interaction with the host during acquisition and release

350 processes can have significant impacts on the population structure and composition of

351 horizontally transmitted symbionts. It is not known whether *A. hessleri* can replenish or

352 recycle its symbionts, or if symbiont acquisition occurs only once upon settlement. For

353 example, hydrothermal vent tubeworms seed the environment with their symbionts only

354 upon death (5), *Bathymodiolus* mussels can acquire and release their symbionts

355 throughout their lifetime (11,63), and *Vibrio fischeri* symbionts are expelled every

356 morning by their sepiolid squid host (64). In *V. fischeri,* it is well established that

357 evolution in the free-living stage—for example, via horizontal gene transfer—impacts

358 the evolution of host-microbe interactions, though the role of novel mutations remains

359 unclear (63). Although *A. hessleri* symbionts were overall more strongly partitioned by

360 geography than by lifestyle, all symbiont samples were genetically distinct from each

361 other and formed separate free-living or host-associated subpopulations. These findings

362 suggest that symbiont exchanges between host and environment throughout the lifetime

363 of the host are limited but might occur occasionally via symbiont uptake or release (49),

364 thereby leading to mixing of host-associated and free-living symbiont pools. Periodic

365 switching of symbiont strains could increase shared genetic variation among intra- and

366 extra-host symbiont populations, while maintaining geographic differentiation in the

367 presence of dispersal barriers and/or environmental selection. All samples from Illium

368   showed a comparable degree of differentiation from each other, while samples from

369   Hafa Adai were notably divergent between free-living and host-associated lifestyles.

370   These patterns could arise from differences in the sampling locations of the free-living

371   symbiont populations (e.g., distance from the snail beds) and/or the age of the

372   *Alviniconcha* host individuals. Although we do not have size-related data for the

373   collected specimens, it is possible that the snail individuals from Hafa Adai were older

374   than those from Illium, giving host-associated symbiont populations more time to

375   diverge from their free-living counterparts. Strong genetic differentiation between host-

376   associated and free-living symbiont populations can be expected if hosts take up similar

377   symbiont strains that have limited exchange with the environment post-infection, while

378   the free-living symbiont population experiences more turnover.

379   The high genetic isolation of symbiont populations observed between vent fields

380   may reflect the influence of both neutral (e.g., dispersal barriers and isolation by

381   distance) and selective processes (e.g., adaptation to habitat differences between vent

382   fields) on symbiont biogeography. Illium, Burke, and Alice Springs are all northern vent

383   fields within the Mariana Back-Arc Basin that are characterized by sites of low-

384   temperature diffuse fluid flow, while Hafa Adai is located further south and contains

385   high-temperature black smokers (25). Illium and Alice Springs are similar

386   geochemically, notably in that they are both low in H_2S concentrations. The close

387   proximity (~360 m) and overlap in geochemical characteristics between Alice Springs

388   and Illium may explain why these vent fields clustered together in our population

389   structure analyses. By contrast, Burke is located about 4 km from both Illium and Alice

390   Springs and has a distinct geochemical signature, which could contribute to the high

391   genetic isolation seen for this vent field. Overall, however, no clear pattern of isolation-

392   by-distance was observed, indicating that ecological factors might play a more important

393   role than dispersal barriers in shaping symbiont population structure, in agreement with

394   the oceanographic connectivity between the northern and central Mariana Back-Arc

395   Basin (65).

396   Interestingly, Hafa Adai-VC1 — while more similar to Hafa Adai-VC2 than any

397   other vent site — represented its own symbiont sub-population, suggesting small scale

398   population structuring of symbionts within vent fields. Local patchiness of symbionts, as

399   observed in our study, mirrors patterns found for host-associated symbionts of cold-

400   seep vestimentiferan tubeworms (66) and *Acropora* corals (67). Although Hafa Adai-

401   VC1 and -VC2 were only ~5 meters apart, it is possible that *Alviniconcha hessleri*

402   symbionts have extremely low dispersal potential that could be further reduced by

403   small-scale circulation within vent sites due to physical structuring in the subseafloor

404   (68,69). Alternatively, micro-niche adaptation driven by locally fluctuating environmental

405   conditions might contribute to these patterns.

406   Among the identified differences in gene content, symbionts from Illium uniquely

407   harbored genes related to iron and sulfur metabolism. As iron concentrations appear to

408   be reduced at northern Mariana Back-Arc vents (70) and are typically lower in diffuse

409   flow than black smoker fluids, it is possible that symbionts at Illium harbor high affinity

410   Fe^{2+} transporters to efficiently obtain this essential element for their metabolism. All

411   symbiont populations, including the low coverage samples from Alice Springs and

412   Burke, showed differences in the presence of genes related to the mobilome and

413   virulence, disease and defense. This suggests that each vent field supports distinct viral

414   communities, as hydrothermal vent viruses have restricted bacterial and archaeal host

415   ranges, and viral communities are typically endemic to a given vent site due to limited

416   dispersal or environmental selection (71,72). The high number of unique genes related

417   to the mobilome may be a consequence of integrated phage-derived genetic material

418   that reflect the local, free-living viral communities.

419

420   **Conclusions**

421   Our research demonstrates that *Alviniconcha hessleri* symbiont populations are

422   primarily structured by geography rather than by their host-associated or free-living

423   lifestyle. Future work using population genomic approaches should help clarify the

424   predominant force(s) shaping the geographic population structure, as recent analyses of

425   the symbionts associated with other *Alviniconcha* species suggests that both genetic

426   drift and local adaptation play a role in symbiont biogeography (18). Although our

427   analyses indicate a weak effect of lifestyle on symbiont genetic structure, it is possible

428   that free-living and host-associated populations are characterized by differences in gene

429   expression. A comparison of gene expression between lifestyles may provide additional

430   clarity on the extent to which these symbiont sub-populations differ functionally. Our

431   work also strengthens previous evidence for horizontal symbiont transmission in

432   *Alviniconcha* species (18,59), despite the fact that almost nothing is currently known

433   about the dynamics of symbiont acquisition and release in these species. Given that *A.*

434   *hessleri* has been classified as "Vulnerable" on the IUCN Red List

435   (https://www.iucnredlist.org) and is a dominant species at vents that are part of the

436   Marianas Trench Marine National Monument, it is critical for future conservation and

437 management that we understand the genetic connectivity of the symbiotic microbes that

438 support this foundation species.

439

Declarations

Data availability

The datasets supporting the conclusions of this article are available in the National Center for Biotechnology Information repository, under BioProject number PRJNA763533. The previously published, free-living raw sequencing reads and corresponding MAGs are available at the National Center for Biotechnology Information under BioProject PRJNA454888.

Competing Interests

The authors declare that they have no competing financial interests.

463   **References**

464   1.   McFall-Ngai M, Hadfield MG, Bosch TCG, Carey H V., Domazet-Lošo T, Douglas

465   AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc

466   Natl Acad Sci U S A. 2013;110(9):3229–3236.

467   2.   Vrijenhoek RC. Genetics and evolution of deep-sea chemosynthetic bacteria and

468   their invertebrate hosts. In: S. Kiel, editor. The Vent and Seep Biota: Aspects from

469   Microbes to Ecosystems. Dordrecht, Netherlands: Springer; 2010. p. 15–49.

470   3.   Russell SL. Transmission mode is associated with environment type and taxa

471   across bacteria-eukaryote symbioses: A systematic review and meta-analysis.

472   FEMS Microbiol Lett. 2019;366(3):1–13.

473   4.   Harmer TL, Rotjan RD, Nussbaumer AD, Bright M, Ng AW, DeChaine EG, et al.

474   Free-living tubeworm endosymbionts found at deep-sea vents. Appl Environ

475   Microbiol. 2008;74(12):3895–3898.

476   5.   Klose J, Polz MF, Wagner M, Schimak MP, Gollner S, Bright M. Endosymbionts

477   escape dead hydrothermal vent tubeworms to enrich the free-living population.

478   Proc Natl Acad Sci. 2015;112(36):11300–11305.

479   6.   Doino JA, McFall-Ngai MJ. A transient exposure to symbiosis-competent bacteria

480   induces light organ morphogenesis in the host squid. Biol Bull. 1995;189(3):347–

481   355.

482   7.   Nyholm SV, McFall-Ngai MJ. Sampling the light-organ microenvironment of

483   *Euprymna scolopes*: description of a population of host cells in association with

484   the bacterial symbiont *Vibrio fischeri*. Biol Bull. 1998;195(2):89–97.

485   8.   Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, et

486   al. Microbial biogeography: Putting microorganisms on the map. Nat Rev

487   Microbiol. 2006;4(2):102–112.

488   9.   Dickey JR, Swenie RA, Turner SC, Winfrey CC, Yaffar D, Padukone A, et al. The

489   utility of macroecological rules for microbial biogeography. Front Ecol Evol.

490   2021;9:633155.

491 10. Nussbaumer AD, Fisher CR, Bright M. Horizontal endosymbiont transmission in

492   hydrothermal vent tubeworms. Nature. 2006;441(7091):345–348.

493 11. Wentrup C, Wendeberg A, Schimak M, Borowski C, Dubilier N. Forever

494   competent: Deep-sea bivalves are colonized by their chemosynthetic symbionts

495   throughout their lifetime. Environ Microbiol. 2014;16(12):3699–3713.

496 12. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of

497   harnessing chemosynthesis. Nat Rev Microbiol. 2008;6:725–740.

498 13. Olins HC, Rogers DR, Frank KL, Vidoudez C, Girguis PR. Assessing the influence

499   of physical, geochemical and biological factors on anaerobic microbial primary

500   productivity within hydrothermal vent chimneys. Geobiology. 2013;11(3):279–293.

501 14. McNichol J, Stryhanyuk H, Sylva SP, Thomas F, Musat N, Seewald JS, et al.

502   Primary productivity below the seafloor at deep-sea hot springs. Proc Natl Acad

503   Sci U S A. 2018;115(26):6756–6761.

504 15. Beinart RA, Gartman A, Sanders JG, Luther GW, Girguis PR. The uptake and

505   excretion of partially oxidized sulfur expands the repertoire of energy resources

506   metabolized by hydrothermal vent symbioses. Proc R Soc B Biol Sci.
507   2015;282:20142811.

508 16. Douglas AE. Conflict, cheats and the persistence of symbioses. New Phytol.
509   2008;177(4):849–858.

510 17. Lan Y, Sun J, Chen C, Sun Y, Zhou Y, Yang Y, et al. Hologenome analysis
511   reveals dual symbiosis in the deep-sea hydrothermal vent snail *Gigantopelta*
512   *aegis.* Nat Commun. 2021;12(1):1–15.

513 18. Breusing C, Genetti M, Russell SL, Corbett-Detig RB, Beinart RA. Horizontal
514   transmission enables flexible associations with locally adapted symbiont strains in
515   deep-sea hydrothermal vent symbioses. Proc Natl Acad Sci U S A.
516   2022;119(14):1–11.

517 19. Beinart RA, Sanders JG, Faure B, Sylva SP, Lee RW, Becker EL, et al. Evidence
518   for the role of endosymbionts in regional-scale habitat partitioning by
519   hydrothermal vent symbioses. Proc Natl Acad Sci U S A. 2012;109(47):1–10.

520 20. Beinart RA, Luo C, Konstantinidis KT, Stewart FJ, Girguis PR. The bacterial
521   symbionts of closely related hydrothermal vent snails with distinct geochemical
522   habitats show broad similarity in chemoautotrophic gene content. Front Microbiol.
523   2019;10:1818.

524 21. Breusing C, Johnson SB, Tunnicliffe V, Vrijenhoek RC. Population structure and
525   connectivity in Indo-Pacific deep-sea mussels of the *Bathymodiolus*
526   *septemdierum* complex. Conserv Genet. 2015;16(6):1415–1430.

527 22. Hinzke T, Kleiner M, Meister M, Schlüter R, Hentschker C, Pané-Farré J, et al.
528   Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis.

529   Elife. 2021;10:1–45.

530   23.   Ücker M. Metagenomic analyses of a deep-sea mussel symbiosis. Dissertation,
531   University of Bremen. 2021.

532   24.   Polzin J, Arevalo P, Nussbaumer T, Polz MF, Bright M. Polyclonal symbiont
533   populations in hydrothermal vent tubeworms and the environment. Proc R Soc B
534   Biol Sci. 2019;286:20181281.

535   25.   Trembath-Reichert E, Butterfield DA, Huber JA. Active subseafloor microbial
536   communities from Mariana back-arc venting fluids share metabolic strategies
537   across different thermal niches and taxa. ISME J. 2019;13(9):2264–2279.

538   26.   Endow K, Ohta S. The symbiotic relationship between bacteria and a
539   mesogastropod snail, *Alviniconcha hessleri*, collected from hydrothermal vents of
540   the Mariana Back-Arc Basin. Bull Japanese Soc Microb Ecol. 1989;3(2):73–82.

541   27.   Breusing C, Johnson SB, Tunnicliffe V, Clague DA, Vrijenhoek RC, Beinart RA.
542   Allopatric and sympatric drivers of speciation in *Alviniconcha* hydrothermal vent
543   snails. Mol Biol Evol. 2020;37(12):3469–3484.

544   28.   Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
545   sequence data. Bioinformatics. 2014;30(15):2114–2120.

546   29.   Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node
547   solution for large and complex metagenomics assembly via succinct de Bruijn
548   graph. Bioinformatics. 2015;31(10):1674–1676.

549   30.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
550   SPAdes: A new genome assembly algorithm and its applications to single-cell
551   sequencing. J Comput Biol. 2012;19(5):455–477.

552 31. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile

553   metagenomic assembler. Genome Res. 2017;27(5):824–834.

554 32. Seah BKB, Gruber-Vodicka HR. gbtools: Interactive visualization of metagenome

555   bins in R. Front Microbiol. 2015;6:1451.

556 33. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-

557   assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578–579.

558 34. Luo J, Lyu M, Chen R, Zhang X, Luo H, Yan C. SLR: a scaffolding algorithm

559   based on long reads and contig classification. BMC Bioinformatics. 2019;20(1):1–

560   11.

561 35. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome

562   Biol. 2012;13(6).

563 36. Xu GC, Xu TJ, Zhu R, Zhang Y, Li SQ, Wang HW, et al. LR-Gapcloser: A tiling

564   path-based gap closer that uses long reads to complete genome assembly.

565   Gigascience. 2018;8(1):1–14.

566 37. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon:

567   An integrated tool for comprehensive microbial variant detection and genome

568   assembly improvement. PLoS ONE. 2014;9(11):e112963.

569 38. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated

570   binning method to recover individual genomes from metagenomes using.

571   Microbiome. 2014;2(26):4904–4909.

572 39. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for

573   genome assemblies. Bioinformatics. 2013;29(8):1072–1075.

574 40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM:

575   assessing the quality of microbial genomes recovered from isolates, single cells,

576   and metagenomes. Genome Res. 2015;25(7):1043–1055.

577 41. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify

578   genomes with the genome taxonomy database. Bioinformatics. 2020;36(6):1925–

579   1927.

580 42. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High

581   throughput ANI analysis of 90K prokaryotic genomes reveals clear species

582   boundaries. Nat Commun. 2018;9(1):5114.

583 43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler

584   A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference

585   in the genomic era. Mol Biol Evol. 2020;37(5):1530–1534.

586 44. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o:

587   An advanced analysis and visualization platformfor 'omics data. PeerJ.

588   2015;2015(10):1–29.

589 45. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al.

590   Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome

591   Biol. 2020;21(1):1–21.

592 46. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer;

593   2016.

594 47. Oksanen AJ, Blanchet FG, Kindt R, Legen- P, Minchin PR, Hara RBO, et al.

595   vegan: Community Ecology Package. 2018. https://CRAN.R-

596   project.org/package=vegan.

597 48. RStudio Team. RStudio: Integrated Development Environment for R. Boston, MA:

598 RStudio, PBC; 2021.

599 49. Romero Picazo D, Dagan T, Ansorge R, Petersen JM, Dubilier N, Kupczok A.

600 Horizontally transmitted symbiont populations in deep-sea mussels are genetically

601 isolated. ISME J. 2019;13(12):2954–2968.

602 50. Kolde R. pheatmap: Pretty Heatmaps. 2019. https://CRAN.R-

603 project.org/package=pheatmap.

604 51. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al.

605 Integrating taxonomic, functional, and strain-level profiling of diverse microbial

606 communities with biobakery 3. Elife. 2021;10:1–42.

607 52. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics

608 2014;30(14):2068–2069.

609 53. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a

610 reference resource for gene and protein annotation. Nucleic Acids Res.

611 2016;44(D1):D457–462.

612 54. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly

613 accurate protein structure prediction with AlphaFold. Nature.

614 2021;596(7873):583–589.

615 55. Heiberger RM, Holland B. Statistical Analysis and Data Display: An Intermediate

616 Course with Examples in R. 2nd edition. New York: Springer; 2015.

617 56. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA

618 ribosomal RNA gene database project: improved data processing and web-based

619 tools. Nucleic Acids Res. 2013;41(D1):590–596.

620 57. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System. Mol Ecol

621   Notes. 2007;7(3):355–364.

622   58.   Nunoura T, Takaki Y, Kazama H, Kakuta J, Shimamura S, Makita H.

623   Physiological and genomic features of a novel sulfur-oxidizing

624   Gammaproteobacterium belonging to a previously uncultivated symbiotic lineage

625   isolated from a hydrothermal vent. PLoS ONE. 2014;9(8):e104959.

626   59.   Suzuki Y, Kojima S, Sasaki T, Suzuki M, Utsumi T, Watanabe H, et al. Host-

627   symbiont relationships in hydrothermal vent gastropods of the genus *Alviniconcha*

628   from the Southwest Pacific. Appl Environ Microbiol. 2006;72(2):1388–1393.

629   60.   Ücker M, Ansorge R, Sato Y, Sayavedra L, Breusing C, Dubilier N. Deep-sea

630   mussels from a hybrid zone on the Mid-Atlantic Ridge host genetically

631   indistinguishable symbionts. ISME J. 2021;15(10):3076–3083.

632   61.   Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally,

633   shaped locally. Nat Rev Microbiol. 2019;17(5):271–283.

634   62.   Vrijenhoek RC. Genetic diversity and connectivity of deep-sea hydrothermal vent

635   metapopulations. Mol Ecol. 2010;4391–4411.

636   63.   Kádár E, Bettencourt R, Costa V, Santos RS, Lobo-da-Cunha A, Dando P.

637   Experimentally induced endosymbiont loss and re-acquirement in the

638   hydrothermal vent bivalve *Bathymodiolus azoricus*. J Exp Mar Bio Ecol.

639   2005;318(1):99–110.

640   64.   Soto W, Travisano M, Tolleson AR, Nishiguchi MK. Symbiont evolution during the

641   free-living phase can improve host colonization. Microbiology (United Kingdom).

642   2019;165(2):174–187.

643   65.   Breusing C, Johnson SB, Mitarai S, Beinart RA, Tunnicliffe V. Differential patterns

of connectivity in Western Pacific hydrothermal vent metapopulations: A comparison of biophysical and genetic models. Evol Appl. 2021;doi: 10.1111/eva.13326.

66. Vrijenhoek RC, Duhaime M, Jones WJ. Subtype variation among bacterial endosymbionts of tubeworms (Annelida: Siboglinidae) from the Gulf of California. Biol Bull. 2007;212(3):180–184.

67. Davies SW, Moreland KN, Wham DC, Kanke MR, Matz M V. *Cladocopium* community divergence in two *Acropora* coral hosts across multiple spatial scales. Mol Ecol. 2020;29(23):4559–4572.

68. Stewart LC, Algar CK, Fortunato CS, Larson BI, Vallino JJ, Huber JA, et al. Fluid geochemistry, local hydrology, and metabolic activity define methanogen community size and composition in deep-sea hydrothermal vents. ISME J. 2019;13(7):1711–1721.

69. Fortunato CS, Larson B, Butterfield DA, Huber JA. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ Microbiol. 2018;20(2):769–784.

70. Ishibashi J ichiro, Tsunogai U, Toki T, Ebina N, Gamo T, Sano Y, et al. Chemical composition of hydrothermal fluids in the central and southern Mariana Trough backarc basin. Deep Res Part II Top Stud Oceanogr. 2015;121:126–136.

71. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. Nature. 2016;536(7617):425–430.

72. Thomas E, Anderson RE, Li V, Rogan LJ, Huber JA. Diverse viruses in deep-sea hydrothermal vent fluids have restricted dispersal across ocean basins.

667   mSystems. 2021;6(3):1–18.

668

Figure Legends

Figure 1. Map of the Mariana Back-Arc Basin adapted from ref. (25), indicating the locations of the two main hydrothermal vent fields sampled in this study, Illium and Hafa Adai. Free-living symbiont samples were obtained at low coverage from two additional vent sites investigated in ref. (25), Alice Springs and Burke. Colors associated with vent fields are used consistently across samples. A light cyan and teal are used for Hafa Adai, representing Voodoo Crater 1 and 2, respectively.

Figure 2. Cladogram branch transformed phylogenomic tree of chemosynthetic Gammaproteobacteria based on 70 single copy core genes. Genome accession numbers for all genomes included in the phylogenomic analysis can be found in Supplementary Table 1. *A. hessleri* symbionts are colored by vent field (Hafa Adai VC2: teal, Hafa Adai VC1: light cyan, Illium: yellow) with dots and triangles indicating free-living or host-associated lifestyle. *Candidatus* Pseudothioglobus singularis was used as outgroup for tree rooting.

Figure 3. Symbiont population structure based on single nucleotide polymorphisms. A.) Heatmap of genome-wide, pairwise fixation indices (F_{ST}) created using the pheatmap package in RStudio. F_{ST} values range from 0 to 1, where a value of 0 indicates no genetic differentiation, while a value of 1 indicates complete isolation among populations. "FL" and "HA" indicate free-living and host-associated symbionts,
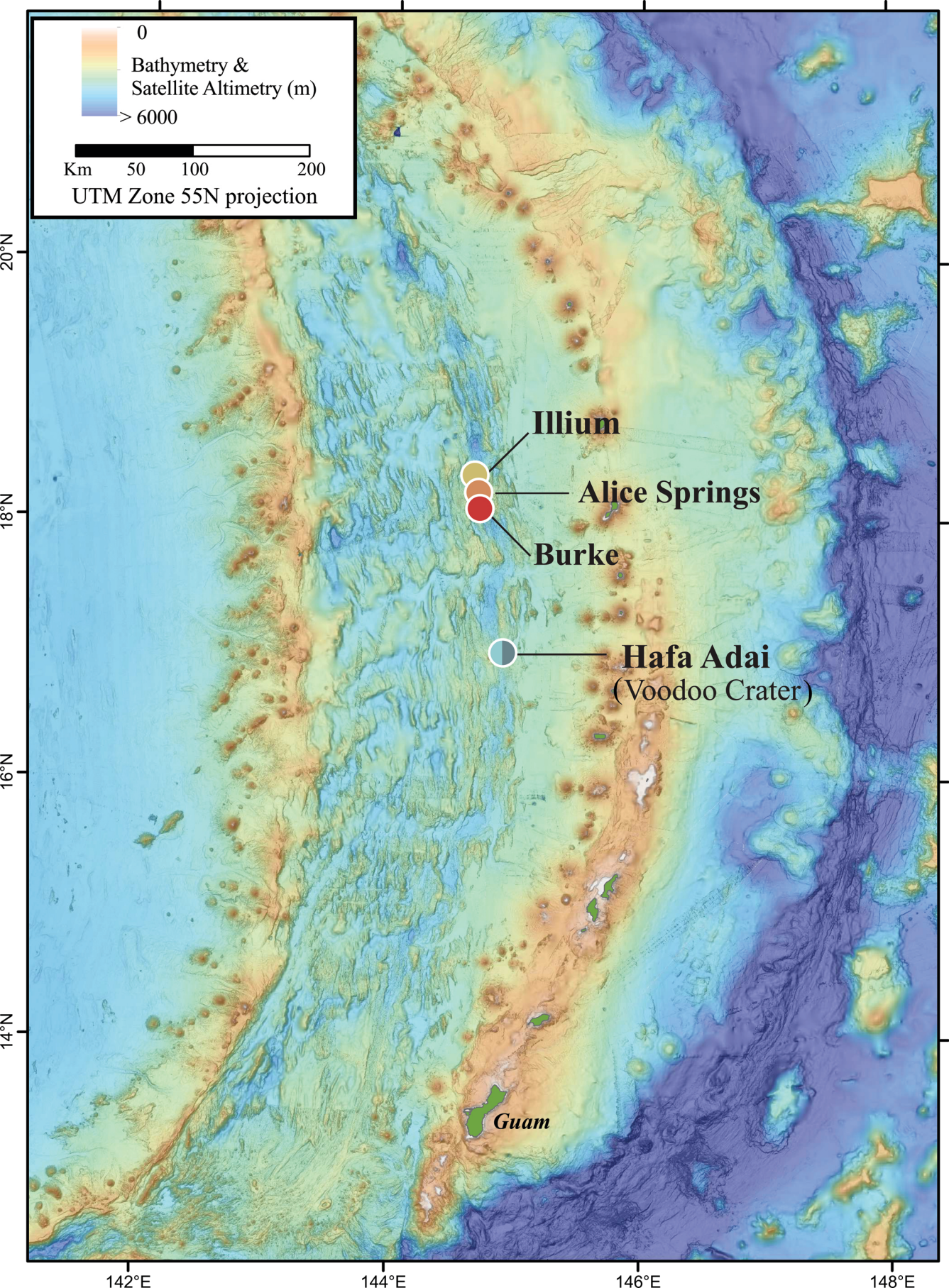
690   respectively. B.) PCoA plot based on Bray-Curtis distances illustrating the population

691   structure of *A. hessleri* symbionts of different lifestyles and vent fields.
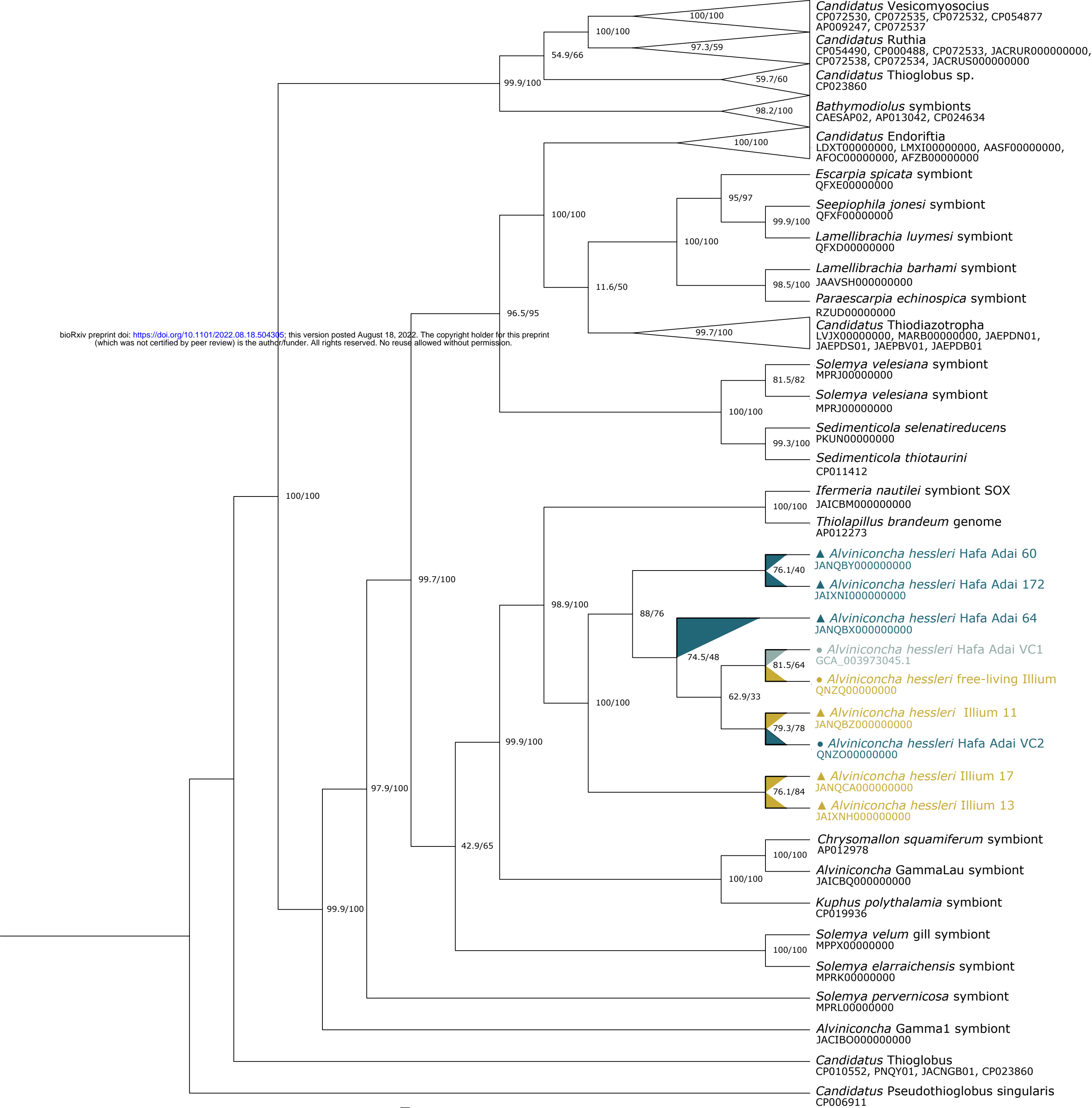
692

693   **Figure 4.** PCoA plot based on Jaccard distances illustrating the difference in gene

694   content between *A. hessleri* symbionts based on both lifestyle and vent field.

695

696   **Figure 5.** Likert plots showing A) number of unique genes between Illium and Hafa Adai

697   (including both VC1 and VC2), and B) number of unique genes between Illium and Hafa
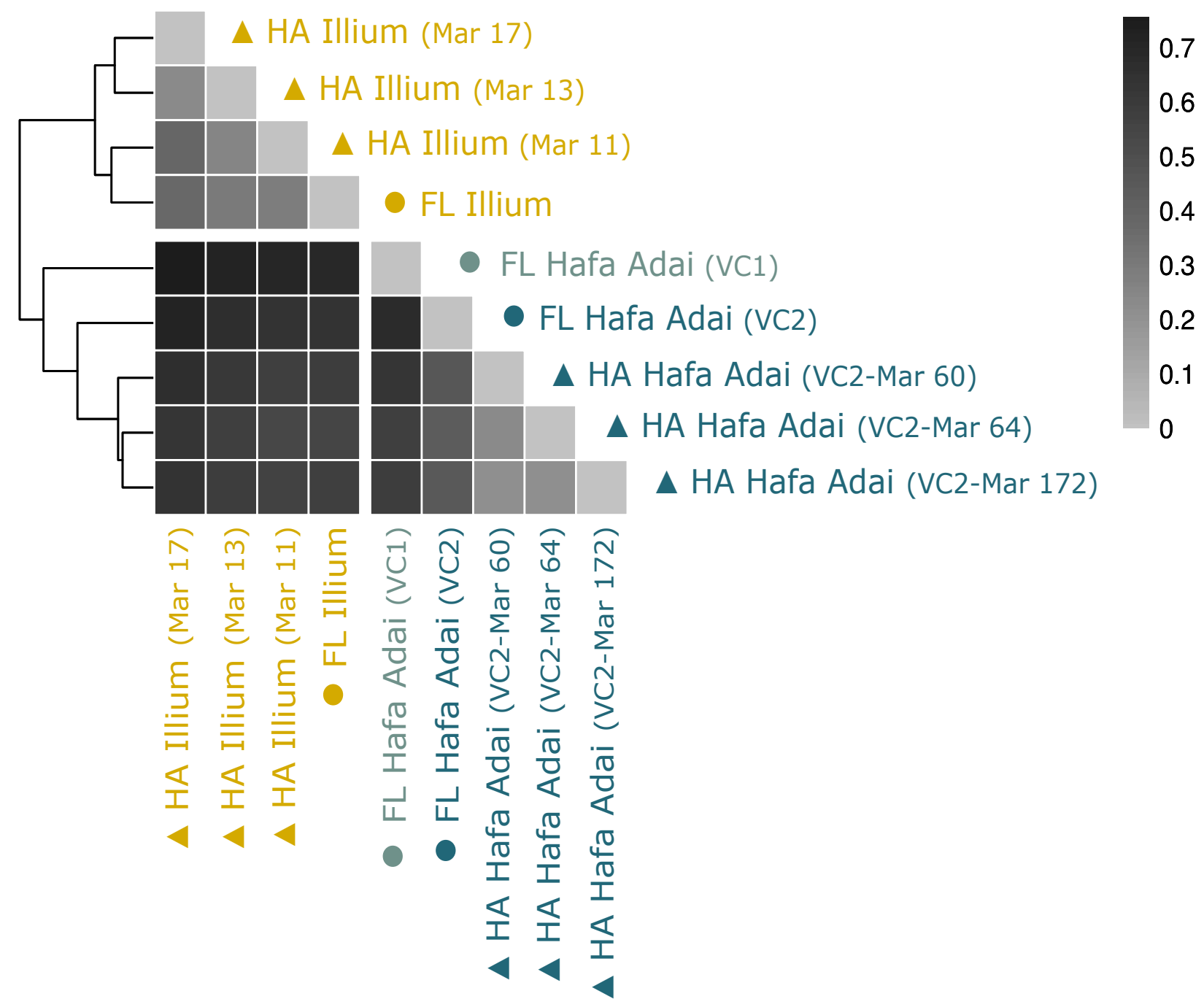
698   Adai (VC2 only).

Table 1. Percentage of unambiguous reads in environmental samples mapping to *Alviniconcha* symbiont 16S rRNA and mitochondrial CO1 genes, as well as the ratio of 16S rRNA to mtCO1.

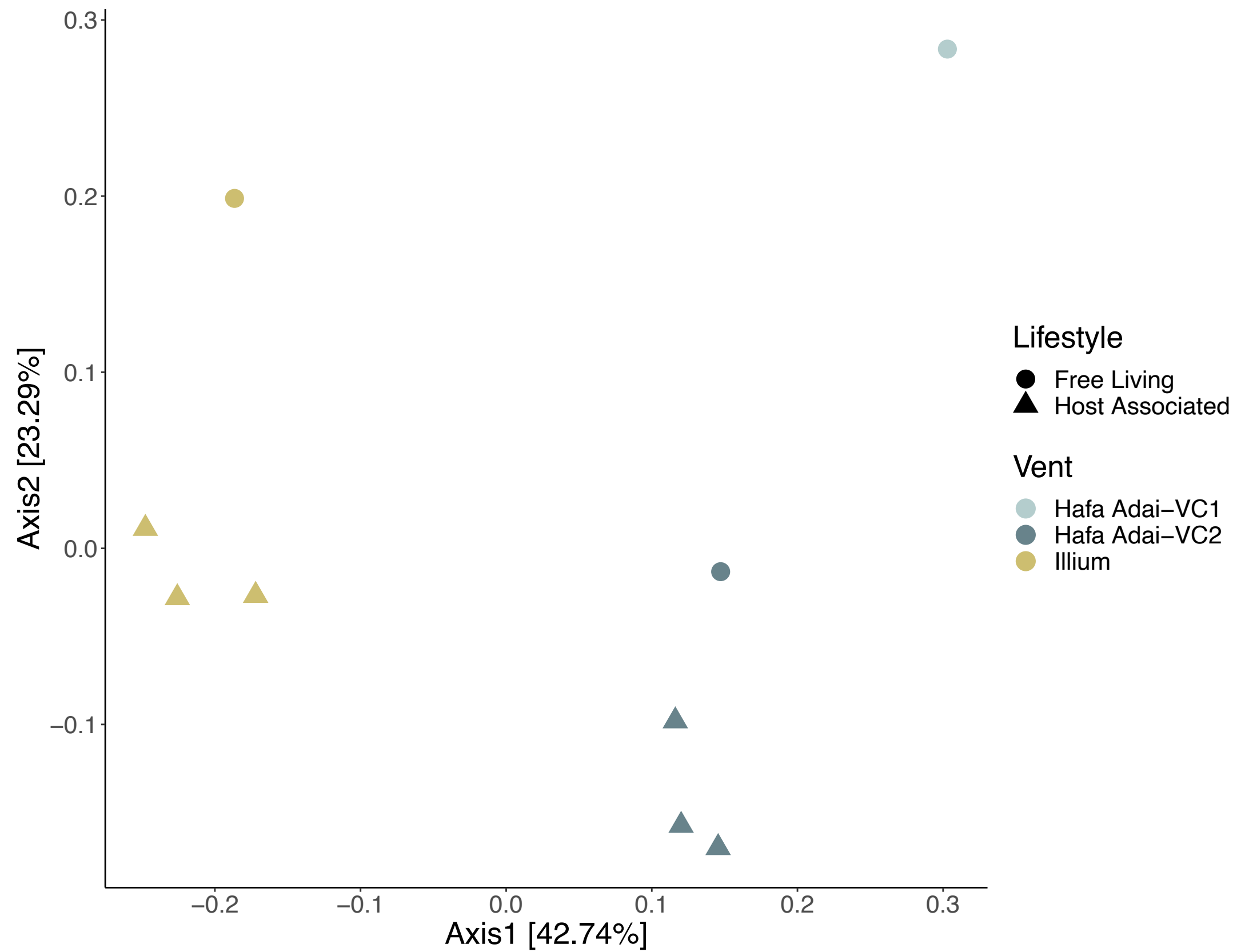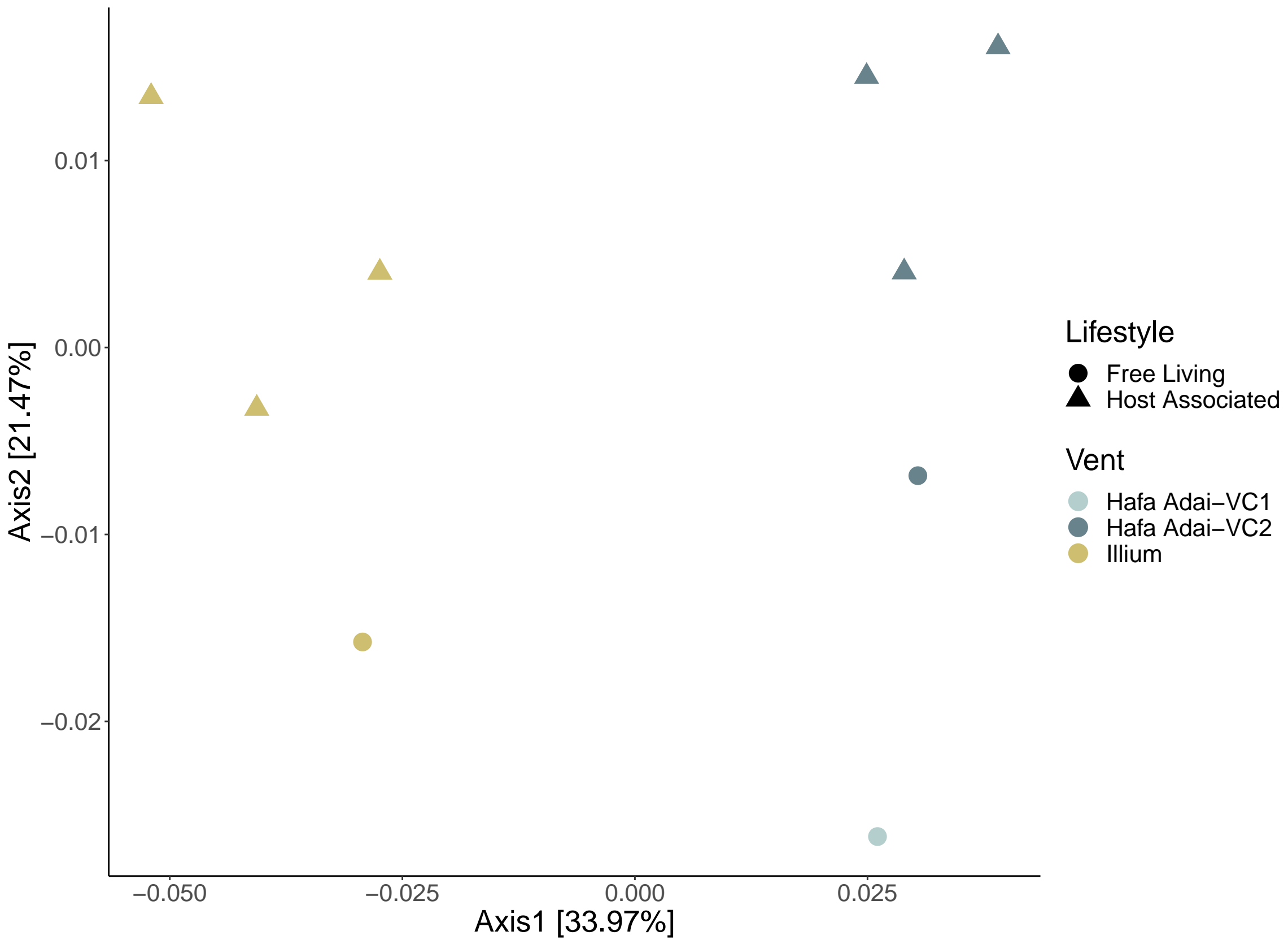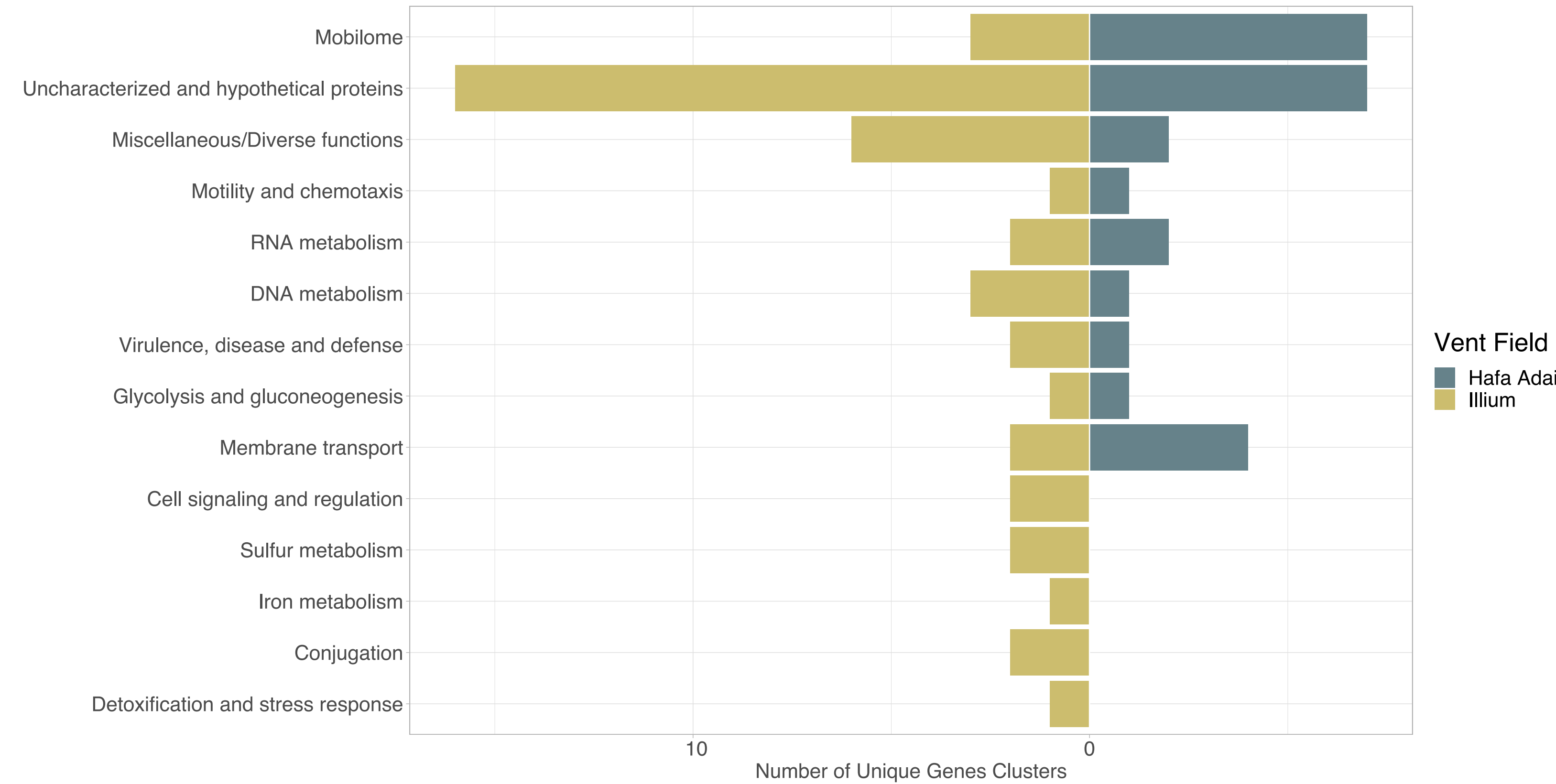| Sample	*Alviniconcha* 16S rRNA	*Alviniconcha* mtCO1	16S rRNA:CO1	Origin
Host-associated Hafa Adai 172	0.01859	0.00319	5.82758621	Gill tissue
Host-associated Hafa Adai 60	0.01627	0.01112	1.4631295	Gill tissue
Host-associated Hafa Adai 64	0.01706	0.01019	1.67419038	Gill tissue
Host-associated Illium 11	0.01116	0.00844	1.32227488	Gill tissue
Host-associated Illium 13	0.02004	0.00802	2.49875312	Gill tissue
Host-associated Illium 17	0.01435	0.00721	1.99029126	Gill tissue
Free-living Illium	0.04302	0.00218	19.733945	Diffuse fluids
Free-living Hafa Adai- VC2	0.08404	0.00193	43.5440415	Diffuse fluids
Free-living Hafa Adai- VC1	0.06526	0.00027	241.703704	Diffuse fluids

a)

- ▲ HA Illium (Mar 17)
- ▲ HA Illium (Mar 13)
- ▲ HA Illium (Mar 11)
- ● FL Illium
- ● FL Hafa Adai (VC1)
- ● FL Hafa Adai (VC2)
- ▲ HA Hafa Adai (VC2-Mar 60)
- ▲ HA Hafa Adai (VC2-Mar 64)
- ▲ HA Hafa Adai (VC2-Mar 172)

b)

Axis1 [42.74%]

Axis2 [23.29%]

Lifestyle
- ● Free Living
- ▲ Host Associated

Vent
- ● Hafa Adai−VC1
- ● Hafa Adai−VC2
- ● Illium

a)

b)